THREE ESSAYS ON THE APPLICATION OF BEHAVIOURAL FINANCE AND TEXTUAL ANALYSIS TO FINANCIAL MARKETS

A thesis submitted for the degree of

Doctor of Philosophy

of The Australian National University

by

Yufei Li

College of Business and Economics

Research School of Finance, Actuarial Studies and Statistics

Canberra, Australia

October 2024

Declaration

I hereby declare that this thesis is my original work except where the work of others has been cited. The contents of this thesis have not been submitted for a degree or diploma at another institution.

_____

Yufei Li

Acknowledgements

I would like to express my sincere gratitude to my advisor, Takeshi Yamada, for his invaluable guidance, encouragement, and support throughout my doctoral studies. His patience and wisdom were instrumental in my ability to complete this dissertation. Takeshi provided not only academic guidance but also personal support during this challenging period.

My profound gratitude extends to the distinguished members of my supervisory panel: Takeshi Yamada, Xin (Kelly) Liu, and Jo Drienko, for their invaluable guidance and support.

I am also deeply grateful to the Australian National University for providing me the opportunity to pursue my doctoral degree in Finance. Despite being unable to return to Canberra due to COVID-19, I received excellent resources and support for remote learning. Special thanks to the Research School of Finance, Actuarial Studies, and Statistics for their support and encouragement throughout my doctoral journey.

During this period, I faced numerous challenges and am deeply indebted to my parents for their unwavering support and encouragement. Their constant presence and unconditional love provided me with the greatest motivation to persevere.

Additionally, heartfelt thanks go to my friends and colleagues – Belinda, Jiancong, Thao, Sayla, Jeremy, Hedda, Shuang, Weijian, and Dan. Their constant support and inspiration were invaluable throughout this lengthy academic journey. I also wish to express deep appreciation to the members of the school's IT team for their invaluable technical support and assistance with the technological challenges I encountered. Additionally, I am deeply grateful to the former HDR RSFAS Research Dean, Timothy Higgins, office administrator Donna Webster, and staff members Emma Schultz and Anna Pickering for their assistance with administrative

matters. Furthermore, I am grateful to teachers Xiaoting, Kun Li, Xianming, Daruo, Qiaoqiao, Phong, Dingding, and Lili for their academic assistance. Their wisdom and guidance were crucial in the completion of this dissertation.

Finally, I would like to express my appreciation to all those who provided me with their assistance and support during this journey. Thank you for being a part of my life and helping me to achieve my academic goals.

To mom, dad, and grandmom

for everything

Abstract

This thesis comprises three chapters. The first, 'Retail investor attention & its effect on stock price reactions to news', describes the impact of individual investor attention on stock market mispricing of news. We found that mispricing is more prevalent after negative news as a result of high abnormal retail attention, in contrast to a prediction from limited attention theory that investor attention reduces stock mispricing and improves market efficiency. Using Google Search volumes as a proxy for investor attention, we found that high unexpected search queries for a target firm do not demonstrate that investors are focusing on firm-specific information. Rather, unusual search volumes may be driven by greater investor uncertainty or the participation of less-informed retail investors, leading to a stronger market underreaction to negative news. We also found that the characteristics of firms with high retail investor attention (small size, low institutional ownership, high analyst dispersion, high idiosyncratic volatility and high short interest) are indicators of market inefficiency and are negatively associated with the speed of information diffusion.

The second chapter, 'MD&A disclosure & investors' reactions to management forecasts', shows that the Management Discussion and Analysis (MD&A) disclosure in the 10-K report may affect price reactions to management forecast news. The study extracted MD&A sections from 10-K reports and used readability, uncertainty and tone change to capture the quality of MD&A disclosures using an approach developed by Loughran and McDonald (2011). Our findings suggest that these factors influence the price reaction to management forecasts and that the price underreaction, measured as the post-announcement cumulative abnormal

return, is significantly associated with MD&A readability and uncertainty. We found that the market reaction to management forecasts was stronger for firms with more uncertain, less readable and more optimistic MD&A disclosures. The price underreaction to management forecast news was stronger for firms with more uncertain and less readable disclosures. We found evidence for a positive relation between MD&A readability and forecast precision, indicating that managers' motives to reduce information asymmetry by issuing clear disclosures align with their motives to enhance forecast precision.

The third chapter, 'How global sentiment affects A–H cross-listed share mispricing in China: mechanisms of global sentiment contagion in a segmented market,' investigates the impact of global and local sentiment on China's A–H cross-listed stock price spread using market accessibility to global investors and risk exposure. The study aimed to provide a global sentiment–driven explanation of A–H price disparities. We investigated how policy options for Chinese equity markets to integrate into international financial markets affected sentiment effects on A–H premiums. We found that the relaxation of foreign ownership restrictions mitigated the negative impacts of global sentiment on A–H price disparities. We also identified the channel of global sentiment contagion in driving portfolio investment. Finally, we found that the higher relative arbitrage costs for A-shares amplify the negative impacts of global sentiment on A–H premiums, indicating an asymmetric influence of global sentiment across different levels of relative arbitrage risk.

Contents

List of Tables

List of Figures

List of Abbreviations

ADR	American depositary receipts
AGSV	abnormal Google search volume
B/M	book-to-market ratio
BHAAR	buy-and-hold average abnormal return.
BHAR	buy-and-hold abnormal return
CAR	cumulative abnormal returns
CIG	company-issued guideline
CRSP	Center for Research in Security Prices
CSMAR	China Stock Market & Accounting Research
CSRC	China Securities Regulatory Commission
DGTW	Daniel, Grinblatt, Titman, and Wermers
EDGAR	Electronic Data Gathering, Analysis, and Retrieval
EPS	earnings per share
GMM	Generalized Method of Moments
HK	Hong Kong
IBES	Institutional Brokers' Estimate System
IOR	Institutional ownership ratio
IPO	initial public offering
MD&A	Management's Discussion and Analysis

MF	Management (earnings per share) forecast
MSCI	Morgan Stanley Capital International
PCA	principal component analysis
QFII	qualified foreign institutional investor
RQFII	Renminbi-Qualified Foreign Institutional Investor
SDC	Securities Data Company
SEC	Securities and Exchange Commission
SEHK	Hong Kong Stock Exchange
SHSC	Shanghai–Hong Kong Stock Connect
SIC	Standard Industrial Classification
SIR	Short interest ratio
SMB	small minus big
SPR	spread gap
SVI	search volume index

Chapter 1: Retail investor attention & its effect on stock price reactions to news

1.1 Introduction

Empirical studies have demonstrated a consistent phenomenon in stock markets whereby stock prices drift following the release of significant news, indicating an initial underreaction to information. Chan (2003) showed that stocks that release news, particularly \negative news, tend to show a price momentum, resulting in a negative drift that can persist for up to 12 months. This finding suggests that the market is slow to incorporate public news into stock prices. While traditional asset pricing models assume instantaneous incorporation of information, the limited cognitive resources of investors can impede their ability to process news promptly. Insufficient attention to an asset can affect stock prices, as indicated by various theoretical and empirical studies (see Barber & Odean, 2008; Engelberg & Parsons, 2011; Seasholes & Wu, 2007). Research has suggested that market underreaction is linked to limited investor attention, with a lack of focus on fundamental news resulting in slow information diffusion. The significance of information only becomes evident in stock prices after investors have paid attention to it (Merton, 1987), and the speed at which information is integrated into stock prices is influenced by the level of investor attention (Loh, 2010). Further, behavioural biases, such as the disposition effect[1], can lead to irrational trading behaviours in individual investors. Attention alone does not guarantee rational pricing, as investor sophistication is another crucial factor.

[1] The disposition effect in behavioral finance, a concept identified by Shefrin and Statman (1985), describes the tendency of investors to sell assets that have appreciated in value and hold onto those that have depreciated.

Because professional knowledge and efficient channels for processing and reacting to news are necessary, individual investors sometimes lack the skills to fully comprehend the information presented and make informed decisions based on it. This can hinder their ability to react immediately to news. The speed of investors' reactions to news depends not only on the presence of attentive investors in the market but also on their ability to process information efficiently. To understand investor attention, one needs an understanding of the different types of investors, as attention can have effects on stock returns that vary with the information-processing abilities of investors. In this study, we aimed to elucidate the impact of attention from individual investors, specifically retail investors, on stock returns following the release of public news.

The literature has primarily investigated the effects of public media news and investor attention on the stock market and has linked media news coverage to investor attention. Our study aimed to explore the relationship between investor attention and the pricing of public news by disentangling information supply and demand. Specifically, we investigated how retail investor attention affects stock price reactions to different types of public news, considering the quantity of public news and public sentiment. This analysis sought to shed light on how retail investor attention can shape investors' behaviour and decision-making in response to public news. Previous studies have used media coverage, such as the number of articles reporting on the firm, as a proxy for investor attention. However, this proxy is not always valid, as news can only attract investors' attention when they have noticed it. Google's online search volume serves as a better measure of retail investor attention, as individual investors rely predominantly on this platform for financial information, compared with institutional investors who have

access to more sophisticated information channels such as Bloomberg (https://www.bloomberg.com; see Ben-Rephael et al., 2017). In this study, we utilised the aggregate search volume index (SVI) from Google Trends[2], a metric developed by Da et al. (2011), to measure retail investor attention.

Other studies have examined specific financial events and assessed the underreaction to news events within a short timeframe. The literature has emphasised the significance of investor attention limitations for post-earnings announcement drift, for example, in studies by Bernard and Thomas (1989), Foster et al. (1984) and Hirshleifer and Teoh (2003). Their findings suggest that higher levels of investor attention correspond to improved signal interpretation and quicker assimilation of earnings news. However, investors need expertise and skill to process information and make informed investment decisions after they notice announcements. Most investors access public media news directly through online channels. Chan (2003) employed a unique approach, examining a database of company news headlines from the Dow Jones Interactive Publications Library, which includes periodicals, newspapers and newswires. The primary aim was to compare the monthly returns of stocks associated with public news ('news stocks') with those not associated with news. To differentiate between positive and negative news effects, Chan established monthly rankings of news stocks based on raw returns. Only stocks that traded during a given month were included, with the highest and lowest third being designated as the 'good news' and 'bad news' groups, respectively. There was a significant drift in stock prices following bad news, suggesting that investors react slowly to negative information. In a similar manner, our study examined the price reaction to the aggregate

[2] For more detailed information and to use Google Trends, you can visit https://trends.google.com/trends/.

intensity of public positive and negative news during a month. After identifying firm-specific news articles, we used sentiment indicators provided by Ravenpack to determine whether the news was good, bad or neutral. We counted the number of positive and negative news articles each month to measure the rate of information arrival. The results revealed that a higher level of abnormal investor attention enhances the current price reaction to public news and also strengthens the news drift effect in the following three to six months. Underreaction is more marked after negative news. These findings suggest that a higher level of retail investor attention may reflect the involvement of individual investors who can be considered uninformed 'noise traders'. According to rational structural uncertainty theory (Brav & Heaton, 2002), the distribution of information in the economy affects market anomalies. A key aspect of this theory is the asymmetry of information distribution among different market participants. For instance, institutional investors typically have more advanced tools and resources for information processing compared to individual traders, leading to an uneven spread of information across the market. The findings of this thesis align with the theory, demonstrating that greater attention from retail investors, often considered less informed, correlates with a pronounced underreaction in the market to public news in the medium term. This suggests that the involvement of these retail investors, who might not have access to or the ability to process information as efficiently as institutional investors, contributes to market inefficiencies. However, we cannot rule out alternative explanations, such as that frictions arising from retail investor searches hinder the incorporation of public information into stock prices, particularly for negative news.

To test the hypothesis that retail investor attention affects the speed of stock price reaction to public news, we used a sample of US stock-month observations from 2004 to 2018. Using Fama and MacBeth's (1973) cross-sectional regressions, our analysis revealed a significant impact of retail investor attention on the price reaction to public news, indicating that higher levels of retail investor attention can strengthen news underreactions. Our analysis of long-short investment strategies further confirmed the underreaction to public news, with a particularly marked underreaction to negative news observed in the high-attention retail investor group. A non-significant underreaction of long–short hedge portfolios to positive news, different from our Fama–MacBeth regression results, may be partly driven by unobserved firm characteristics that have asymmetric sensitivity to positive and negative news. Our portfolio strategy results may be unable to capture this directly. Our findings are consistent with research by Chan (2003) that highlights the prevalence of underreaction specifically to negative news. Our findings align with the hypothesis that the dissemination of firm-specific information, particularly negative information, occurs only gradually in the investing public (Hong et al., 2010).

In addressing the concerns regarding the potential conflation of Google search volumes with the salience of news content, the study employs a comprehensive empirical strategy that differentiates these influences on stock returns. Notably, it incorporates interaction terms between abnormal Google search volumes (AGSV) and both positive and negative news variables, and controls for lagged effects of news to isolate the unique contributions of heightened retail investor attention from the inherent impact of the news itself. Further, by orthogonalizing AGSV relative to news variables, this study effectively separates the

component of investor attention that is independent of immediate news content, allowing for a precise analysis of the unique impact of exogenous investor attention on stock returns. These methods address the intricacies involved in disentangling these effects, where both salient news can prompt increased searches and heightened search volumes may amplify the perceived importance of news. The results underscore the critical role of independent investor attention in driving stock price dynamics, suggesting that while news content inherently influences market reactions, the orthogonalized measure of AGSV captures a purer form of investor behavioral response. These findings highlight the need for future research to employ cleaner measures and innovative methodologies, such as natural experiments or machine learning techniques, to further refine our understanding of how digital footprints of investor attention interact with traditional news metrics to shape market dynamics

In further analyses, we investigated the profits of positive and negative news trading strategies for various levels of retail investor attention, applying the Fama–French three-factor model (Fama & French, 1992), the four-factor Fama–French model with momentum (Carhart, 1997) and the five-factor Fama–French model with investment and profitability. Our findings suggest that only a long–short strategy for negative news can generate significant positive alpha, conditional on a high level of retail investor attention. The negative news effect was more persistent than the positive news effect; there is no evidence that a positive news trading strategy can generate significant profits. Contrary to the literature, our study found that high retail attention focused on a firm does not necessarily lead to improved market efficiency but can further enhance the mispricing of negative news.

We ensured the robustness of our primary finding using various approaches. Although retail investor attention is potentially useful as an indicator of the speed of information diffusion, it is probable that it encompasses other variables that could confound our results. We replicated the regression analysis while controlling for firm characteristics, including analyst earnings forecast dispersion, institutional ownership, firm size, idiosyncratic volatility and short interest. We conducted a portfolio analysis and examined whether negative news mispricing is solely driven by firms' characteristics. We found that the abnormal returns from a negative news trading strategy were significant for firms with high levels of abnormal retail investor attention, particularly those with small size, low institutional ownership, high analyst dispersion, high idiosyncratic volatility and high short interest. These characteristics are all indicators of market efficiency or factors associated with the speed of information diffusion. For each characteristic, firms with high levels of retail attention generated statistically and economically significant alpha as assessed by the Fama–French three-factor model, the four-factor model with momentum and five-factor model with investment and profitability. In a separate analysis, we substituted positive and negative news intensity with news coverage and news sentiment and performed Fama–MacBeth and Fama–French regressions to eliminate the possibility that news intensity or sentiment was responsible for our results. The results confirmed that the negative news drift (momentum) effect and news mispricing are not driven solely by investors' underreaction to news tone or news intensity. The results revealed an asymmetry in investors' sensitivity to positive and negative news shocks, with the price underreaction to negative news being more pronounced. In a final analysis, we split our sample into two subperiods. We obtained similar results.

The remainder of Chapter 1 is organised as follows. A literature review is presented in Section 1.2, followed by the data and sample description in Section 1.3. The empirical results and discussion are given in Section 1.4. Additional analysis and robustness checks are presented in Section 1.5. Finally, the conclusions of the study are presented in Section 1.6.

1.2 Literature review

Extensive research supports underreactions of stock prices following specific news events. The literature has primarily focused on the incorporation of news into stock pricing, particularly examining underreaction after significant financial events (see Mitchell & Mulherin, 1994; Tetlock et al., 2008). Various financial news events have been explored, such as dividend initiations or omissions (Michaely et al., 1995), earnings surprises or announcements (Bernard & Thomas, 1990), changes in analyst recommendations (Womack, 1996), seasoned equity offerings (Loughran & Ritter, 1995), share repurchases (Ikenberry et al., 1995) and price change events (Pritamani & Singal, 2001). Underreactions have been explained by two competing theories: behavioural theory (Daniel et al., 1998; DeLong et al., 1990) and rational structural uncertainty theory (Brav & Heaton, 2002). Behavioural theory suggests that investors tend to underreact to public signals and overreact to private information and pure price movements. Limited attention theory, as a variant of behavioural theory, provides a theoretical model to predict the underreaction to public information (see Hirshleifer & Teoh, 2003). Limited attention theory suggests that underreaction to news is positively associated with investor inattention (see Hirshleifer et al., 2009). In contrast, rational structural uncertainty theory suggests that information distribution influences underreaction by financial markets. In this theoretical framework, the arrival rate of informed traders and the level of structural

uncertainty predict news drift. Hence, news attributes, investor irrationality and market participant types influence the speed and magnitude of market reactions to news.

Ben-Rephael et al. (2017) emphasised the significance of institutional investor attention for the assimilation of earnings news into stock prices but treated the impact of retail investor attention as limited. Welagedara et al. (2017) explored the differences between retail and non-retail investor attention using SVI as an indicator of retail investor attention. That study showed that stocks receiving greater attention from individual investors exhibited a more pronounced underreaction following analyst recommendation upgrades compared with stocks attracting institutional investor attention. The finding implied that individual investors, despite their attentiveness, do not consistently exhibit rational behaviour. It emphasised the importance of investor attention and sophistication to price discovery within the market.

Some researchers, such as Engelberg and Parsons (2011), have utilised media or news coverage as a proxy for investor attention. However, this approach has certain limitations. First, news coverage may not accurately reflect investor attention, given the abundance of information in the modern era, which can lead to a scarcity of attention. Second, the influence of the media on stock markets is complex. The media plays a crucial role in disseminating information and can help reduce capital costs for firms, as noted by Fang and Peress (2009). However, the tone of media coverage, especially when negative, can affect stock prices, as demonstrated by Tetlock et al. (2008). Engelberg and Parsons (2011) have suggested that more media coverage generally leads to better-informed trading, but Tetlock (2007) and Garcia (2013) found that it can distort prices temporarily because of sentiment. The precise role of the media in markets remains a subject of debate, deemed beneficial in some respects but potentially

misleading in others. Third, different types of news may attract varying levels of attention, with negative news more likely to attract investors' interest. Negativity bias theory suggests that negative events of equal intensity tend to have a stronger psychological impact than neutral or positive ones (see Kanouse & Hanson, 1972). Hence, distinguishing between news demand (the attention investors pay to news) and news supply (the quantity and nature of the news reported) is essential. This distinction is crucial for understanding investor responses to news, recognising the importance of not equating media coverage with investor attention. Several indirect proxies for attention have been proposed, including advertising expenses (Chemmanur & Yan, 2009; Lou, 2014), trading volume (Gervais et al., 2001), ceiling price (Seasholes & Wu, 2007) and excessive trading volume and stock returns (Barber & Odean, 2008). The internet has revolutionised information searches and dissemination, making it an indispensable tool in finance studies. Studies by Da et al. (2011, 2015 and Dzielinski (2012) have incorporated internet search volume as a key variable. There are three advantages of using Google Search volume rather than other proxies. First, this measure can capture retail investors' search interests, avoiding the inclusion of economic outcomes associated with firms rather than with investor attention. Second, Google Search volume is a measure for assessing the attention of retail investors. It is often used with Bloomberg search volume to differentiate between individual and institutional investors (Ben-Rephael et al., 2017), as institutional investors predominantly rely on professional financial outlets to access company information. Third, the SVI is not influenced by passive attention measures such as advertising expenses and news coverage (Ding & Hou, 2015).

This study aimed to address two fundamental research questions: What is the impact of positive and negative news on the stock market? Does retail investor attention exacerbate the market's underreaction to public news? If there is a link between retail investor attention and an efficient allocation of attention to firm-specific information, there should be a negative correlation between retail investor attention and the market's underreaction to public news. Such a finding would be in line with the limited attention theory. Conversely, if retail investor attention predominantly involves the entry of uninformed individual investors into the market or greater uncertainty regarding a firm, there should be a positive correlation between retail investor attention and news underreaction. Our results have shown that greater retail investor attention strengthens news drift effects, particularly for negative news. Specifically, our study used Google Search volume as a proxy for retail investor attention and found that a higher Google Search volume can reveal the entry of uninformed retail investors into the market, which exacerbates the market's underreaction to news. This finding supports the individual trading hypothesis (Bernard & Thomas, 1990) regarding the influence of individual investor trading on market efficiency.

1.3 Data & sample

The sample period ran from January 2004 to December 2018. To eliminate microstructure concerns, the sample excluded stock-month observations with prices lower than $3.00 or a trading period of less than 12 months. Stocks with non-valid Google Search indexing in the sample period were also removed. The final dataset included 2,719 stocks. Return, trading volume and market capitalisation data were obtained from the Center for Research in Security Prices (CRSP), and the accounting and short-interest data were sourced from Compustat. The

analyst coverage and analyst forecast dispersion[3] data were taken from the I/B/E/S database, while the institutional ownership data were acquired from 13F filings. The public news data were obtained from the Ravenpack database. For a detailed overview of variable definitions, please refer to Appendix A.

1.3.1 News measures

Rather than focusing on specific financial events and evaluating underreaction within a short time window, this study examined the price reaction to the aggregate intensity of public positive and negative news over a month. To accomplish this, a comprehensive news measure was employed that extracted data from various public news sources. Aggregating news articles on a monthly basis is desirable for several reasons. First, obtaining sufficient news variation data for firms on a daily basis is challenging, as most firms do not have news reported daily. Secondly, daily news effects are more likely to be influenced by the nature and content of the news, resulting in a noisier daily news measure. Finally, a monthly news measure is better suited to a news-sorted investment strategy and the evaluation of news mispricing. This study identified material news using the Ravenpack news database, which includes a broad range of news sources and allows for comprehensive coverage of the information sources of individual investors. Ravenpack employs advanced machine learning algorithms to process news from various sources, including Dow Jones Newswires, the Wall Street Journal, Barron's, direct regulatory feeds and social media websites. It recognises over 1,200 types of firm-specific news,

[3] To determine the analyst forecast dispersion, the analyst forecast for a given company–fiscal period end pair was carried forward to the next estimate release date for the same pair by the same analyst or to a date 105 days prior to the earnings announcement date.

with data available from 2000.[4] Additionally, Ravenpack records news arrival with Greenwich Mean Time time stamps, providing millisecond precision.[5]

Following the methodology of Shi et al. (2016), this study used four news-related measures (shown below) to measure news quantity and news sentiment:

$$\text{News\_Positve}_{i,t} = \sum_{all} \tau \frac{I(CSS_{i,\tau} > 50)|CSS_{i,\tau} - 50|REL_{i,\tau}}{100} \quad (1)$$

$$\text{News\_Negative}_{i,t} = \sum_{all} \tau \frac{I(CSS_{i,\tau} < 50)|CSS_{i,\tau} - 50|REL_{i,\tau}}{100} \quad (2)$$

$$\text{News\_Coverage}_{i,t} = \sum_{all} \tau \frac{REL_{i,\tau}}{100} \quad (3)$$

$$\text{News\_Sentiment}_{i,t} = \frac{\sum_{all} \tau \frac{REL_{i,\tau}(MCQ - 50)}{100}}{T} \quad (4)$$

where $\text{News\_Positive}_{i,t}$ and $\text{News\_Negative}_{i,t}$ reflect the number of positive and negative news stories, respectively, for stock *i* in month *t*. $\text{News\_Coverage}_{i,t}$ gives the total number of news stories covering stock *i* in month *t*. $\text{News\_Sentiment}_{i,t}$ gives the average sentiment score of news stories for stock *i* in month *t*, where *T* is the number of days in the month. A positive (negative) value for $\text{News\_Sentiment}_{i,t}$ indicates an overall positive (negative) sentiment towards the stock, where a value of zero indicates neutral sentiment. Relevance (REL) indicates how relevant a news story is to a firm from 0 to 100, with higher values denoting greater relevance. The sentiment score (CSS) measures the sentiment of a news story based on intraday stock price reactions, modelled empirically using tick data from approximately 100 large-cap

[4] More extensive information on Ravenpack news analytics tool may be found at www.ravenpack.com.
[5] The processed data are available to paying market participants almost instantaneously; non-paying academic researchers can access this information only later, usually after several months.

stocks. A value greater than 50 suggests a positive impact on intraday stock prices, while one below 50 suggests a negative impact. I(.) is the indicator defining the news type; it takes a value of 1 if the condition in parentheses is satisfied. The multi-classifier score (MCQ) quantifies sentiment in news articles related to specific companies. It takes values of 0, 50 or 100 for negative, neutral or positive sentiment, respectively. In contrast to CSS, which reflects news impact on stock prices, MCQ directly assesses the tone of news stories.[6]

1.3.2 Attention measures

Google is currently the most widely used internet search engine. Researchers have used the Google Trends SVI for financial market research (see Moat et al., 2013; Preis et al., 2013), and Google searches have been employed as a proxy for retail investor attention (see Andrei & Hasler, 2014; Da et al., 2011). This study leveraged SVI, focusing on specific stock ticker keywords, to quantify retail investor attention. SVI provides the frequency of a given keyword search compared with all other keyword searches in the same time frame. It is a normalised measure that ranges from 0 to 100. As a relative measure, SVI captures both cross-sectional and time-series variations to allow analysis of keyword search patterns. For instance, the SVI of a particular keyword does not increase when the search frequency during a given period is less than that of other keywords. Further, the SVI for a particular month may vary based on the time horizon used to obtain it. In this study, all SVI data for ticker keywords were downloaded in the same month to ensure consistency. The transformation technique used by Google eliminates time trend effects, which makes absolute values harder to analyse and hinders the analysis of investor attention variations. Therefore, this study used abnormal Google search

[6] If CSS is replaced by MCQ to assess the effects of positive or negative news, we obtain similar findings.

volume (AGSV) as a proxy for abnormal attention. AGSV was measured as the aggregate SVI in a month minus the median SVI in the previous three months. In this way, AGSV captured retail investors' unexpected attention in a given month. Tickers with generic meanings, such as 'JAVA', 'LION' or 'BEAR', were excluded to avoid misinterpretation, as these could reflect non-stock-related attention. Using ticker symbols instead of firm names provides a more focused evaluation of investors' interest in financial information. A firm name search could address various non-financial purposes, whereas a ticker search on Google is more indicative of financial interest in a particular stock. Finally, Google Search allows for filtering of searches submitted within the US, which was employed in this study to reflect the local retail investors' search interest.

1.3.3 Descriptive statistics

Table 1.1 provides an overview of the key variables of our sample dataset. The AGSV variable had a mean of 1.105 and a standard deviation of 16.056. Its skewness value of 0.961 indicates a right-skewed distribution. The News_Positive and News_Negative variables have relatively high means of 127.219 and 101.455, respectively, with substantial variability, as evidenced by standard deviations of 888.303 and 868.951, respectively. The News_Sentiment variable, with a mean of 0.146 and a median of 0.318, suggests that the public news sentiment is slightly positive on average. The considerable variation observed in the AGSV and news measures indicates the potential for detailed analysis.

Table 1.1: Summary statistics

Variable	Obs	Mean	SD	Median	Min	Max	Skew
AGSV	190,150	1.105	16.056	0.000	−99.000	100.000	0.961
News_Positive	183,310	127.219	888.303	24.920	0.000	5,951.840	24.837
News_Negative	183,310	101.455	868.951	10.290	0.000	13,462.900	36.788
News_Sentiment	187,935	0.146	0.318	0.053	−31.587	4.428	−12.76
News_Coverage	183,359	99.699	644.097	23.07	0	39,749.58	22.71
ME	197,729	3.241	17.344	0.269	0.000	794.426	18.327
Vol	197,729	168.588	1,101.333	17.236	0.000	111,839.300	42.662
B/M	187,935	0.146	0.318	0.053	−31.587	4.428	−12.76
Analysts	134,778	1.500	0.847	1.609	0.000	3.807	−0.087
IOR	194,320	0.460	0.316	0.442	0.000	7.561	0.512
Disp	98,909	0.012	0.078	0.002	0.000	8.174	40.524
SIR	188,985	0.034	0.050	0.017	0.000	0.767	3.312

Note. The table reports the summary statistics of key variables during the sample period. AGSV is non-orthogonalised abnormal attention. Vol is expressed in thousands of dollars, and market equity (ME) is expressed in millions of dollars.

1.4 Empirical results & discussion

1.4.1 Relationship between AGSV & other variables of interest

1.4.1.1 Correlations

The study began by examining the correlation between AGSV and other variables of interest. Correlations were measured in the time series for each stock and then averaged across all stocks in the sample, following Engelberg and Parson's (2011) methodology. The results, presented in Table 1.2, reveal that AGSV has a weak positive correlation with negative and positive news and news coverage. Interestingly, while positive news showed a positive correlation with stock returns, negative news showed a negative correlation, suggesting that markets react differently to these two types of news. Trading volume and analyst coverage, often used as proxies for investor attention, had different relationships with AGSV. The correlation matrix indicates a positive correlation between trading volume and AGSV, but no strong correlation was observed between analyst coverage and AGSV. Overall, the relatively weak correlations between AGSV and the other variables imply that the reasons why individuals search for specific stock tickers may extend beyond stock performance and public news events to encompass a range of other factors.

Table 1.2: Correlations between AGSV and other stock-related factors

Variable	AGSV	Return	News_Positive	News_Negative	Analysts	Forecast Disp.	News_Coverage	Vol
AGSV	1	0.011	0.073	0.059	0	0.008	0.085	0.107
Return	0.011	1	0.041	−0.124	−0.02	−0.012	−0.02	−0.028
News_Positive	0.073	0.041	1	0.349	0.093	0.02	0.765	0.239
News_Negative	0.059	−0.124	0.349	1	0.078	0.084	0.578	0.246
Analysts	0.000	−0.02	0.093	0.078	1	0.049	0.106	0.119
Forecast Disp	0.008	−0.012	0.02	0.084	0.049	1	0.042	0.057
News_Coverage	0.085	−0.02	0.765	0.578	0.106	0.042	1	0.31
Vol	0.107	−0.028	0.239	0.246	0.119	0.057	0.31	1

Note. For each stock, we first measured the correlations in the time series at a monthly frequency, then took the average of all stocks in the sample following the methodology of Engelberg and Parsons (2011). Stocks were included in the sample if there were at least 12 trading months.

1.4.1.2 Determinants of abnormal attention

In the literature, media coverage in newspapers such as *The New York Times*, the *Wall Street Journal* and the *Washington Post* has been commonly used as a measure of investor attention. However, considering that retail investors are unlikely to subscribe to all these newspapers, this approach may not capture the actual demand for investor information. In today's digital age, investors can flexibly and conveniently obtain news from various sources via internet searches. To investigate the determinants of retail investor attention, this study examined the monthly lead–lag relationship between AGSV and other relevant variables, including negative and positive news arrivals, using a panel vector autoregression model. The model incorporates five standardised variables: AGSV, trading volume (Vol), stock return, positive news (News_Positive) and negative news (News_Negative) and treats positive and negative news as exogenous variables.

The results, shown in Table 1.3, indicate that stock returns contribute to AGSV, with stocks that perform well attracting more retail investor attention. Interestingly, the coefficient for lagged negative news with AGSV was significantly positive, suggesting that retail investors allocate more attention to stocks with negative news. This result implies that investors are more attentive to negative than positive news. However, the influence of positive news on AGSV was not significant. Additionally, the analysis showed that negative news lagged by one month correlates negatively with current-month stock returns, whereas positive news lagged by one month has a positive association with current-month stock returns. These findings emphasise that both the quantity and the tone of news play a role in influencing stock prices, reflecting the different impacts of positive and negative news on market dynamics.

Table 1.3: Panel VAR model

Variable	(1) AGSV	(2) Vol	(3) Return
L.AGSV	0.0604***	−0.0051***	0.0050**
	(0.0047)	(0.0011)	(0.0026)
L.Vol	−0.0045	0.8346***	0.0164**
	(0.0044)	(0.0615)	(0.0073)
L.Return	0.0306***	−0.0093***	−0.0060
	(0.0032)	(0.0031)	(0.0058)
L.News_Positive	−0.0099	0.0048	0.1072***
	(0.0074)	(0.0552)	(0.0247)
L.News_Negative	0.0372***	0.1737***	−0.0794***
	(0.0050)	(0.0577)	(0.0219)
Observations	170,554	170,554	170,554

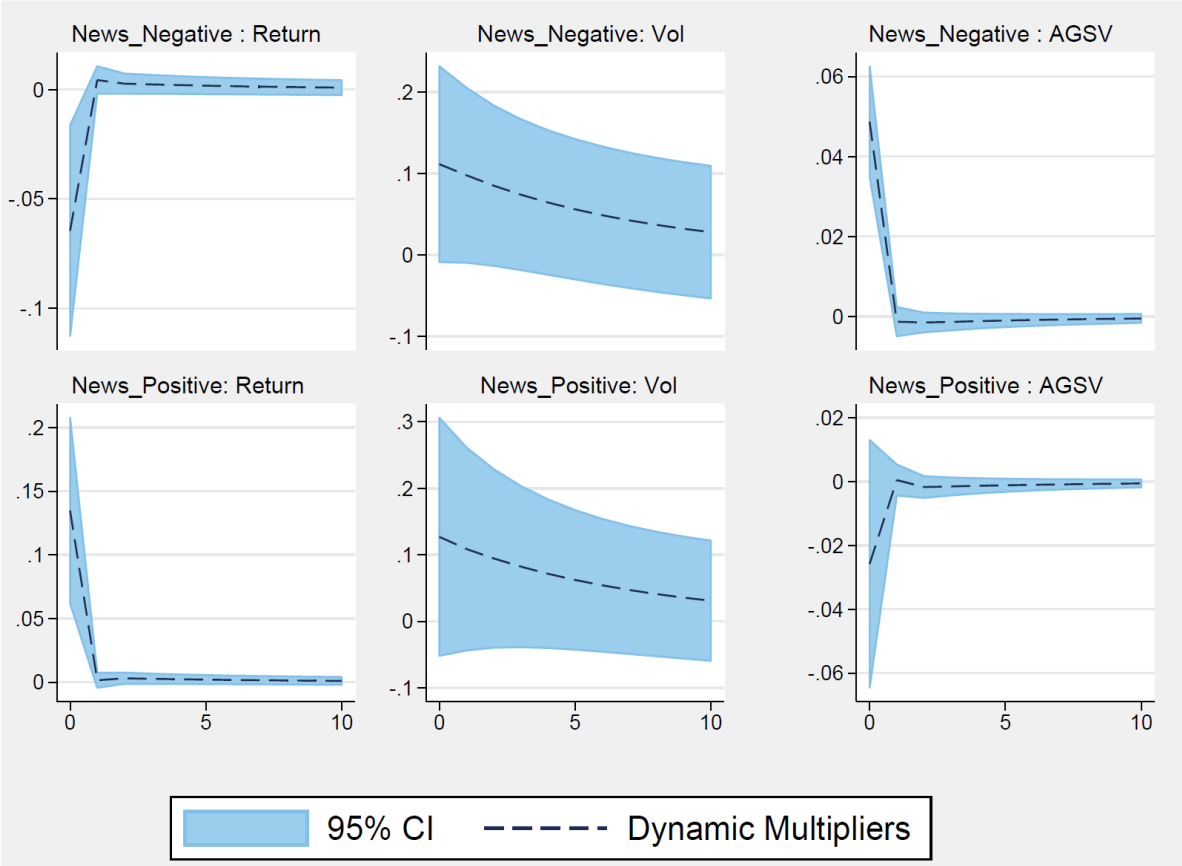
Note. We used a panel vector autoregression (VAR) model using GMM estimation as proposed by Holtz-Eakin et al. (1988), specifying positive and negative news as exogenous variables. Panel-specific fixed effects were removed using the Helmert transformation. Variables prefixed with 'L.' indicate the use of lag terms as instrumental variables. For example, 'L.News_Negative' signifies the lag term of the 'News_Negative'. All variables were standardised to have a mean of 0 and a standard deviation of 1. Standard errors are given in parentheses. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

1.4.1.3 Panel impulse-response functions

To further examine the response of investor attention to news shocks, the study used impulse-response functions with confidence intervals estimated using Monte Carlo simulations. The results, shown in Figure 1.1, demonstrate the responses of several factors to exogenous news shocks. Stock returns responded negatively to negative news and positively to positive news. This aligns with expectations, as news conveys information about a firm's fundamentals, influencing investors' projections of future cash flows and, consequently, stock prices. These responses typically lasted for approximately one month. Trading volume reacted positively to negative news, suggesting that negative news spurs trading activity and increases market liquidity. This reaction emphasises the heightened sensitivity of investors to negative over positive news. The trading volume response was persistent, tapering off over one to 10 months.

For abnormal investor attention, there was a positive response to negative news, while the response to positive news was negative in the following month. The impact of negative news is found to be twice as large as positive news, indicating a tendency for investors to allocate more abnormal attention to stocks affected by negative news. These findings emphasise the asymmetric effects of news shocks on retail investors' attention, with negative news having a more substantial influence than positive news.

Figure 1.1: Panel impulse-response functions



Note. Figure 1.1 presents the dynamic panel impulse-response functions. Positive and negative news shocks were treated as exogenous variables. All variables were standardised. The horizontal axis represents the number of months.

This study examined the relationship between AGSV and other variables to explore factors influencing investor attention. Table 1.4 presents panel regression results, with the

dependent variable being AGSV in month *t*. The regression model incorporated the main firm characteristics linked in the literature to investor attention, including the standardised trading volume, firm size, analyst forecast following and stock returns. In Column 1, the results indicate a significant positive correlation between the trading volume and retail investors' abnormal attention. Similarly, stock return was significantly positively associated with abnormal attention. For Column 2, which includes total news coverage and news sentiment, the results show that News_Coverage is significantly positively correlated with abnormal attention, suggesting that firms with more news coverage attract more retail investor attention. Column 3 incorporates both positive and negative news. Here, News_Negative is significantly positively associated with abnormal investor attention, while News_Positive has a significant negative association. This implies that public investors are more attentive to stocks with negative news. Interestingly, the coefficient of News_Sentiment is statistically non-significant, indicating that news sentiment does not have a substantial impact on abnormal investor attention. In Column 4, the regression model includes idiosyncratic volatility, analyst dispersion, institutional ownership ratio and short interest ratio (SIR). The incorporation of these variables resulted in the coefficient of market capitalisation becoming positive, in line with the expectation that larger firms attract more attention. This result suggests that lower institutional investor ownership, analyst coverage and SIR, along with higher idiosyncratic volatility, are associated with increased retail investor attention. These factors, as indicators of information integration efficiency, imply that firms with more information uncertainty are more likely to draw retail investors' attention.

Table 1.4: Abnormal attention and related variables

	Dependent Variable: AGSV			
Variable	(1)	(2)	(3)	(4)
News_Positive			−0.0002***	−0.0003***
			(0.0001)	(0.0001)
News_Negative			0.0004***	0.0004***
			(0.0001)	(0.0001)
News_Sentiment		0.002	0.002	−0.001
		(0.005)	(0.005)	(0.007)
News_Coverage		0.0002*		
		(0.0001)		
Vol	1.317***	1.313***	1.301***	0.00004
	(0.068)	(0.068)	(0.068)	(0.00003)
LnME	−0.241***	−0.261***	−0.258***	0.194***
	(0.026)	(0.028)	(0.028)	(0.044)
Analysts	0.032	0.030	0.031	−0.205*
	(0.062)	(0.062)	(0.062)	(0.108)
Return	3.137***	3.150***	3.203***	1.607***
	(0.469)	(0.470)	(0.470)	(0.558)
Idio				85.866***
				(6.222)
Disp.				0.000
				(0.000)
IOR				−0.703***
				(0.180)
SIR				−3.792***
				(1.004)
Observations	124,889	124,889	124,889	92,900
Month fixed effect	YES	YES	YES	YES
R^2	0.0154	0.0155	0.0156	0.0193
Adjusted R^2	0.014	0.014	0.0142	0.0174

Note. The table reports the panel data regression using abnormal attention (AGSV) as the dependent variable. Robust standard errors clustered by individual firms are given in parentheses. The sample period is from 2004 to 2018. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

1.4.2 AGSV & underreaction to public news

The study aimed to examine the relationship between AGSV and reactions to news through hypothesis testing, specifically by estimating the interaction coefficients (b4 and b5) in Equation 5. The control variables include LnME, LnBM, Vol and Analysts, representing company size, book-to-market ratio, trading volume and analyst coverage. $\text{BHAR}_{i,t,N}$ represents the buy-and-hold abnormal return (BHAR), with N denoting the holding periods.

$$\text{BHAR}_{i,t,N} = a_0 + b_1 {AGSV_{i,t}}^{+} + b_2 News\_Positive_{i,t} + b_3 News\_Negative_{i,t}$$

$$+ b_4 {AGSV_{i,t}}^{+} \times News\_Positive_{i,t} + b_5 {AGSV_{i,t}}^{+} \times News\_Negative_{i,t} + b_6 Z_{i,t} \quad (5)$$

Table 1.5 presents the results of Fama–MacBeth cross-sectional regression. We considered that abnormal attention, measured as AGSV, could be influenced by other firm characteristics, including trading volume, firm size and news coverage, previously proposed as key explanatory factors of investor attention. To isolate the effect of AGSV from these factors, denoted as X_j, this study extracted the orthogonalised portion of AGSV that is independent of X_j. Specifically, it estimated cross-sectional regressions of AGSV against X_j to obtain the residual abnormal attention, $AGSV^+$. $AGSV^+$ provides a more accurate measure of abnormal attention, because it controls for the influence of other identified factors on AGSV.

$$AGSV_{i,t} = \beta X_{i,t,j} + \delta_{it} \quad (6)$$

$$AGSV_{i,t}{}^{+} = AGSV_{i,t} - \widehat{\beta} X_{i,t,j} \quad (7)$$

The dependent variable is the cumulative DGTW BHAR from month *t + 1* to month *t + 3*, with month *t* as the month of news arrival. To obtain the monthly BHAR, we calculated the buy-and-hold return for each stock and then subtracted the buy-and-hold return over the same period from the matched risk-adjusted portfolio.[7] All independent variables were standardised to a mean of 0 and a standard deviation of 1. The results showed that the coefficient of $AGSV^+$ was not significant, suggesting that abnormal attention cannot predict the three-month BHAR ($BHAR_{i,t,3}$). The findings in Table 1.5 demonstrate a negative news drift effect, as indicated by the significant negative coefficient of News_Negative in Column 2. This negative impact persisted in BHAR over a three-month holding period. Further, Column 3 in Table 1.5 reveals an interaction between negative news and abnormal attention, which suggests that retail investor attention impedes the incorporation of negative news into prices, reinforcing

[7] We calculated the characteristic-based benchmarks following Daniel et al. (1997).

the underreaction to negative news. Similarly, positive news positively affected BHAR over the subsequent three months; the interaction between positive news and abnormal attention had a significantly positive coefficient, suggesting that greater abnormal retail attention strengthens the positive news drift effect. In summary, the results imply that higher abnormal attention leads to a stronger underreaction to news, resulting in more pronounced drift effects for both positive and negative news. When total news arrival is used as a proxy for total news coverage in Column 4, the coefficient of News_Coverage is positive. However, the interaction term between News_Coverage and abnormal attention is not significant,[8] possibly because of the counteracting effect of abnormal attention on both positive and negative news.

[8] We repeat all regressions using the unorthogonalised abnormal attention AGSV and obtained similar findings.

Table 1.5: Fama–MacBeth regression of three-month BHAR

	Dependent variable: DGTW three−month BHAR after news arrival ($BHAR_{i,t,3}$)				
Variable	(1)	(2)	(3)	(4)	(5)
$AGSV^{+}$	-0.00004	-0.00004	-0.00005	-0.0001	0.00000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
News_Positive		0.007*	0.010^{**}		
		(0.004)	(0.004)		
News_Negative		-0.008**	-0.010^{***}		
		(0.004)	(0.004)		
News_Coverage				0.001^{**}	
				(0.001)	
News_Sentiment					0.0004^{***}
					(0.0001)
$AGSV^{+}$×News_Positive			0.001^{**}		
			(0.0005)		
$AGSV^{+}$×News_Negative			-0.001^{**}		
			(0.001)		
$AGSV^{+}$×News_Coverage				−0.0002	
				(0.0001)	
$AGSV^{+}$×News_Sentiment					−0.00001
					(0.0001)
LnME	0.002^{**}	0.004^{***}	0.004^{***}	0.004^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
LnBE	0.005^{**}	0.005^{**}	0.005^{**}	0.005^{**}	0.005^{**}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
GrProf	0.021^{***}	0.024^{***}	0.024^{***}	0.024^{***}	0.024^{***}
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Vol	−0.002	−0.002	−0.002	−0.003	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Analysts		-0.007^{***}	-0.007^{***}	-0.007^{***}	-0.007^{***}
		(0.002)	(0.002)	(0.002)	(0.002)
Constant	-0.008^{***}	-0.008^{***}	-0.008^{***}	-0.008^{***}	-0.010^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	113,147	113,147	113,147	113,147	113,147
R^2	0.035	0.042	0.043	0.041	0.044

Note. The table reports the result from Fama–MacBeth regression. The dependent variable is the cumulative DGTW BHAR from month *t + 1* to month *t + 3*, with month *t* as the month of news arrival. The independent variables were standardised to a mean of 0 and a standard deviation of 1. AGSV was orthogonalised to trading volume, total news coverage and the natural logarithm of market size (LnME) to obtain the residuals from the cross-sectional regressions $AGSV^{+}$. To control for other variables, the study included the natural logarithm of the book-to-market ratio (LnBM), Vol and analyst coverage (Analysts) in the regression.

***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Newey-West standard errors with three lags are employed

We next conducted a regression analysis across various time horizons, with the current month return (Ret), next month return (Ret_{t+1}) and $BHAR_{i,t,N}$ as the dependent variables. $BHAR_{i,t,N}$ represents the BHAR from month $t + 1$ to $t + N$, excluding the current month t. The regression of Column 3 of Table 1.5 was repeated for each time horizon. Table 1.6 presents the Fama–MacBeth coefficients of the key variables for each time horizon. These coefficients were obtained by monthly regression of BHAR against firm characteristics. Coefficients averaged across months were used, and Newey-West standard errors with three lags are employed. Our results show that both positive and negative public news exert an effect over a six-month horizon, as evidenced by the significant coefficient of news arrival. The coefficient of news arrival suggests that the arrival of negative news negatively affects the next month's return and $BHAR_{i,t,N}$ over the following three to six months. Similarly, positive news positively drives the next month's stock return, and the effect continues in the following periods. Further, the regression analysis for the next month's stock returns (see Panel A) indicates that the interaction between $AGSV^{+}$ and the news effect is positively associated with positive news and negatively associated with negative news. Hence, higher abnormal attention amplifies the impact of both types of news, suggesting that investors' reactions to a firm's public news are stronger when they allocate more unexpected attention to the firm. For $BHAR_{i,t,N}$ in the two to six months after news arrival, we found that the interaction of $AGSV^{+}$ with negative news was negative and the interaction of $AGSV^{+}$ with positive news was positive, implying that abnormal retail investor attention exacerbates investors' underreaction to both types of news. Hence, a higher level of abnormal retail investor attention does not improve market efficiency or accelerate information absorption into prices.

Table 1.6: Fama–MacBeth coefficients over different time horizons

Variable	Result	$AGSV^+$	News_Positive	News_Negative	$AGSV^+$ ×News_Positive	$AGSV^+$ ×News_Negative
Ret_t	Coeff	2.00e−04	0.0403	−0.0474	0.0034	−0.0041
	p	0.00	0.00	0.00	0.00	0.00
Ret_{t+1}	Coeff	-0.0000	0.0068	-0.007	0.0001	-0.0008
	p	0.0548	0.00	.00	0.1604	.00
$\mathrm{BHAR}_{i,t,2}$	Coeff	0e+00	0.0055	-0.0067	0.0005	-0.0009
	p	0.1747	0.00	0.00	0.00	0.00
$\mathrm{BHAR}_{i,t,3}$	Coeff	0e+00	0.0098	-0.0104	0.001	-0.0013
	p	0.2712	0.00	0.00	0.00	0.00
$\mathrm{BHAR}_{i,t,4}$	Coeff	0e+00	0.0119	-0.0126	0.0013	-0.0018
	p	0.5975	0.00	0.00	0.00	0.00
$\mathrm{BHAR}_{i,t,5}$	Coeff	0e+00	0.0116	-0.0134	0.0013	-0.0017
	p	0.441	0.00	0.00	0.00	0.00
$\mathrm{BHAR}_{i,t,6}$	Coeff	-1.00E-04	0.0129	-0.0156	0.0014	-0.0017
	p	0.394	0.00	0.00	0.00	0.00

Note. The table reports Fama–MacBeth coefficients on the key variables over the different time horizons based on the same regression as in Column 3 of Table 1.5. Ret represents a regression using the current return as the dependent variable. The dependent variable $\mathrm{BHAR}_{i,t,N}$ is the BHAR from time *t + 1* to *t + N*, excluding the current month *t*. To obtain the $\mathrm{BHAR}_{i,t,N}$ each month, we calculated the buy-and-hold return of each stock, then subtracted the buy-and-hold return of the matched risk-adjusted portfolio over the same horizon (see Daniel et al., 1997). The benchmark returns were based on portfolios matched on size, book-to-market value and momentum. Newey-West standard errors with three lags are employed.

1.4.3 Portfolio analysis

This study aimed to explore the news-dependent outcomes of a long–short investment strategy, using risk-adjusted buy-and-hold average abnormal returns (BHAAR). To achieve this, we constructed two types of news strategies to analyse the market reaction to positive and negative news. For each month *t*, we grouped stocks into three portfolios based on their AGSV^{+} rankings. For a given month, the AGSV^{+} break points were the thirtieth and seventieth percentiles. Within each attention group, we sorted the stocks into 11 portfolios based on the quantity of negative or positive news. Portfolio 0 (P0) represents a portfolio with no positive or negative news in month *t*. Firms with positive or negative news coverage in a given month are further divided Into ten groups. Then, we created 11 portfolios for each news strategy. Decile 10 represented the portfolio with the highest news arrival. Within each AGSV^{+} group, we held each of the 11 news portfolios for months $t + 1$ to $t + N$. To obtain BHAR values, we calculated the buy-and-hold return for each stock in the portfolio in each month, excluding the formation month. This return was then subtracted from the buy-and-hold return of the characteristic-based benchmark stocks (Daniel et al., 1997) over the same period. We averaged these abnormal returns across the stocks in each portfolio to obtain $\mathrm{BHAARs}_{t,N}$. The negative news strategy involved creating a long–short spread by going long on $\mathrm{BHAARs}_{t,N}$ with no negative news (P0) and short on $\mathrm{BHAARs}_{t,N}$ with the highest negative news (Decile 10). Similarly, the positive news strategy involved going long on $\mathrm{BHAARs}_{t,N}$ in with highest positive news (Decile 10) and short on $\mathrm{BHAARs}_{t,N}$with no positive news (P0). Table 1.7 reports the time-series average of the monthly long–short profit of the average current return and $\mathrm{BHAARs}_{t,N}$ for each abnormal attention group over the sample period. Group 1 of AGSV^{+} represents the highest attention

group, and Group 3 represents the lowest attention group. Group 1–3 represents the long–short difference between the high- and low-attention groups. Panel A of Table 1.7 shows that the negative news strategy can generate a persistent positive return over 2 to 6 months for the high-attention group. Specifically, the high-attention group had a greater profit than the low-attention group. The long–short profits associated with BHAARs peaked for the three-month holding period, with a significant *t*-statistic of 2.66 for the high-attention group, yielding profits of up to 1.86%. Profitability remained significant for the three- and four-month holding periods for the high-attention group. Over a longer period, BHAR became negative, although the *t*-statistic was not significant. This result suggests that prices may eventually move towards a reasonable level; there is no clear evidence for overreaction, as indicated by the price reversals. In the half-year holding period, the long–short profit of the negative news strategy increased with the attention level, indicating that the underreaction to news is strongest for stocks receiving the highest level of abnormal attention. This finding suggests that market mispricing following the arrival of negative news is asymmetric across different levels of retail investors' abnormal attention ($AGSV^{+}$). We observed a significant difference between the high- and low-attention groups for $BHAR_{t,3}$, with a significant *t*-statistic of 2.79 and profits of 2.53%. In enhancing the initial analysis, this study incorporates proportion statistics to illustrate market reactions. Specifically, the proportion of returns post-news release, calculated as 1.863 out of a total of 4.815 (combining a $BHAR_{t,3}$ of 1.863 and the current month return of 2.952), indicates that approximately 38.7% of the total stock performance over this three-month period followed the release of significant negative news by high attention groups. This underscores a notable market underreaction to negative news, as a substantial portion of the returns accrued post-release.

Such behavior aligns with observations by Da, Engelberg, and Gao (2011), who highlighted market delays in responding to news, often attributable to investor hesitancy or the gradual dissemination of information.

Unlike the negative news strategy, the positive news strategy did not show a linear relationship between the attention level and the long–short profit. The positive news strategy did not yield a positive return in month *t*; stocks with a higher volume of good news coverage did not outperform those with no positive news coverage. It appears that the underreaction to positive news is less sensitive to retail attention levels than the negative news reaction, as evidenced by a narrower spread of BHAAR between high- and low-$AGSV^{+}$ groups for the positive news strategy.

Because the hedge portfolio for both news strategies produced a significant long–short profit for the three-month holding period, we selected the three-month holding period for further analysis.

Table 1.7: Time-series means of long–short strategy for three-month BHAARs

News_Negative AGSV+	Ret	$BHAR_{t,2}$	$BHAR_{t,3}$	$BHAR_{t,4}$	$BHAR_{t,5}$	$BHAR_{t,6}$	$BHAR_{t,12}$
Group 1	2.952	0.515	1.863	1.683	1.316	0.941	−0.773
	(6.3)	(0.937)	(2.66)	(1.999)	(1.284)	(0.804)	(−0.493)
Group 2	1.74	0.795	1.271	1.349	1.239	1.061	0.45
	(5.534)	(1.736)	(2.122)	(1.958)	(1.65)	(1.223)	(0.362)
Group 3	1.343	−0.273	−0.667	−0.091	0.43	0.127	0.96
	(4.423)	(−0.535)	(−1.153)	(−0.133)	(0.526)	(0.13)	(0.688)
Group 1–3	1.609	0.788	2.53	1.774	0.887	0.815	−1.733
	(2.882)	(1.051)	(2.786)	(1.636)	(0.677)	(0.535)	(−0.826)
News_Positive AGSV+	Ret	$BHAR_{t,2}$	$BHAR_{t,3}$	$BHAR_{t,4}$	$BHAR_{t,5}$	$BHAR_{t,6}$	$BHAR_{t,12}$
Group 1	−0.259	0.311	1.449	1.674	1.292	1.339	−0.898
	(−0.884)	(0.745)	(2.606)	(2.144)	(1.683)	(1.674)	(−0.885)
Group 2	0.147	0.624	0.871	0.083	0.143	0.194	1.977
	(0.599)	(1.737)	(1.804)	(0.135)	(0.219)	(0.291)	(1.927)
Group 3	−0.613	1.19	1.344	0.915	1.591	1.916	1.387
	(−2.684)	(2.989)	(2.78)	(1.691)	(2.494)	(2.561)	(1.217)
Group 1–3	0.354	−0.879	0.105	0.759	−0.299	−0.577	−2.284
	(0.952)	(−1.522)	(0.142)	(0.799)	(−0.299)	(−0.527)	(−1.498)

Note. For each month t, we grouped stocks into three portfolios based on the AGSV+ ranking. The AGSV+ break points were the thirtieth and seventieth percentiles in month t. Group 1 represents the highest abnormal retail attention group, and Group 3 represents the lowest abnormal retail attention group. We further sorted the stocks into 11 portfolios based on the quantity of negative or positive news. For each AGSV+ decile and the 11 news portfolios, we held each portfolio for months t + 1 to t + N. To obtain the buy-and-hold abnormal return, we calculated the buy-and-hold return for each stock in the portfolio during each month, excluding month t. This return was then subtracted from the buy-and-hold return of the characteristic-based benchmarks (Daniel et al., 1997) over the corresponding period. We further averaged these abnormal returns across the stocks within each portfolio to obtain the buy-and-hold average abnormal returns (BHAARs). The negative news strategy was to obtain the long–short spread that is long on the BHAARs for zero negative news (P0) and short on the BHAARs in the highest decile (Decile 10) of negative news for each month. The positive news strategy was to be long on the BHAARs in Decile 10 of positive news and short on the BHAARs in zero positive news (P0) for each month. The table reports the time-series average return of the long–short BHAARs over two to 12 months following month t. Standard deviations are given in parentheses. Ret represents the time-series average of the current period risk-adjusted return (DGTW 1997).

Table 1.8 displays average abnormal returns over a three-month holding period from a long–short zero-investment strategy. The returns are categorised by $AGSV^{+}$ and were adjusted using the Fama–French three-factor model and the Fama–French model with an additional momentum factor. To ensure adequate stock diversification, the portfolios were formed from two subsamples of stocks: high abnormal attention stocks (with abnormal attention above the seventieth percentile in a given month) and low abnormal attention stocks (with abnormal attention below the thirtieth percentile). Within each attention group, the stocks were divided into 11 portfolios based on the volume of negative or positive news. The negative news strategy involved creating a hedge portfolio that went long on the portfolio with no negative news (P0) and short on the portfolio with the highest number of negative news items (Decile 10). The positive news strategy involved going long on the portfolio with the highest number of positive news items (Decile 10) and short on the portfolio with no positive news (P0). Additionally, a 'negative–positive' news strategy was developed to measure the relative influx of negative news in month *t*. This was achieved by calculating the difference between the numbers of negative and positive news items. We sorted stocks into ten portfolios based on the relative volume of negative news and created a hedge portfolio that was long on the lowest group (Decile 1) and short on the highest group (Decile 10). The evaluation of each news strategy involved a three-month holding period from month $t + 1$ to $t + 3$. The portfolio comprised three cohorts, each corresponding to a different month of formation. Each month, a cohort representing the previous three months was removed from the portfolio, and a new cohort was added, reflecting the most recent month. The return for each news strategy was calculated as the average return across the three cohorts. To obtain the average abnormal return, we

performed regressions of the average returns for the three cohorts using the Fama–French three-factor model, the four-factor model with momentum, and the Fama–French five-factor model incorporating profitability and investment factors.

Panel A of Table 1.8 shows that the long–short strategy for negative news produced significant alpha with high-attention ($AGSV^{+}$) stocks, suggesting that abnormal returns are dominant in this group. In contrast, Panel B reveals that the long–short strategy for positive news did not yield significant alpha for either of the attention ($AGSV^{+}$) groups. This suggests that the underreaction to news is more pronounced for negative news, especially for high-attention ($AGSV^{+}$) stocks. Panel C demonstrates that the negative-positive news strategy resulted in significant abnormal returns for both the high- and low-attention ($AGSV^{+}$) groups. Notably, the high-attention portfolio achieved a higher alpha, up to 0.93%, when analysed using the three-factor model. After we isolated the net negative news impact from the positive news impact, the mispricing became more pronounced. This emphasises the particularly significant underreaction of investors to net negative news shocks.

Table 1.8: Abnormal returns of long–short profits based on negative and positive news strategies

Panel A: Negative news

Fama–French factors				Fama–French and momentum factors			Fama–French five factors		
Decile 0–10	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High attention	**0.557**	2.385	.018	**0.533**	2.287	0.023	**0.606**	3.024	0.003
Low attention	−0.096	−0.469	.64	−0.132	−0.656	0.513	0.072	0.380	0.704
High−low	**0.652**	2.743	.007	**0.665**	2.785	0.006	**0.534**	2.945	0.004

Panel B: Positive news

Fama–French factors				Fama–French and momentum factors			Fama–French five factors		
Decile 0–10	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High attention	−0.235	−0.821	0.413	−0.192	−0.679	0.498	-0.261	-1.073	0.285
Low attention	0.15	0.621	0.535	0.18	0.75	0.454	-0.048	-0.218	0.828
High−low	−0.384	−1.279	0.203	−0.372	−1.233	0.219	-0.213	-0.850	0.396

Panel C: Negative–positive news

Fama–French factors				Fama–French and momentum factors			Fama–French five factors		
Decile 0–10	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High attention	**0.928**	3.262	0.001	**0.88**	3.138	0.002	**0.898**	3.862	0.000
Low attention	**0.573**	2.599	0.01	**0.556**	2.52	0.013	**0.689**	3.420	0.001
High−low	0.355	1.376	0.171	0.324	1.26	0.209	0.209	1.044	0.298

Note. Table 1.8 shows the average abnormal return over three months following the initiation of a long–short zero-investment news strategy sorted by $AGSV^{+}$ and using the Fama–French three-factor model and the Fama–French model with momentum. The return for each news strategy portfolio was the average return of these three cohorts. The table reports the estimates of α from the regression for Equations:

$$R_{i,t} = a + bMKT_t + cSMB_t + dHML_t + u_t \quad (8)$$

$$R_{i,t} = a + bMKT_t + cSMB_t + dHML_t + eUMD_t + u_t \quad (9)$$

$$R_{i,t} = a + bMKT_t + cSMB_t + dHML_t + eCMA_t + RMW_t + u_t \quad (10)$$

where $R_{i,t}$ is the average return of the long–short spread over the three months after the initiation of the long–short zero-investment strategy, sorted by news type.

1.5 Additional analysis & robustness checks

1.5.1 Controlling for firm characteristics

To address the heterogeneity of news impact, this study considered other relevant characteristics using regression controls and sorting. The goal was to identify the subset of stocks for which the news effect is most pronounced and determine whether the abnormal attention effect persists after accounting for relevant firm characteristics. Following Hong et al. (2010), we controlled for characteristics that are known indicators of market efficiency and have been used in the literature as proxies for firm-specific information diffusion, including firm size, institutional ownership, analyst forecast dispersion, idiosyncratic volatility and short interest.

1.5.1.1 Analyst dispersion

Zhang's (2006) research has demonstrated that news underreaction is especially significant in situations characterised by high levels of information uncertainty. In this context, analyst forecast dispersion has been used as a proxy for the divergence in investors' opinions. Miller (1977) found that stocks with high analyst dispersion can serve as proxies for the heterogeneity of investors' opinions. If abnormal attention is associated with the arrival of uninformed investors, then, after analyst dispersion is controlled for, the $AGSV^{+}$ effect may be less significant for news strategy anomalies. To examine this possibility, we used a measure of analyst forecast dispersion proposed by Garfinkel (2009) to construct an analyst-based proxy for investors' divergence of opinion. Table 1.9 gives the average abnormal return over three months following the initiation of the long–short zero-investment strategy, sorted by firm characteristics and orthogonalised abnormal attention ($AGSV^{+}$) using the Fama–French three-

factor model with momentum. For each news strategy, we divided stocks into high, medium and low groups sorted by firm characteristics. For each portfolio, we further sorted stocks into high- and low-AGSV groups. Within each of the six portfolios, we divided the stocks into 11 portfolios sorted by the volume of negative or positive news. The negative news strategy was to form a hedge portfolio that is long on the portfolio with zero negative news (P0) and short on the portfolio in the highest decile (Decile 10) of negative news for each month. The positive news strategy was to be long on the portfolio in Decile 10 of positive news and short on the portfolio with zero positive news (P0). We additionally created a negative–positive news strategy that took the difference between the volume of negative and positive news in month *t* to capture the relative arrival of negative news, then sorted stocks into ten portfolios based on the relative negative news volume and formed a hedge portfolio that was long on the lowest group (Decile 1) and short on the highest group (Decile 10). For example, in Panel A of Table 1.9, we first sorted stocks into high-, medium- and low-dispersion groups and then examined the news strategy. Within each portfolio,[9] we further sorted stocks into high- and low-$AGSV^{+}$groups. To ensure that each portfolio was well diversified, we used median break points instead of seventieth and thirtieth percentile break points as in Table 1.7. For each of the six portfolios, the same news strategy was implemented, and the portfolio was held from month *t + 1* to *t + 3*. We regressed the time-series average return using the Fama–French three-factor model and the four-factor model with momentum to calculate the average abnormal return.

In Panel A of Table 1.9, the results show that the negative news long–short strategy produced significantly positive alpha only for stocks with high analyst dispersion and high

[9] The result was also robust if we used unconditional sorting for all characteristic groups.

$AGSV^{+}$. On average, firms with high analyst dispersion had a greater alpha compared with those in the medium and low analyst dispersion groups. These findings suggest that the negative news effect for high-$AGSV^{+}$ stocks can be partially subsumed under the analyst dispersion effect, consistent with literature findings that analyst dispersion is associated with heterogeneity in investors' opinion; therefore, it can amplify investors' underreactions to negative news and drive a pricing anomaly. Within the high analyst dispersion group, the high-low $AGSV^{+}$portfolio produced significant alpha, further confirming the asymmetry of negative news mispricing effects across different retail attention levels. The negative news strategy anomaly exhibited statistical significance only for stocks characterised by high levels of analyst forecast dispersion and high abnormal attention allocation. This provides evidence that greater allocation of unexpected attention and increased analyst uncertainty regarding a stock can impede the speed at which public news information is incorporated into the stock price. The finding emphasises the crucial role played by abnormal attention ($AGSV^{+}$) and analyst forecast dispersion in the market information assimilation process. The positive news strategy did not generate significant abnormal returns in any portfolio.

Table 1.9: Characteristic controls

Panel A: Long–short profit spread for the three-factor analysis plus momentum, controlling for analyst forecast dispersion

		News_Negative			News_Positive			Negative–positive		
Dispersion	AGSV⁺	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High	High	0.696	1.718	.08	0.154	0.227	.821	**0.887**	2.396	.018
High	Low	-0.421	-0.966	.336	0.416	0.537	.592	**1.672**	3.598	.00
High	High-low	**1.06**	2.07	.04	-0.91	-0.95	.344	-0.785	-1.444	.15
Medium	High	0.029	0.111	.912	0.118	0.253	.801	0.107	0.408	.684
Medium	Low	-0.292	-0.8	.425	0.711	1.446	.151	0.083	0.29	.772
Medium	High-low	0.313	0.737	.462	-0.43	-0.802	.425	0.024	0.065	.948
Low	High	0.07	0.315	.753	-0.355	-0.663	.508	0.144	0.684	.495
Low	Low	0.152	0.768	.443	0.358	1.045	.298	**0.442**	2.267	.025
Low	High-low	-0.101	-0.403	.687	-0.826	-1.407	.163	-0.298	-1.094	.275

Panel B: Long–short profit spread for the three-factor analysis plus momentum, controlling for institutional ownership

		News_Negative			News_Positive			Negative–positive		
IOR	AGSV⁺	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High	High	0.372	1.387	.167	-0.628	-1.317	.191	**0.581**	1.99	.048
High	Low	0.238	0.818	.415	-0.401	-0.599	.55	**0.919**	3.668	.00
High	High-low	0.157	0.455	.649	-0.336	-0.504	.616	-0.338	-1.088	.278
Medium	High	**0.916**	3.307	.001	-0.549	-1.883	.061	**1.102**	3.966	.00
Medium	Low	-0.004	-0.015	.988	0.408	1.261	.209	0.74	3.075	.002
Medium	High-low	**0.92**	3.096	.002	-0.879	-2.354	.02	0.361	1.111	.268
Low	High	**2.085**	4.983	.00	**-1.395**	-3.359	.001	**1.259**	4.088	.00
Low	Low	**1.201**	2.851	.005	**-1.229**	-3.326	.001	**0.758**	2.224	.027
Low	High-low	0.884	1.502	.135	-0.166	-0.313	.755	0.502	1.084	.28

Panel C: Long–short profit spread for the three-factor analysis plus momentum, controlling for firm size

		News_Negative			News_Positive			Negative–positive		
Size	AGSV⁺	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High	High	0.369	1.261	.209	-0.277	-0.633	.528	0.136	0.632	.528
High	Low	0.24	0.833	.406	0.065	0.141	.888	0.264	1.404	.162
High	High-low	0.245	0.847	.398	-0.478	-1.038	.303	-0.128	-0.6	.549
Medium	High	**1.216**	3.317	.001	**-0.924**	-2.273	.024	0.982	3.928	.00
Medium	Low	0.428	1.153	.25	-0.192	-0.456	.649	0.575	2.652	.009
Medium	High-low	**0.783**	2.142	.034	-0.735	-1.743	.083	0.407	1.389	.167
Low	High	**1.68**	3.006	.003	**-1.257**	-2.325	.021	**1.49**	4.82	.00

Low	Low	**1.236**	2.748	.007	**-1.43**	-3.655	.00	**1.092**	2.971	.003
Low	High-low	0.443	0.751	.454	0.093	0.168	.867	0.398	0.809	.42

Panel D: Long–short profit spread for the three-factor analysis plus momentum, controlling for idiosyncratic volatility

		News_Negative			News_Positive			Negative–positive		
Idio	AGSV⁺	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High	High	**1.149**	2.84	.005	-0.523	-1.168	.244	**1.307**	3.542	.001
High	Low	0.278	0.625	.533	-0.515	-1.101	.272	**1.312**	3.392	.001
High	High-low	0.872	1.817	.071	0.027	0.048	.962	-0.005	-0.01	.992
Medium	High	0.397	1.864	.064	-0.335	-0.774	.44	**0.46**	2.313	.022
Medium	Low	-0.026	-0.11	.913	0.212	0.703	.483	**0.724**	4.249	.00
Medium	High-low	0.423	1.656	.099	-0.544	-1.067	.287	-0.264	-0.994	.321
Low	High	**0.33**	2.089	.038	-0.374	-1.863	.064	0.188	1.611	.109
Low	Low	0.133	0.868	.387	0.055	0.263	.793	0.191	1.662	.098
Low	High-low	0.197	1.444	.151	-0.514	-2.336	.021	-0.003	-0.02	.984

Panel E: Long–short profit spread for the three-factor analysis plus momentum, controlling for short interest

		News_Negative			News_Positive			Negative–positive		
Idio	AGSV⁺	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High	High	**0.987**	2.249	.026	-0.029	-0.053	.957	0.6	1.717	.088
High	Low	-0.117	-0.309	.758	-0.486	-0.858	.392	**0.921**	2.962	.004
High	High-low	**1.103**	2.285	.024	0.108	0.148	.882	-0.321	-0.747	.456
Medium	High	-0.033	-0.172	.864	0.283	0.754	.452	**0.816**	3.64	.00
Medium	Low	-0.234	-1.107	.27	0.463	1.147	.253	0.42	1.739	.084
Medium	High-low	0.201	0.776	.439	-0.124	-0.229	.819	0.396	1.299	.196
Low	High	0.44	1.33	.185	-0.036	-0.093	.926	**0.938**	3.299	.001
Low	Low	**0.524**	2.178	.031	-0.301	-1.181	.239	**1.058**	4.217	.00
Low	High-low	-0.083	-0.215	.83	0.265	0.574	.567	-0.121	-0.328	.743

Note. Table 1.9 gives the average abnormal return over three months following the initiation of the long–short zero-investment strategy, sorted by firm characteristics and orthogonalised abnormal attention (AGSV⁺) using the Fama–French three-factor model with momentum. We regressed the time-series average return using Fama–French four-factor models to obtain the average abnormal return. The table reports the estimates of α from the regression.

where $R_{i,t}$ is the average return of the long–short spread over the three months following the initiation of the long–short zero-investment strategy, sorted by news type.

1.5.1.2 Institutional ownership

We argue that retail attention is an important determinant of individual investors' reactions to public news, but this does not apply equally to institutional investors. Institutional investors tend to limit their focus to firms that meet particular criteria in a specific industry. Therefore, the attention resources available to them are not as scarce as those available to individual investors. Further, institutional investors are less inclined to acquire information about a firm through Google searches. The price reactions of stocks to news may thus vary with the level of institutional ownership.

Panel B of Table 1.9 reveals that both negative and positive news strategies can generate significant alpha in firms with low institutional ownership. In the low institutional ownership group, high-$AGSV^+$ stocks had the most significant alpha for the negative news strategy. However, in the same group, the alpha for high-$AGSV^+$ stocks was significantly negative for the positive news strategy. Stocks with low institutional ownership and a high positive news volume underperformed relative to those with no positive news in the three-month holding period. In the high institutional ownership group, neither the negative or the positive news strategy generated significant profit across different $AGSV^+$ levels. This suggests that firms with higher institutional investor holdings show higher market efficiency and may encounter fewer problems related to limited attention. Combined with the finding in Table 1.4 that institutional ownership is negatively correlated with AGSV, this finding confirms that the effect of retail investor attention is consistent with but not entirely subsumed under the effect of institutional ownership. For the negative–positive news strategy, alpha was significant for high analyst dispersion and the high-$AGSV^+$ portfolio, indicating that the negative news mispricing

prevails. The negative-positive news strategy also generated significantly positive alpha for the group with low institutional ownership and high $AGSV^+$.

1.5.1.3 Firm size

Firm size serves as a useful indicator of information diffusion rate. We controlled for firm size in the analysis of the news trading strategy. Panel C of Table 1.9 shows that both the negative and positive news trading strategies yielded significant alpha for small firms. For the negative news strategy, abnormal returns were more pronounced for high-$AGSV^+$ stocks in each size group. This finding is consistent with our previous result in table 1.8 that underreaction to negative news is more prevalent with high $AGSV^+$ allocation. The alpha for the positive news strategy was significantly negative for high-$AGSV^+$ groups and small firms. Small-firm stocks with high positive news seem to underperform those with no positive news in the three-month holding period. The negative-positive news strategy generates significant positive alpha for small firms, with greater alpha in the high-$AGSV^+$ group compared with the low-$AGSV^+$ group. Thus, the underreaction to net negative news is more apparent for stocks with high $AGSV^+$ and small firm size.

1.5.1.4 Idiosyncratic volatility

West (1988) has argued that rapid information incorporation reduces idiosyncratic volatility. In our study, we explored the relationship between the speed of news incorporation into stock prices and idiosyncratic volatility. Panel D of Table 1.9 supports West's argument. Within each idiosyncratic volatility category, the abnormal return was much larger for the high-$AGSV^+$ portfolio. For instance, the negative news strategy alpha is 1.15% for the high idiosyncratic volatility, high-$AGSV^+$ portfolio. The positive news strategy is not profitable in

any of the six portfolios. The negative-positive news strategy generated a significantly positive alpha for stocks with large idiosyncratic volatility. This result further indicates that the net negative news strategy Is profitable for stocks with high idiosyncratic volatility.

1.5.1.5 Short interest

The literature provides evidence that short sales improve market efficiency by accelerating price adjustments to both private and public firm-specific information and also contribute to enhancing overall information absorption by the market (see C. X. Chen & Rhee, 2010). Because of the high transaction costs of short sales, short sellers are likely to possess greater sophistication than investors holding long positions. This sophistication means short sellers have a greater capacity to identify and capitalise on arbitrage opportunities. Short-sale constraints may impede certain informative trades, and prohibiting traders from shorting can decelerate price adjustments in response to private information, especially negative news (Diamond & Verrecchia, 1987). Therefore, based on the literature, the degree of news underreaction is likely to be asymmetric for different levels of short interest. If short sellers actively engaging in arbitrage opportunities and online searches artificially drive the retail investor attention effect, then controlling for short interest might cause this effect to diminish.

To investigate the impact of short sales, this study controlled for SIR, defined as the shares sold short divided by the shares outstanding in a given month. Panel E of Table 1.9 shows that the results support the hypothesis that the alpha for the negative news strategy is significant within the group with a high SIR and high $AGSV^{+}$. The value of alpha is close to 1%. The long–short difference between high and low $AGSV^{+}$ in high-SIR groups is statistically significant, driven mainly by the long side. An ambiguous finding emerges here, as alpha is also significant

for the low-SIR, low-$AGSV^+$ group, but the value of alpha is not as large as for the high-SIR, high-$AGSV^+$ group. Our finding supports the hypothesis that short sale constraints can impede price adjustments in response to public information, particularly bad news. The retail attention effect does not completely disappear even after we control for short interest. Further, the value of alpha for high SIR with high $AGSV^+$ is almost twice as large as the high-$AGSV^+$ group in Table 1.8 in the absence of characteristic sorting.

To summarise, after controlling for relevant firm characteristics, we found that the negative news trading strategy still generated significant positive abnormal returns for portfolios with high retail attention, particularly for firms with high analyst dispersion, high idiosyncratic risk, low institutional investor holdings, small firm size and high short interest, all of which are indicators of slower information incorporation and lower market efficiency. The efficiency of public negative news absorption was negatively correlated with the level of retail attention allocated to the stock and also with firm characteristics that are indicative of market inefficiency.

As a final analysis, we repeated the Fama–MacBeth regression of Table 1.5 while including additional controls for firm characteristics and interaction terms between these and both positive and negative news. The dependent variable was the BHAR from month *t + 1* to *t + 3*, using the same control variables as for Column 3 of Table 1.5. Each panel in Table 1.10 corresponds to a separate Fama–MacBeth regression. The results reveal that, after controlling for relevant firm characteristics, the interaction effect for retail investor attention with news remains significant, except for analyst dispersion and idiosyncratic risk. This finding confirms that the influence of retail investor attention on news underreaction cannot be ascribed solely

to the indicators of market inefficiency discussed earlier. The sign of each interaction term between news and the various characteristics is consistent with our hypothesis that the degree of news underreaction is inversely related to the indicators of market efficiency for both positive and negative news. Firms characterised by high idiosyncratic risk, low institutional ownership, small size and high SIR demonstrate a more pronounced news drift effect in the ensuing three-month holding period.

Table 1.10: Fama–MacBeth regression interaction terms with firms' characteristics

Firm characteristic	$AGSV^{+}$ ×News_Positive	$AGSV^{+}$ ×News_Negative	News_Negative ×Characteristic	News_Positive ×Characteristic
Panel A: Analyst dispersion	0.001^{***}	-0.001^{***}	-0.116^{***}	0.416^{***}
	(0.001)	(0.001)	(0.002)	(0.004)
Panel B: Idiosyncratic risk	0.001^{***}	-0.001^{***}	-1.359^{***}	1.903^{***}
	(0.001)	(0.001)	(0.085)	(0.121)
Panel C: Institutional ownership	0.001^{***}	-0.002^{***}	0.140^{***}	-0.103^{***}
	(0.001)	(0.001)	(0.007)	(0.008)
Panel D: Market size	0.001^{*}	-0.001^{**}	0.006^{***}	-0.007^{***}
	(0.001)	(0.001)	(0.000)	(0.000)
Panel E: Short interest	0.0002^{**}	-0.0003^{***}	−0.181	0.455^{***}
	(0.000)	(0.000)	(0.0268)	(0.0425)

Note. The table lists the result of a Fama–MacBeth regression similar to the regression in Column 3 of Table 1.5 but with further interaction of each firm characteristic with positive or negative news. The dependent variable is the buy-and-hold abnormal return from month $t + 1$ to $t + 3$. We repeated the same regression for four to six months and obtained similar results. Each panel represents a single regression with the relevant characteristics. Newey-West standard errors with three lags are employed

1.5.2 Explaining the news effect: robustness test

The primary goal of this section is to confirm that the findings on the news effect are not spurious and do not depend solely on a specific setting. To ensure the robustness of our findings, this subsection examines how replacing the quantity of negative and positive news with total news coverage and news sentiment within the original framework affects the results.

1.5.2.1 News coverage effects

We have shown that abnormal attention can amplify the positive and negative news drift effects over the three months following news arrival, with investors exhibiting different

reactions to positive and negative news. If positive and negative news are not distinguished but aggregated into total news coverage, the heterogeneous impact of attention on news reactions could be obscured or diluted. To assess the robustness of our findings about the news effect, we modified our original settings by replacing the positive and negative news categories with total news coverage as shown in Equation 11.

$$\mathrm{BHAR}_{i,t,N} = a_0 + b_1 {AGSV_{i,t}}^{+} + b_2 News\_coverage_{i,t} + b_3 {AGSV_{i,t}}^{+} \times News\_coverage_{i,t} + b_4 Z_{i,t} \quad (11)$$

In Column 4 of Table 1.5, the coefficient for total news items (News_Coverage) was positive. However, the interaction term between total News_Coverage and AGSV^{+} was not significant. This result suggests that the findings in Table 1.5 might have been influenced by the counteracting effects of abnormal attention on positive and negative news. To examine this, we conducted a long–short news coverage strategy in high- and low-attention groups. For month *t*, stocks were classified into ten portfolios[10] based on total news coverage in each attention group. The strategy applied entailed going long on portfolios in the highest decile of news coverage (Decile 10) and short on those in the lowest decile (Decile 1) and holding these positions from month *t + 1* to *t + 3*. Panel A of Table 1.11 shows the abnormal returns calculated using the Fama–French and Fama–French with momentum models. The abnormal returns were not significant for the two attention groups. This result confirms that the significant abnormal returns observed for the negative news strategy do not merely result from investors' underreactions to firms with high news coverage, independent of the news type.

[10] To maintain balanced portfolio sizes (deciles), we excluded the zero news portfolio strategy for each company, as there were very few months with no news stories.

1.5.2.2 News sentiment effects

Investors react differently to positive and negative news intensity, and both news sentiment (tone) and news intensity can influence market outcomes. If news sentiment is the dominant factor for market outcomes, employing the average sentiment score as a proxy for positive and negative news variables might yield results similar to our original analysis. Therefore, we modified the original settings by replacing the positive and negative news intensity variables with the average news sentiment score for each firm in a given month.

$$\mathrm{BHAR}_{i,t,N} = a_0 + b_1 AGSV_{i,t}{}^{+} + b_2 News\_sentiment_{i,t} + b_3 AGSV_{i,t}{}^{+} \times News\_sentiment_{i,t} + b_4 Z_{i,t} \quad (12)$$

The results in Column 5 of Table 1.5 indicated a significant and positive relationship between news sentiment and the price drift over three months following the arrival of news. However, the interaction term between retail investor attention and news sentiment did not have statistical significance. These findings suggest that abnormal investor attention affects stock price reactions to news primarily by differentially affecting the marginal effects of positive and negative news intensity.

To examine the impact of news sentiment on the stock market, we employed a long–short strategy based on news sentiment for high- and low-attention groups. For month *t*, stocks were divided into ten portfolios, each sorted by their average sentiment score in the respective attention groups. We went long on the portfolio with the highest decile of news sentiment (Decile 10) and short on the portfolio with the lowest decile of news sentiment (Decile 1), and held the portfolio from month *t + 1* to *t + 3*. Panel B of Table 1.11 gives the abnormal returns using the Fama–French and Fama–French with momentum models. Our results show no

significant alpha for the two attention groups, indicating that news sentiment plays a limited role in influencing news mispricing. Combined with our previous findings that news mispricing is predominantly driven by investors' underreaction to negative news, we conclude that asymmetry in investors' sensitivity to positive and negative news, together with the impact of news intensity on stock prices, contributes to the mispricing in the negative news strategy and the absence of significant mispricing in the news sentiment strategy.

In summary, our findings indicate that the observed effects of news are not solely due to either news sentiment or news quantity alone. Specifically, we found that the underreaction to negative news predominated over the effects of positive news and the aggregate impact of total news. For the news sentiment strategy, the alpha was non-significant for both factor models employed. Therefore, the observed mispricing following the arrival of news is influenced primarily by a combination of news quantity and tone, coupled with the asymmetric sensitivity of investors to positive and negative news. Comparing the reactions to the two types of news, we found that the price underreaction to negative news was more pronounced and coherent. The retail attention effect on market reactions to news was also more significant and pronounced in affecting the degree of negative news underreaction.

Table 1.11: Abnormal returns of long–short profits based on news coverage and news sentiment

Panel A: News coverage

Fama–French factors				Fama–French and momentum factors			Fama–French five factors		
Decile 0–10	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High attention	0.13	0.626	0.532	0.173	0.856	0.393	0.183	0.851	0.396
Low attention	0.324	1.57	0.118	0.363	1.795	0.074	0.349	1.627	0.106
High−low	−0.194	−0.915	0.362	−0.19	−0.891	0.374	-0.166	-0.752	0.453

Panel B: News sentiment

Fama–French factors				Fama–French and momentum factors			Fama–French five factors		
Decile 0–10	alpha	statistic	*p*	alpha	statistic	*p*	alpha	statistic	*p*
High attention	0.275	1.543	0.125	0.255	1.437	0.153	0.276	1.486	0.139
Low attention	0.268	1.378	0.17	0.266	1.36	0.176	0.285	1.405	0.162
High−low	0.007	0.03	0.976	−0.01	-0.044	0.965	-0.009	-0.036	0.971

Note. Table 1.11 gives the average abnormal return over three months following the initiation of the long–short zero-investment news strategy sorted by $AGSV^{+}$ using the Fama–French three-factor model, the four-factor model with momentum and the five-factor model with investment and profitability. The return for the news strategy portfolio was the average return of these three cohorts. The table reports the estimates of α for the regression corresponding to Equations 13,14 and 15:

$$\mathrm{R}_{i,t} = a + bMKT_t + cSMB_t + dHML_t + u_t \quad (13)$$

$$\mathrm{R}_{i,t} = a + bMKT_t + cSMB_t + dHML_t + eUMD_t + u_t \quad (14)$$

$$\mathrm{R}_{i,t} = a + bMKT_t + cSMB_t + dHML_t + eCMA_t + RMW_t + u_t \quad (15)$$

where $\mathrm{R}_{i,t}$ is the average return of the long–short spread for the three months following the initiation of the long–short zero-investment strategy sorted by news type.

1.5.2.3 Effects of two subsample periods

To rule out the possibility that our findings on the news and attention effects were influenced by unique events such as the global financial crisis, itself characterised by irrational investor reactions to public news and atypical online search habits, we divided our sample into two periods: pre-2008 and post-2010 and conducted Fama–MacBeth regression analyses for each period. The dependent variables in these analyses were the current month's return and the BHAR from month *t + 1* to *t + 3* ($BHAR_3$). We found that our results remained consistent across the two periods, as shown in Table 1.12. This result reinforces the robustness of our findings, confirming that they were not skewed by any specific event or period.

Table 1.12: Fama–MacBeth regression in two subsample periods

	Dependent variable			
Variable	Ret		$BHAR_{i,t,3}$	
Subperiod	Before 2008	After 2010	Before 2008	After 2010
$AGSV^+$	0.0002***	0.0003***	−0.00000	−0.00004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
News_Positive	0.076***	0.026***	0.043***	0.007***
	(0.0040)	(0.0010)	(0.0050)	(0.0010)
News_Negative	-0.079***	-0.038***	-0.038***	-0.008***
	(0.0040)	(0.0010)	(0.0040)	(0.0010)
$AGSV^+$×News_Positive	0.003***	0.004***	0.003***	0.001***
	(0.0004)	(0.0001)	(0.0005)	(0.0001)
$AGSV^+$×News_Negative	-0.002***	-0.005***	-0.002***	-0.001***
	(0.0003)	(0.0001)	(0.0003)	(0.0000)
lnME	0.007***	0.004***	0.004***	0.004***
	(0.0010)	(0.0004)	(0.0010)	(0.0010)
lnBM	-0.018***	-0.011***	-0.001	0.006***
	(0.0010)	(0.0010)	(0.0020)	(0.0010)
GrProf	-0.003	-0.004	0.017*	0.015***
	(0.0080)	(0.0030)	(0.0090)	(0.0050)
Vol	-0.013***	0.0001	-0.0003	-0.002***
	(0.0050)	(0.0003)	(0.0040)	(0.0004)
Analysts	-0.009***	-0.008***	-0.009***	-0.007***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Constant	−0.012***	−0.004***	−0.008*	−0.001
	(0.002)	(0.001)	(0.004)	(0.003)
Observations	35,926	59,582	35,926	59,582
R^2	0.075	0.075	0.028	0.043

Note. Table 1.12 shows Fama–MacBeth regression results for subsample periods before 2008 and after 2010. The two dependent variables are the current month's return and the buy-and-hold abnormal return from month *t + 1* to *t + 3* ($BHAR_{i,t,3}$). ***/**/* denotes significance at the 0.1%/1%/5% level. Newey-West standard errors with three lags are employed

1.6 Conclusion

This study aimed to investigate how retail investor attention affects the stock price response to public news. To achieve this, we utilised Fama and MacBeth's (1973) cross-sectional regressions to analyse the reaction of stock prices to positive and negative news. The results showed a persistent drift in stock prices after both positive and negative news. The market tended to underreact to public news for around six months. Our study also revealed that the price reaction to public news differs with the level of retail investor attention. We found that higher abnormal attention amplified both the immediate price reaction and the post-news underreaction effect, particularly for negative news. Our analysis reveals compelling economic significance through the interaction terms between retail investor attention (AGSV) and news types. The significant interaction coefficient between AGSV and negative news (e.g., 0.001) indicates that heightened retail attention amplifies the underreaction to negative news. This suggests that as retail investor attention increases, the market's response to adverse information becomes more muted, underscoring a systematic inefficiency driven by uninformed retail traders.

Our analysis reveals that a long–short strategy focused on negative news generated significant alpha for stocks with high retail investor attention, while a corresponding strategy for positive news did not yield significant abnormal returns for either high or low attention groups. These findings suggest that elevated levels of retail investor attention do not enhance market efficiency or the absorption of public information into stock prices. On the contrary, retail investor attention appears to hinder the incorporation of public news into prices. Specifically, our long-short investment strategy demonstrates a robust alpha of 0.557% (t-

statistic 2.66) for high retail attention stocks following negative news. This alpha magnitude not only exceeds the typical benchmarks reported in prior literature, which often range around 0.5%, but also highlights a significant economic opportunity for investors. The ability to achieve such excess returns through a strategy leveraging retail investor attention underscores the economic relevance of our findings.

We further considered relevant firm characteristics to better understand the source of news mispricing. We found that alpha for a long–short negative news strategy was significant for firms with high retail investor attention, especially small firms, firms with high analyst forecast dispersion, high idiosyncratic volatility, low institutional ownership and high short interest. This is consistent with behavioural finance theories, which propose that companies with lower institutional investor coverage and higher uncertainty (as evidenced by high forecast dispersion and high idiosyncratic volatility) are more susceptible to investor sentiment and behavioural biases. We also found that abnormal retail investor attention was positively associated with idiosyncratic risk and negatively associated with institutional ownership, supporting the argument that a higher allocation of retail investor attention to a firm may indicate the presence of individual investors considered uninformed noise traders. The use of Google SVI as a source of investment information poses significant challenges for investors seeking insight. The problem arises because of the likelihood that SVI reflects the search patterns of less sophisticated individual investors. Hence, Google Search trends may not correlate with informed decision-making. A lack of sophistication or an unwillingness to access more comprehensive information may lead to temporary mispricing and increased price volatility. Retail investors may delay the incorporation of news into prices because they lack

sufficient knowledge of a firm and seek information online because of uncertainty. They may need more time to process and interpret a large amount of information. This aligns with the gradual information diffusion model and suggests that investors may selectively ignore value-relevant stories in a noisy information environment. However, alternative explanations should be considered, such as frictions in incorporating public information into stock prices caused by retail investors' searches, particularly for negative news.

This study has implications for the literature on behavioural finance. Contrary to the traditional limited investor attention model of Hirshleifer and Teoh (2003), which predicts that greater investor attention can reduce the mispricing of accruals, we found a positive association between retail investor attention and underreaction to public news. We primarily focused on the role of retail investor attention in the sensitivity of stock returns to public news, specifically examining market mispricing in response to public news intensity and tone. We have proposed a behavioural interpretation to explain the effects of retail investor attention on price reactions to public news. While this study has provided one perspective, it does not exclude the possibility of alternative explanations.

Chapter 2: MD&A disclosure & investors' reactions to management forecasts

2.1 Introduction

Traditional finance has primarily focused on quantitative factors such as earnings, dividends, and financial ratios to explain stock price formation and predict future performance. However, research by Cutler et al. (1989) and Shiller (1981) highlights the limitations of relying solely on these measures, as stock prices often deviate from predictions based purely on quantitative models. Amir and Lev (1996) further argue that while traditional financial metrics remain useful, they do not always fully capture the underlying drivers of stock price movements. Consequently, there has been growing recognition of the importance of qualitative factors—such as media tone, managerial uncertainty, and news sentiment—which can provide insights beyond those offered by quantitative data alone. These qualitative elements enable investors and analysts to form a more nuanced understanding of market dynamics, helping them make more informed investment decisions.

The increasing focus on qualitative factors is particularly significant in the context of management forecasts, which are designed to offer insights into a firm's anticipated performance, risks, and opportunities. However, these forecasts can be subject to biases, especially when managerial discretion comes into play. Managers may use complex or vague language to obscure negative information, leading to potentially misleading investor interpretations. As a result, investors must critically assess the credibility of management forecasts by considering other information sources, most notably the Management's Discussion and Analysis (MD&A) section of financial reports. The U.S. Securities and Exchange

Commission (SEC) mandates the inclusion of MD&A sections in Forms 10-K and 10-Q, recognizing their critical role in providing management's perspective on a firm's current and future performance (Arnold et al., 2008). While safe-harbour provisions allow managers some legal protection in making forecasts, the language used in MD&A disclosures offers valuable qualitative insights into management's expectations and level of confidence.

Research by Bryan (1997), Clarkson et al. (1999), Francis et al. (2002), and Amel-Zadeh and Faasse (2016) has established that MD&A disclosures can influence stock price reactions and predict future financial variables. Studies such as Bryan (1997) and Feldman et al. (2010) demonstrate that the tone, readability, and clarity of MD&A disclosures can affect market perceptions, shaping both short-term stock price reactions and long-term performance expectations. However, much of the existing literature has focused on quantitative aspects of disclosures, such as earnings and capital expenditures, with less attention paid to qualitative linguistic factors like tone, readability, and complexity of managerial language. Li (2010) expanded this discussion by examining the qualitative content of MD&A disclosures, emphasizing the importance of these factors in providing a more holistic view of a firm's performance, risks, and future prospects.

Despite these contributions, a research gap remains in understanding how MD&A linguistic factors specifically affect market reactions to management forecasts. This study aims to address this gap by investigating how the uncertainty, readability, and tone change in MD&A disclosures influence stock price responses to management forecasts. Building on prior work by Baginski et al. (2004) and Li (2010), this study explores whether these qualitative elements provide explanatory power beyond traditional quantitative measures, particularly in shaping

both immediate and delayed market reactions to forecast surprises. The findings will contribute to the literature on voluntary disclosure and offer deeper insights into how the strategic use of language in MD&A sections impacts stock price formation, especially in the context of management forecasts. This key finding highlights the critical role of qualitative factors embedded in MD&A disclosures in shaping investor reactions independently of the forecast's expected quantitative impact. The distinction is crucial as it demonstrates that even when forecast surprises are controlled for, the intrinsic quality of the MD&A—its clarity, uncertainty, and tone—can significantly sway investor perceptions and market behavior. Such insights suggest that investors value transparent and straightforward disclosures, which can mitigate the uncertainty typically associated with managerial forecasts. Moreover, this emphasizes the need for management to focus on the quality of disclosure, not just the content of the forecasts themselves. Future research could further dissect these interactions by examining specific industry sectors or incorporating longitudinal studies to track these effects over time.

Our research methodology involved examining the MD&A sections of 10-K reports from a large sample of companies. Linguistic variables were extracted using Python software and Loughran and McDonald's (2011) word list. To capture qualitative factors, we considered three widely discussed MD&A linguistic factors: uncertainty, readability and tone change. To evaluate firms' management forecast records, we used the company-issued guideline (CIG) dataset of the Institutional Brokers' Estimate System (IBES). For a sample of management forecasts issued from 2002 to 2017, we used regression analysis to demonstrate the influence of MD&A uncertainty, readability and tone change on the stock price reaction to management forecasts. Our results revealed a positive association of MD&A uncertainty, complexity (as

opposed to readability) and tone change with contemporaneous stock price reactions to forecast news.

A second set of tests explored the potential cross-sectional effects of MD&A linguistic factors on post-management abnormal returns. The results confirmed that MD&A complexity and uncertainty can exacerbate price underreactions during the event window. We carried out tests to ensure the robustness of the findings and confirmed that the post–management forecast drift effect is not driven solely by other factors, such as earnings announcements. Our study revealed that MD&A linguistic factors partially influence the market reaction to management forecast news, even after controlling for prior return drift (momentum), earning surprises, firm size, institutional ownership and analyst following.

Our study also examined the asymmetric market reaction to good and bad news of management forecasts. Prior research shows that markets typically react more strongly to negative than positive news, driven by uncertainty and sentiment (Williams, 2009; Kim et al., 2010). We hypothesize that MD&A factors—such as linguistic complexity, uncertainty, and tone—affect market reactions differently depending on the type of forecast. To explore this, we analyze market responses to both good and bad news. Our findings confirm that MD&A complexity amplifies the negative reaction to bad news, while good news accompanied by less readable or more uncertain MD&A disclosures leads to a weaker positive response compared to clearer disclosures.

A fourth set of tests examined the impact of management forecast news and MD&A linguistic factors on investors' searches for 10-K filings in the management forecast announcement period. We collected daily non-missing 10-K filings and computed the daily

volume of EDGAR (Electronic Data Gathering, Analysis, and Retrieval) downloads. We employed two measures to evaluate information acquisition activities after the announcement date: average raw EDGAR searches and average abnormal EDGAR searches, with the latter indicating unexpected attention. Our study revealed that both extremely good and extremely bad forecast news is positively associated with the average raw EDGAR search for 10-K filings. Negative news was also positively associated with unexpected EDGAR search activities, suggesting that negative news is more likely to trigger unexpected information acquisition for the 10-K report. Furthermore, the study confirms the direct and indirect effects of MD&A linguistic factors on post-forecast Edgar search activities. The results suggest a complex interplay between news, disclosure complexity, and investor information demand, warranting further investigation.

A fifth set of tests examined whether MD&A linguistic factors are associated with management forecast precision. Managers often adjust the precision of their earnings forecasts based on their strategic objectives, particularly in influencing investor reactions and stock prices. Prior studies have demonstrated that forecast precision can be a tool for managing market expectations. For instance, less precise forecasts might be used to provide flexibility, allowing managers to adjust future disclosures without being held strictly to their initial estimates. This can be particularly advantageous when conveying uncertain or negative news, as it provides managers a buffer against negative market reactions (Baginski et al., 1993; Rogers & Stocken, 2005). Conversely, more precise forecasts are typically employed when managers are confident about the firm's future performance or when they wish to signal credibility to reduce information asymmetry (Nagar et al., 2003; Ng et al., 2013). The strategic selection of forecast

precision can therefore be understood as part of the broader narrative managers construct through voluntary disclosures, particularly in the Management Discussion and Analysis (MD&A) sections of financial reports, where readability and tone also play critical roles in shaping market reactions (Li, 2010). We proposed that managers should be motivated to reduce information asymmetry by providing a clear and easily understandable MD&A disclosure, which could result in improved forecast precision. Our findings suggest that linguistic factors in the MD&A influence the following quality of management forecasts. Managers may have an incentive to obscure their forecasts by using complex language in the MD&A. Overall, our findings are that managers use their discretion in shaping forecasts through the MD&A and strategically choose forecast precision for self-serving motives.

Finally, we assessed the robustness of our main results by controlling for potential latent variables extracted from the 'Quantitative and Qualitative Disclosures about Market Risk' and 'Risk Factors' sections of the 10-K report. Our approach aimed to mitigate the potential confounding effects of managers' general attitudes or linguistic styles on the market's reaction to management forecasts. The results of the placebo regression analysis confirmed the robustness and consistency of our findings, indicating that the effect of linguistic factors in the MD&A section on stock prices was not a result of the overall linguistic style of managers.

The remainder of this chapter is organised as follows. A literature review is presented in Section 2.2. The data and samples are discussed in Section 2.3. The research design and empirical results are described in Section 2.4. The robustness checks are given in Section 2.5. Section 2.6 concludes the chapter.

2.2 Literature review

2.2.1 Prior literature on the analysis of tone, readability & uncertainty

Our study forms part of the literature on textual analysis for financial markets. The tone used in quantitative information can affect investor behaviour by framing financial performance within a positive market environment, encouraging investors to consider growth relative to a reference point. Research by Tetlock (2007) has shown that pessimistic news in the *Wall Street Journal* can depress stock indices. Li's (2010) research has indicated that optimistic disclosures in MD&A sections result in higher future earnings. Feldman et al. (2010) showed a correlation between tone changes in SEC filing periods and short-term market reactions. Demers and Vega (2008) examined the emotional tone in corporate earnings announcements, finding that unexpected sentiment shifts can explain variations in post-announcement earnings drift. Henry (2008) has shown that a more positive tone in press releases leads to higher abnormal market returns. Our study builds on existing research by investigating the direct association between the tone expressed by management in MD&A sections and contemporaneous and delayed price reactions following management forecasts.

We have also addressed readability in financial reporting. Prior research has found substantial uncertainty in reported earnings. This uncertainty is often attributed to unknown or complex accounting choices and assumptions and leads to doubts about the reliability and sustainability of reported earnings. Mandatory disclosure documents, such as those required by the SEC, have been critiqued as overly complex, lengthy and difficult to understand (ABT SRBI, 2008). Poor readability can impede investors' ability to extract and process valuable information. Clear and straightforward MD&A sections are generally perceived as more

informative. Studies have indicated that report length and readability can act as proxies for disclosure transparency. Longer and less readable reports are viewed as more challenging to analyse and understand (Li, 2008). The Flesch index and the fog index have been employed to quantify financial report readability (see Shroeder & Gibson, 1990; Smith, 1971). The readability of these reports is challenging for many shareholders (Jones & Shoemaker, 1994). Other studies have linked report readability to financial indicators such as firm profitability and return on capital (see Courtis, 1986; Subramanian et al., 1993). Our study is motivated by the findings of You and Zhang (2009) and Callen et al. (2010), who indicate that investor reactions to information are influenced by its levels of complexity and uncertainty and that lower information quality leads to greater stock price delays. We aim to extend the literature on text readability by examining the relationship between MD&A readability and market reactions to management forecasts.

For textual uncertainty, Jiang et al. (2005) and Zhang (2006) have investigated the impact of information uncertainty on stock price drift. Loughran and McDonald (2013) found that initial public offerings (IPOs) with high levels of textual uncertainty experienced higher first-day returns. Building on research by Demers and Vega (2014), our study employed Loughran and McDonald's (2011) word lists to quantify uncertainty in MD&A documents. This approach allows us to isolate the effect of uncertainty and analyse its influence on market reaction to management forecasts.

2.2.2 Management forecasts & market reactions

Our research contributes to the literature about management forecasts, which are documents that should address the information asymmetry between managers and investors or

analysts. Litigation risk and self-interest have been identified as potential factors affecting management forecasts (Cao & Narayanamoorthy, 2005; Nagar et al., 2003). Past forecasting performance is also important, as the accuracy of previous management forecasts affects the credibility of current forecasts (A. Hutton & Stocken, 2007; Williams, 1996). The consequences of forecasts, particularly in terms of stock market reaction, have long been studied. Forecasts affect stock prices (Patell, 1976; Penman, 1980), and while there is extensive literature on price reactions to earnings (see Ball & Brown, 1968; Bernard & Thomas, 1989), investors' reactions to management forecast news warrant further exploration. Forecast credibility and accuracy (Hirst et al., 1999), forecasting reputation (A. Hutton & Stocken, 2007), the nature of news (Kothari et al., 2005), the forecast form (Baginski et al., 1993) and the time horizon (Pownall et al., 1993) have been studied in relation to stock market reactions to management forecasts. The supplementary information accompanying management earnings forecasts has also been found to affect the response, with standalone forecasts being more informative than forecasts combined with earnings announcements (Atiase et al., 2005).

In summary, the literature has shown that the market reaction to management forecasts depends upon characteristics and additional information that may affect the informativeness of these forecasts. For example, verifiable forward-looking information that enhances the informativeness of management forecasts tends to have a more substantial impact on market reactions. To our knowledge, our study represents the first attempt to incorporate textual analysis into firm-level management forecasts and examine its influence on stock price reactions. We examine the relationship between the quality of MD&A disclosures and abnormal returns following management forecasts for different time windows. Our results demonstrate

that textual uncertainty and readability in the MD&A section have a significant effect on post–management forecast outcomes. This result goes beyond factors established in the literature, offering insights into the consequences of management forecasts in the stock market.

2.3 Data & sample

2.3.1 Sample construction and data sources

The data for this study were obtained from several sources. Management forecast data were sourced from the CIG dataset within IBES. A management earnings forecast may take the form of a point, range, or open-ended forecast. Our analysis focused on annual and quarterly earnings per share (EPS) forecasts, excluding open-ended forecasts and those without specific numerical values. Multiple forecasts issued on the same day for the same forecast period were removed to avoid redundancy. The final sample included 74,475 management forecasts from 2001 to 2017.

We restricted the sample to firms with common stocks listed on the NYSE, AMEX, or NASDAQ. Return data were sourced from CRSP, and firm-specific financial data were collected from Compustat. We applied several filtering criteria to ensure data quality: only firms with a CUSIP identifier, an equity market value exceeding $5 million, positive total assets and sales, and a share price above $5 at the end of the quarter were included. Additionally, forecasts with confounding events, which could result in discontinuity in the EPS, were excluded. After matching the IBES, CRSP, and Compustat datasets, our final sample comprised 45,790 forecasts.

Figure 2.1

Data collection process

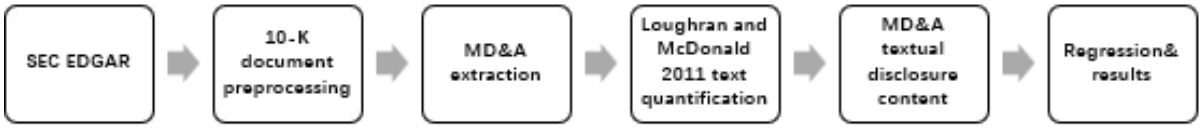


Table 2.1: Data sources

Source	Obs
EDGAR	48,786
10K/10-K405	
IBES first call	74,475
Annual and quarterly management earnings forecasts for the fiscal years from 2001 to 2017	
Point and range management earning forecasts	
Removal of duplicate observations or observations with multiple EPS forecasts for the same forecast period on the same forecast date	
Removal of open-ended and qualitative forecasts	
Compustat	548,840
Valid Committee on Uniform Securities Identification Protocol number (CUSIP)	
Equity market value exceeding $5 million at end of quarter	
Positive average total assets and sales during the quarter	
Price per share greater than $5.00 at end of quarter	
Combined data	45,790
Matched EDGAR MD&A, IBES and Compustat data	

To ensure the relevance of the MD&A disclosures, we extracted the most recent 10-K reports within a 300-day window before each forecast announcement. Since the enactment of the Regulation Fair Disclosure law in 2013, firms have been permitted to disclose certain information via social media. However, for this study, all SEC filings from 2001 to 2017 (classified as 10-K, 10-K/A, 10-K405/A, 10-K405, 10-KSB, 10-KSB/A, 10KSB, 10KSB/A, 10KSB40 or 10KSB40/A.) were collected using a Python web crawler. The MD&A sections (Items 7 and 7a) were extracted from these filings for analysis. Examples of the MD&A overview sections are included in the appendix for quick reference.

2.3.2 Construction of Key Variables

The key variables in this study include MD&A disclosure metrics and management forecast variables.

MD&A Linguistic Metrics: We used two linguistic factors to assess the readability of the MD&A disclosure: the word count and the fog index. The word count is the number of words in the MD&A section, and the complexity of its text content was evaluated using the fog index. The fog index employs two variables: the average sentence length (as the number of words) and the number of complex words (words with more than two syllables).[11] Together, these factors are used to estimate the educational grade level necessary for understanding the text. The fog index is calculated as shown in Equation 16:

$$Fog\ index = 0.4 \times (words\ per\ sentence + \%\ complex\ words) \qquad (16)$$

We also investigated two additional factors: uncertainty and tone change. Uncertainty is quantified as the proportion of uncertain words, while tone change is determined as the percentage change in net optimism from the previous year to the current year. Net optimism is calculated as the ratio of positive to negative words. We used tone change rather than tone level to assess managers' net optimism, following Davis et al. (2012), Demers and Vega (2008) and Tetlock et al. (2008).[12]

---

[11] Beyond the components of the fog index, empirical research has demonstrated the significance of language features in documents, such as their level of concreteness (see Elliott et al., 2015) and vividness (see Hales et al., 2011).

[12] Given that managers often rely on previous MD&As when creating a new MD&A with small incremental changes, MD&As from years that are close in proximity may exhibit similarities, especially the manager's tone. Existing research indicates that a measure of tone change may offer a more robust means of evaluating a manager's sentiment compared with a measure of tone level. Therefore, our tone change measure may be a better indicator than the tone level measure of Kothari and Short (2003).

Management Forecast Variables: Management forecast surprises were computed as the difference between the announced EPS forecast and the consensus EPS forecast, scaled by the pre-announcement stock price.

2.3.3 Descriptive Statistics

Table 2.2 gives the summary statistics for the event-level sample. The median word count in the MD&A sections was 3,713.5, and the median proportion of complex words was 0.372. This pattern suggests a preference for companies to employ complex words in their MD&A disclosures. The average proportion of negative words surpassed positive words, pointing to a tendency for managers to use pessimistic language in the disclosures. The MD&A factors showed a slight skewness, with means slightly larger than medians, other than for the ratio of positive words. Our sample encompassed firms of various sizes with an average earnings surprise of 39.41 and an average earnings persistence of 0.228. The management forecast surprise ranged from a maximum of 0.059 to a minimum of−5.437. The control variables used are also presented in the table.

Table 2.2: Summary statistics for firms' variables

Variable	Obs	Min	Max	Median	Mean	SD	Q1	Q3
MD&A variable								
Word count	45,790	0	5,300,446	3,713.5	4,478.559	83,485.252	2,272.25	5,773.25
Uncertainty	45,790	0	0.125	0.016	0.018	0.004	0.011	0.022
Negative ratio	45,790	0	0.09	0.016	0.017	0.004	0.01	0.022
Positive ratio	45,790	0	0.039	0.008	0.008	0.002	0.006	0.011
Fog index	45,790	0	420.409	8.156	9.06	10.634	7.348	9.353
Complexity	45,790	0	0.812	0.372	0.375	0.160	0.344	0.403
Control variable								
MF	45,790	−5.437	0.059	0.000	0.000	0.025	0.000	0.001
Earnp	45,419	−10.828	8.5	0.182	0.228	0.39	−0.026	0.481
Sue	43,150	−3.07	1,692,004	0.001	39.414	8,145	−0.002	0.004
B/M	45,789	0	13.36	0.45	0.549	0.449	0.27	0.7
ME	45,790	6.84	572,530	1,766	7225	19,401	596.693	5,622.557
Turnover	45,790	0	0.114	0.006	0.008	0.005	0.004	0.01
Loganalyst	43,717	0.693	3.892	1.946	1.956	0.659	1.386	2.485
IOR	45,790	0	2.015	0.82	0.784	0.209	0.674	0.923
Pastret	45,790	−0.806	1.561	0.01	0.009	0.103	−0.043	0.061
Momentum	45,751	−0.968	15.62	0.113	0.165	0.524	−0.099	0.332
Vola	45,790	0.005	0.174	0.021	0.023	0.011	0.015	0.028
Loghorizon	43,717	4.5	6.277	4.771	4.801	0.137	4.718	4.836

Note. Table 2.2 gives summary statistics for firm variables from 2001 to 2017. Earnings surprise (Sue) is the actual earnings per share (EPS) minus the consensus forecast EPS, adjusted by the pre-announcement price. Earnings persistence (Earnp) is the four-year quarterly EPS first-order autocorrelation with a four-observation minimum. Book-to-market ratio (B/M) is the firm's ratio at the end of the previous fiscal year. ME is the beginning-of-month market equity value ($ million). Pastret is the stock return of the firm during the month preceding the forecast announcement. Turnover is the 250-day average daily trading volume (over 250 days) from $t - 255$ to $t - 5$. LogAnalysts is the logarithm of one plus the number of analysts following the firm during the most recent quarter. IOR is the institutional ownership ratio: the proportion of shares held by institutions at the end of the most recent calendar quarter. Momentum is the 12-month cumulative raw return ending one month before the forecast. Vola is the return volatility standard deviation from $t - 255$ to $t - 5$. Loghorizon is the natural logarithm of the number of days between the forecast announcement date and the end of the fiscal year plus 90. *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

2.4 Research design & empirical results

2.4.1 Hypothesis Development

Management forecasts are pivotal in shaping investor expectations by providing insights into a company's future performance. Investors rely on these forecasts for short-term trading decisions, leading to price movements around announcement dates. The effectiveness of investor interpretation depends not only on the forecast content but also on the clarity and complexity of the accompanying Management Discussion and Analysis (MD&A) disclosures.

The MD&A section contextualizes the forecast by detailing company operations, risks, and forward-looking statements. When these disclosures are complex—characterized by high word counts, low readability (measured by the Fog Index), and elevated uncertainty—they can obscure key information about a firm's prospects, making it harder for investors to accurately assess the company's financial health (Loughran and McDonald, 2014). This obscurity increases the information processing costs for investors, leading to heightened uncertainty and potentially more pronounced short-term market reactions.

Prior research supports this notion. Miller (2010) found that increased reporting complexity leads to higher information asymmetry and greater trading volume around earnings announcements, suggesting that investors react more strongly when disclosures are harder to interpret. Similarly, Li (2008) and You and Zhang (2009) demonstrated that less readable financial reports are associated with greater investor uncertainty, which in turn heightens the market's sensitivity to new information. The difficulty in interpreting complex disclosures creates ambiguity, prompting investors to react more strongly to subsequent management forecasts as they strive to resolve this uncertainty.

Moreover, Rogers and Stocken (2005) found that investors react more intensely to management forecasts when prior information is ambiguous or difficult to interpret. Additionally, Jiang, Lee, and Zhang (2005) suggest that heightened uncertainty in financial disclosures leads to stronger price reactions as investors adjust their expectations for future firm performance. Investors may perceive increased complexity and uncertainty as signals that the forecast carries more risk or significance, leading them to rely more heavily on management forecasts to make investment decisions. This reliance can result in more pronounced price movements in the short-term period surrounding the announcement.

Therefore, we expect that the complexity and uncertainty of MD&A disclosures are positively correlated with the magnitude of the market reaction to management forecasts. More specifically, firms with higher Fog Index scores, longer MD&A sections, and greater uncertainty measures should see stronger short-term market reactions. This expectation aligns with the idea that investors, facing greater challenges in processing complex information, respond more strongly to management forecasts that provide much-needed clarity.

Hypothesis 1: Companies with more complex and uncertain MD&A disclosures experience stronger short-term market reactions to management forecasts.

While short-term market reactions are immediate, the complexity of MD&A disclosures may also have a longer-term impact by contributing to market underreaction to management forecasts. This phenomenon occurs when investors fail to fully incorporate new information into stock prices immediately after the forecast, leading to a drift in stock prices as the market gradually absorbs the implications over time.

Complex and uncertain MD&A disclosures increase the cognitive load for investors, making it harder for them to process and act on the information promptly. Lower readability and greater uncertainty in these disclosures elevate the information processing costs when evaluating forecast news. As a result, stock prices may not adjust fully to the new information immediately following the forecast. Over time, as investors continue to interpret the complex disclosures, stock prices may drift, reflecting a delayed adjustment to the management forecast.

Research by Li (2008) and You and Zhang (2009) supports the idea that more complex and less readable disclosures hinder the timely incorporation of new information into stock prices. These studies demonstrate that when financial reports are difficult to understand, it takes longer for investors to process the available information, leading to slower price reactions. In the context of management forecasts, the combination of complex MD&A disclosures and forecast news adds an extra layer of uncertainty, causing investors to underreact to the new information. Bernard and Thomas (1989) documented that a significant portion of post-announcement drift—where stock prices continue to move after the initial announcement—occurs within the first 60 trading days. Building on this insight, we hypothesize that this drift is more pronounced when MD&A disclosures are more complex and uncertain. The additional processing time required for investors to comprehend the full implications of complex forecasts leads to a slower adjustment in stock prices. Therefore, we hypothesize that the complexity and uncertainty of MD&A disclosures exacerbate the market's underreaction to management forecasts.

Hypothesis 2: The market underreaction to management forecasts is more pronounced when the forecasts are accompanied by higher levels of complexity and uncertainty in the MD&A disclosures.

2.4.2 Event analysis: MD&A linguistic factors & market reactions to management forecasts

We measured the management forecast (MF) for EPS using the surprise formula (Equation 17). This formula quantifies the forecast surprise by dividing the difference between the MF and the actual EPS figure (AF) by the stock price (P).

$$\text{MF surprise} = (X - AF)/P \tag{17}$$

The variable P denotes the stock price two days prior to the announcement date, while MF surprise reflects the management's EPS forecast surprise. The determination of X depends on the nature of the management EPS forecast. For point forecasts, X is the lower bound of the forecast. For range forecasts, X is the mean of the upper and lower bounds. AF is the analyst consensus median forecast prior to the management forecast, as sourced from the non-split-adjusted IBES summary file.

To ensure consistency, we aligned the initial non-split-adjusted management forecasts obtained from the IBES first call with the non-split-adjusted analyst forecasts from IBES. We achieved this by adjusting the EPS figures with share split factors obtained from the CRSP database. This approach ensured that both sets of forecasts were based on an equivalent number of outstanding shares. We calculated the cumulative abnormal returns (CAR) from a firm's management forecast announcement by comparing the firm's buy-and-hold return with that of a portfolio matched by size and book-to-market (B/M) ratios over the window periods [−1, 1]

(see Equation 18) and [2, T] (see Equation 19) relative to the trading days around the announcement date.

$$CAR[-1,1]_{I.t} = \prod_{k=t-1}^{t+1}(1+R_{i,k}) - \prod_{k=t-1}^{t+1}(1+R_{p,k}) \quad (18)$$

$$CAR[2,T]_{I.t} = \prod_{k=t+2}^{t+T}(1+R_{i,k}) - \prod_{k=t+2}^{t+T}(1+R_{p,k}) \quad (19)$$

where $R_{i,k}$ is the return of firm *I* and $R_{p,k}$ is the return of the size-matched B/M portfolio following the methodology of Fama and French (1992, 1993), on day k, with t being the announcement date of the management forecast for each quarter.

We conducted regression analysis using all announcements and used the model in Equation 20 to estimate the three-day abnormal returns (CAR[−1,1]).

$$MF\_CAR_{-1,1} = a_1 D.MF + a_2 D.MD\&A + a_3(D.MF * D.MD\&A) + \sum_{i=1}^{k} b_i Z_i + \sum_{i=1}^{k} c_i (D.MF * Z_i) + \varepsilon \quad (20)$$

Based on Hypothesis 1, we expect that the complexity and uncertainty of MD&A disclosures are positively correlated with the magnitude of the market reaction to management forecasts ($a_3 > 0$). More specifically, firms with higher Fog Index scores, longer MD&A sections, and greater uncertainty measures should see stronger short-term market reactions. This is consistent with the idea that investors, facing greater challenges in processing complex information, rely more heavily on subsequent management forecasts to make investment decisions.

The dependent variable in our analysis was the size- and B/M-adjusted three-day CAR ($MF\_CAR_{-1,1}$) from the day before the management announcement ($t - 1$) to the day after the announcement ($t + 1$). To mitigate potential outliers and non-linearities, we use scaled decile ranks of management forecast surprise in our regression analyses.[13] We categorised quarterly management forecast announcements into ten groups based on their degree of surprise and created the variable D.MF, which represents the decile rank of management forecast surprises in each quarter. Similarly, we divided each MD&A linguistic factor into ten groups and generated the variable D.MDA to denote the decile rank of these factors in each quarter. The decile rank was established by independently sorting announcement observations on a quarterly basis. In the regression analyses, we included firm characteristic variables as control variables. These encompassed earnings persistence, decile rank of earnings surprise, stock turnover, momentum, forecast horizon, institutional ownership, return volatility, the logarithm of analyst coverage, B/M, firm size and past one-month returns. Earnings surprise is the difference between the actual EPS and the consensus analyst forecast of EPS before the earnings announcement. This difference was adjusted by the pre-earnings-announcement adjusted share price. Earning persistence was measured as the first-order autocorrelation coefficient of quarterly EPS over the past four years and required a minimum of four observations. The institutional ownership ratio (IOR) was the percentage of shares held by institutions at the end of the most recent calendar quarter. Return volatility was calculated as the standard deviation of daily returns over the previous 250 days, from day $t - 255$ to day $t - 5$. Turnover was the

[13] Research has shown a nonlinear relationship between abnormal returns on announcement days and earnings surprises, particularly with notable impacts from small negative surprises (see Kothari, 2001). Accordingly, we followed previous studies and utilised the decile rank of the management forecast error rather than the raw forecast error.

average daily trading volume over the previous 250 days. LogAnalyst was the logarithm of the number of analysts following the firm during the most recent quarter plus 1. Momentum was the cumulative raw return over a 12-month period, ending one month prior to the management forecast month. Size was the market value of equity at the beginning month of the management forecast date ($ million). Forecast horizon was the natural logarithm of the number of days between the forecast announcement and the fiscal year end plus 90. The B/M ratio is the ratio of the book value to the market value of equity at the end of the previous fiscal year (t).

The study examined the market's short-term reaction to management forecast news through regression analysis while considering the influence of MD&A linguistic factors. Year and industry effects were controlled for using the Fama–French 49 industry classification framework. Columns 1 and 2 of Table 2.3 present the main regression effect without the inclusion of MD&A linguistic factors. CAR [−1,1] is expressed as a percentage. We found the coefficient of D.MF to be significantly positive, indicating a substantial difference in adjacent deciles of forecast surprise for short-term market reaction, with investors reacting positively to positive forecast news and negatively to negative news. Columns 3 to 6 of Table 2.3 give the MD&A variables and their interaction terms in the regression. The result is consistent with our expectations. After the inclusion of MD&A factors in the regression, the interaction term between D.MD&A and D.MF was significantly positive for all four MD&A factors. This finding indicates that investors' short-term reactions are stronger for forecasts accompanied by less certain, less readable and more optimistic MD&A disclosures. Specifically, the interaction term between D.MF and D.Uncertainty in Column 3 is positive (0.0335) and significant at the 1% level. Similarly, the interaction term between D.MF and D.Fog in Column 4 is significantly

positive (0.0274), suggesting that the difference in short-term market reactions to adjacent deciles of forecast surprise is greater by 0.028% for firms with more complex MD&A disclosures. Columns 5 and 6 exhibit positive and statistically significant coefficients for D.MF and D.Wordcount and D.MF and D.Tonechg, respectively. Column 6 shows that the market reacts more strongly to forecast news from firms with more optimistic MD&A disclosures. The short-term market reaction for adjacent deciles of forecast surprise is 0.33% higher for firms with more optimistic MD&A disclosures in their 10-K reports. The magnitude of this marginal effect, expressed as the coefficient of D.MF and D.MD&A, accounts for nearly 3% of the main effect in our results.

The MD&A factors provide incremental informational content beyond financial measures. Table 2.3 presents evidence that all four MD&A factors exhibit significant negative correlations with short-term abnormal returns, even after controlling for other financial measures. Notably, the coefficient of D.Fog is negative (-0.14) and significant, indicating that firms with more complex MD&A disclosures underperform those with more readable disclosures during the three-day announcement window. This evidence suggests that MD&A disclosures carry valuable information that impacts short-term returns, independent of the forecast surprise. Firms with less readable (characterized by longer word length and higher fog index), less certain, and more optimistic MD&A disclosures tend to underperform those with more readable, more certain, and more pessimistic disclosures in the immediate post-announcement period. Our analysis shows that MD&A disclosure quality not only directly influences management forecast announcement returns but also interacts with forecast surprises to shape these returns. By controlling for additional firm characteristics, we demonstrate that

MD&A factors provide investors with incremental value beyond the forecast itself. The finding highlights the dual influence of forecast surprise and MD&A disclosure quality on stock price movements.

Table 2.3: Regression analysis: contemporaneous market reactions to management forecasts

	MD&A factor					
	(1)	(2)	(3)	(4)	(5)	(6)
CAR[−1,1]	Main effect		Uncertainty	Fog index	Word count	Tone change
D.MF	0.919***	0.407***	4.007***	3.990***	4.000***	4.167***
	(0.013)	(0.083)	(0.610)	(0.610)	(0.610)	(0.648)
D.MD&A			−0.156***	−0.146***	−0.134***	−0.173***
			(0.037)	(0.035)	(0.035)	(0.038)
D.MD&A×D.MF			0.0335***	0.0274***	0.0261***	0.0329***
			(0.006)	(0.005)	(0.005)	(0.006)
Earnp	0.0814	0.176	0.0571	0.0684	0.0683	−0.16
	(0.084)	(0.213)	(0.269)	(0.267)	(0.267)	(0.296)
Sue	0.556***	0.623***	0.633***	0.633***	0.633***	0.633***
	(0.012)	(0.028)	(0.034)	(0.034)	(0.034)	(0.036)
Loganalyst	0.007	−0.194	0.0446	0.0332	0.0346	0.0707
	(0.057)	(0.143)	(0.174)	(0.174)	(0.174)	(0.184)
IOR	1.327***	−0.0244	−0.059	−0.109	−0.116	−0.608
	(0.166)	(0.383)	(0.572)	(0.570)	(0.570)	(0.597)
Pastret	−2.886***	−1.598	−1.951	−1.933	−1.93	−2.227
	(0.471)	(1.077)	(1.442)	(1.439)	(1.440)	(1.573)
Momentum	2.161***	2.968***	3.737***	3.736***	3.732***	3.646***
	(0.153)	(0.363)	(0.454)	(0.450)	(0.451)	(0.498)
Vola	4.902	−97.16***	−96.82***	−99.66***	−99.68***	−99.23***
	(6.491)	(13.360)	(16.390)	(16.100)	(16.150)	(18.370)
Controls interacted with MF	NO	YES	YES	YES	YES	YES
Constant	−5.349**	−2.876	−24.74***	−24.58***	−24.66***	−28.50***
	(2.436)	(2.532)	(5.228)	(5.243)	(5.248)	(3.924)
Observations	64,009	64,009	43,141	43,141	43,141	38,602
R^2	0.198	0.21	0.237	0.237	0.237	0.239

Note. Table 2.3 presents the results of multivariate tests used to investigate the influence of linguistic factors in the Management Discussion and Analysis (MD&A) section on the relationship between contemporaneous market reaction and management forecast surprises. Columns 1 and 2 give the primary regression effects without incorporating MD&A linguistic factors. Columns 3 to 6 separately introduce individual independent variables of D.MD&A representing the decile ranks of MD&A linguistic factors into the regression model. CAR[−1,1] is given as a percentage. Our study controlled for year and industry effects. The industry classification used the Fama–French 49-industry framework. Standard errors were adjusted for heteroskedasticity and clustering by firm and quarter. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

$$MF\_CAR_{2,T} = a_1 D.MF + a_2 D.MD\&A + a_3(D.MF * D.MD\&A) + \sum_{i=1}^{k} b_i Z_i + \sum_{i=1}^{k} c_i (D.MF * Z_i) + \varepsilon \quad (21)$$

Our second hypothesis was that the market underreaction to management forecasts will be more pronounced when the forecasts are associated with higher levels of complexity and uncertainty in the MD&A disclosure. Lower readability and greater uncertainty in MD&A disclosures increase the information processing costs for investors when evaluating forecast news. This may lead to a stronger underreaction after the forecast. To test this, we chose a drift window from Day 2 to Day 45 as a benchmark, as the MD&A effect is most significant during this period.[14] Research by Bernard and Thomas (1989) has indicated that the majority of earnings drift (continued price movement after an earnings announcement) occurs within the first 60 trading days (approximately three calendar months). To reduce the impact of possible errors in the announcement date on return sensitivities, we calculated abnormal returns for the announcement and post-announcement periods using various time windows. The study demonstrated consistent results when employed within a two-month window. We chose CAR[2, 45] as the dependent variable for the primary analysis.

Columns 1 and 2 of Table 2.4 present the main effect, where the coefficient of D.MF indicates the degree of underreaction to the management forecast. A positive coefficient implies that positive news demonstrates better performance in subsequent periods compared with

[14] We repeated the regression analysis using different CAR[2, T] values as the dependent variable for a two-month window after the announcement. The results were consistent with our primary analysis, but the underreaction effect was no longer significant beyond the two-month window.

negative news, indicating an underreaction to management forecasts. Columns 3–6 show the marginal effects of MD&A linguistic factors on the underreaction. The coefficient of the interaction term between D.MF and D.MD&A represents the difference in market underreaction for adjacent deciles of forecast surprise. Column 3 shows that the interaction term for D.MF and D.Uncertainty is significant and that the market underreaction for adjacent deciles of management forecast is greater by 0.023% for firms with more uncertain MD&A disclosures. Similarly, Column 4 shows that the market underreaction to adjacent deciles of management forecast is greater by 0.023% for firms with a more complex MD&A section. The analysis also included other MD&A factors, such as word count and tone change. Column 5 reveals that a longer MD&A section correlates with a stronger price underreaction to the management forecast. This finding aligns with the result in Column 4, suggesting that readability acts as a mitigating factor for the drift effect. Additionally, the analysis controlled for the interactions of other factors with D.MF, including earnings persistence, firm size, institutional ownership, decile rank of earnings surprise, B/M ratio, turnover, analyst coverage, momentum and volatility. After controlling for these factors, most linguistic parameters, other than tone change, still showed a significantly positive coefficient for D.MF and D.MD&A.

Table 2.4: Regression analysis: management forecasts and post-announcement returns

	MD&A factor					
	(1)	(2)	(3)	(4)	(5)	(6)
CAR[2,45]	Main effect		Uncertainty	Fog index	Word count	Tone change
D.MF	0.919***	0.305**	0.336*	0.326*	0.324*	0.345*
	(0.013)	(0.142)	(0.177)	(0.18)	(0.18)	(0.189)
D.MD&A			-0.141**	-0.169***	-0.175***	-0.0406
			(0.0588)	(0.0562)	(0.0568)	(0.0602)
D.MD&A×D.MF			0.0232**	0.0226**	0.0235***	0.0051
			(0.00916)	(0.00883)	(0.00892)	(0.00939)
Earnp	0.149	0.0923	-0.241	-0.227	-0.229	0.21
	(0.229)	(0.543)	(0.671)	(0.667)	(0.667)	(0.613)
Sue	0.145***	0.0564	0.0439	0.0435	0.0437	0.0579
	(0.023)	(0.049)	(0.0614)	(0.0614)	(0.0614)	(0.0631)
Loganalyst	0.255**	0.354	0.384	0.378	0.38	0.331
	(0.103)	(0.247)	(0.3)	(0.3)	(0.3)	(0.314)
IOR	0.838***	2.610***	1.572	1.552	1.557	1.321
	(0.324)	(0.709)	(0.973)	(0.968)	(0.968)	(1.022)
Pastret	−5.211***	-7.030***	-5.772***	-5.779***	-5.783***	-4.004*
	(0.897)	(2.068)	(2.234)	(2.232)	(2.232)	(2.336)
Momentum	1.184***	1.326***	1.673***	1.675***	1.671***	1.921***
	(0.176)	(0.373)	(0.486)	(0.488)	(0.487)	(0.524)
Vola	61.81***	74.56***	62.87**	61.43**	62.06**	35.46
	(12.65)	(24.61)	(27.56)	(27.17)	(27.18)	(27.88)
Controls interacted with MF	NO	YES	YES	YES	YES	YES
Constant	4.694	3.444	-13.6	-13.36	-13.36	-5.175***
	(4.394)	(4.485)	(8.681)	(8.694)	(8.678)	(1.394)
Observations	64,027	64,027	42,924	42,924	42,924	38,539
R^2	0.013	0.014	0.016	0.017	0.017	0.018

Note. Table 2.4 presents multivariate tests for the impact of linguistic factors in the Management Discussion and Analysis (MD&A) section on the relationship between post-announcement cumulative abnormal returns and management forecast surprises. Our analysis focused on trading days in the [*t + 2*, *t + 45*] window after the management forecast announcement. The dependent variable CAR[2,45] is presented as a percentage. We controlled for year and industry effects using the Fama–French 49 industry classification framework. Standard errors were adjusted for heteroskedasticity and clustering by the firm and quarter. Significance (*, ** and ***) is indicated at the 10%, 5% and 1% levels, respectively.

2.4.3 Event analysis: the asymmetric impacts of good & bad news

The literature indicates that the market reaction to information release is asymmetric, with a stronger reaction to negative news than positive news. An intriguing explanation for this asymmetry lies in higher uncertainty in information dissemination, which can instil pessimism in market participants (Bird & Yeung, 2012; Williams, 2009). Another perspective is that the initial asymmetric response is shaped by market sentiment at the time of the announcement. That is, positive news elicits a stronger market response during periods of positive sentiment, while negative news prompts a stronger response in times of negative sentiment (Jiang, 2011; Mian & Sankaraguruswamy, 2008). Our research supports the idea that a manager's state of mind (as reflected in the tone and uncertainty of the MD&A) and the complexity of the MD&A information that investors must process jointly influence market reactions to management forecasts. When there is high uncertainty and information complexity in MD&A reports, investors may struggle to comprehend the implication of management's forecasts for firm valuation, particularly at times of prevailing negative sentiment. Hence, it is not surprising that a stronger negative drift occurs following the disclosure of unfavourable news.

In this study, we began by examining differential underreactions to positive and negative news forecasts. The literature documents an asymmetric market response to positive and negative earnings surprises. Given this primary asymmetry, we cannot assume that the influence of MD&A factors on market reactions will be symmetrical for positive and negative management forecast news. To investigate asymmetry in the impact of MD&A factors on market reactions to management forecast surprises, we categorised negative earnings surprises in five equal-sized groups, Quantiles 1 to 5. Zero surprises were assigned to Quantile 6. We

allocated positive surprises to Quantiles 7 to 11, following the approach proposed by DellaVigna and Pollet (2009). The variable Good news was represented by a dummy value of 1 for Quantiles 9 to 11 and 0 otherwise. Similarly, the variable Bad news had a value of 1 for Quantiles 1 to 3 and 0 otherwise. The variable Neutral news had a value of 1 for Quantiles 5 to 7 and 0 otherwise. We controlled for year and industry effects.

Column 1 of Table 2.5 shows a statistically significant short-term market response to both positive and negative news without incorporating MD&A factors in the regression analysis. However, after controlling for MD&A factors and including interaction terms between the two types of news and MD&A factors, our results indicate a more significant market reaction to negative than positive forecasts. The regression results presented in Columns 2 to 5 offer further evidence that firms with higher uncertainty, lower readability and greater optimism in their MD&A disclosures experience a more pronounced price response to both positive and negative forecast news. Specifically, the coefficients of the interaction terms for the four D.MD&A factors with positive news are significantly positive, whereas those of the interaction terms for the D.MD&A factors with negative news are significantly negative. These findings indicate that investors have a stronger reaction to forecasts with higher levels of uncertainty, complexity and optimism in the MD&A disclosure. For the effect on the post-announcement drift, we observed an asymmetric and ambiguous effect of the two types of news. The main finding from Column 6 is that there is a significant underreaction to positive news forecasts. However, after accounting for MD&A factors and incorporating their interaction with the two news types, our analysis revealed a significant underreaction to negative news, while the underreaction to positive news is not statistically significant. These results stand in contrast to findings by Jeffrey

and Rodrigo (2013) of a significant underreaction to positive management forecast news and a non-significant underreaction to negative forecasts.[15] Our results provide evidence of asymmetry in the news drift and MD&A effects. Specifically, negative interaction effects between bad news and the MD&A factors, D.Uncertainty, D.Fog and D.Wordcount, in Columns 7, 8 and 9 are statistically significant at the 1% level with magnitudes of −0.117,−0.145 and−0.130, respectively. The coefficient of the interaction term between D.MD&A and positive news is non-significant in the CAR[2, 45] regression. Although the evidence regarding the asymmetry of the news effect remains mixed, our results indicate a positive association between MD&A uncertainty and complexity and the degree of underreaction to negative news. This finding is consistent with our previous results in table 2.4 and emphasises the importance of considering the intricate details of financial disclosures when analysing market reactions.

[15] In our regression analysis, when we did not control for MD&A factors, we found a significant underreaction to positive news and an insignificant underreaction to negative news, consistent with Jeffrey and Rodrigo's (2013) findings. Our results also support their finding that the short-term market reaction to negative forecast news has a greater magnitude than positive forecast news.

Table 2.5: Asymmetric impacts of good and bad news: additional analysis

	MD&A factor									
	CAR[−1,1]					CAR[2,45]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variable	Main	Uncertainty	Fog index	Word count	Tone change	Main	Uncertainty	Fog index	Word count	Tone change
Good news	3.541***	−0.458	−0.456	−0.462	−0.249	0.586***	0.339	0.58	0.501	0.954
	−0.0805	(0.567)	(0.569)	(0.569)	(0.608)	(0.145)	(1.026)	(1.033)	(1.033)	(1.076)
Bad news	−3.664***	−3.303***	−3.312***	−3.380***	−2.815***	0.0153	−2.193**	−2.007*	−2.074*	−1.572
	(0.074)	(0.631)	(0.635)	(0.635)	(0.666)	(0.125)	(1.106)	(1.117)	(1.118)	(1.168)
D.MD&A		0.0161	−0.00652	−0.00595	0.0361*		0.015	0.0135	0.00134	−0.00561
		(0.019)	(0.018)	(0.018)	(0.020)		(0.035)	(0.033)	(0.034)	(0.037)
Good news ×D.MD&A		0.142***	0.120***	0.123***	0.0758**		0.0372	−0.0289	−0.00817	0.0199
		(0.0304)	(0.0284)	(0.0285)	(0.0319)		(0.0527)	(0.0501)	(0.0504)	(0.0542)
Bad news ×D.MD&A		−0.0924***	−0.0751**	−0.0614*	−0.140***		−0.117**	−0.145***	−0.130**	−0.0374
		(0.035)	(0.033)	(0.033)	(0.036)		(0.057)	(0.054)	(0.054)	(0.058)
Controls interacted with MF	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.99	−2.371	−2.238	−2.226	−2.995***	4.429	−10.47	−10.52	−10.47	−2.849***
	(2.478)	(3.510)	(3.533)	(3.541)	(0.551)	(4.309)	(8.198)	(8.157)	(8.167)	(1.022)
Observations	63,731	42,924	42,924	42,924	38,539	63,730	42,924	42,924	42,924	38,539
R2	0.151	0.016	0.017	0.017	0.018	0.011	0.018	0.018	0.018	0.019

Note. Table 2.5 reports the results of a regression analysis that replaces the management forecast surprises with good or bad news. The multivariate tests were repeated as described earlier. We controlled for year and industry effects. Standard errors were adjusted for heteroskedasticity and clustering by firm and quarter. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

2.4.4 Portfolio analysis: market reactions to management forecasts

To investigate the impact of MD&A textual factors on the market's underreaction to forecast news, we first verified the existence of an underreaction to management forecast news using a univariate test. We classified management forecasts into ten groups based on the surprise values each quarter, utilising quarterly and yearly announcements from 2001 to 2017. We calculated the average three-day announcement CAR (CAR [−1,1]) and the drift returns for various windows, including CAR[2,45], CAR[2,60] and CAR[2,m3], as presented in Table 2.6. CAR[2,m3] is the size- and B/M-adjusted buy-and-hold CAR, expressed as a percentage, across three months beginning on the second day after the management forecast date. We calculated MF10–1, the difference between the extreme surprise deciles (MF10 for good news and MF1 for bad news), using heteroskedasticity-adjusted standard errors for *p*-value calculations.

Table 2.6 gives the average abnormal returns across different time windows for the ten management forecast surprise groups. The difference in abnormal announcement day returns (CAR[−1,1]) between Deciles 10 and 1 (MF10–1) reflects the contemporaneous reaction of stock prices to the forecast news. A larger spread indicates a stronger investor reaction on the announcement date. Consistent with prior research (see A. P. Hutton et al., 2003), our findings show a stronger market reaction to bad news than good news, with a CAR[−1,1] of −5.77% for bad news and 5.26% for good news. Column 2 of Table 2.6 reveals a highly significant mean spread of 11% between firms with good news (MF10) and those with bad news (MF1) in the three-day CAR (CAR[−1,1]). For post–management forecast abnormal returns, the extremely good news decile generated a positive CAR[2,45] of 1.31%, with a CAR[2,60] of 1.34% and a CAR[2,3m]) of 1.24%. The extremely bad news decile generated a negative CAR[2,45] of

−0.46%, a CAR[2,60] of −0.75% and a CAR[2,3m] of −0.66%. These returns are significant both statistically and economically. The spread in post-abnormal returns between Deciles 10 and 1 of forecast surprise quantifies the extent of underreaction to management forecasts, as evidenced by the subsequent drift. Our findings reveal a significant spread of 1.769% in the average post-announcement CAR[2,45] between the highest deciles of good and bad management earnings forecast news. Table 2.6 also gives the post–management forecast CAR as CAR[2, 60] and CAR[2, m3]. Our analysis focused on the 45-day return drift following the management forecast announcement, as the hedge portfolio's abnormal returns were primarily observed during the first 45 days post-announcement, followed by a decline.[16]

Table 2.6: Portfolio analysis: CAR of management forecast surprises

Decile	CAR[−1,1]	CAR[2,45]	CAR[2,60]	CAR[2,m3]
1	−5.774***	−0.457*	−0.748**	−0.659**
2	−3.719***	0.193	−0.211	−0.087
3	−2.096***	−0.626***	−0.976***	−0.837***
4	−0.972***	−0.522***	−0.798***	−0.749***
5	−0.321***	−0.533***	−0.998***	−0.81***
6	0.233***	−0.56***	−0.762***	−0.613***
7	1.156***	−0.226.	−0.437**	−0.331*
8	2.304***	−0.133	−0.382*	−0.313*
9	3.632***	0.294.	0.185	0.216
10	5.259***	1.312***	1.336***	1.235***
MF10–1	11.032***	1.769***	2.084**	1.894***

Note. For each calendar quarter, we categorised forecasts into ten groups based on management forecast surprises. Using quarterly and yearly management forecast announcements from 2001 to 2017, we computed measures of abnormal returns, including the average three-day post-announcement cumulative abnormal returns (CAR[−1,1]) and the drift returns for different time periods (CAR[2,45], CAR[2,60] and CAR[2,m3]). CAR[2,T] is the size- and B/M-adjusted cumulative abnormal return, expressed as a percentage, within the [t + 2, t + T] window following the management forecast announcement. CAR[2,m3] is the three-month buy-and-hold size- and B/M-adjusted cumulative abnormal return, as a percentage, for the three months from two days after the management forecast date. MF10–1 is the difference between the extreme surprise deciles (MF10: good news, MF1: bad news). Values of p were calculated using standard errors adjusted for heteroskedasticity. *, **, *** indicate two-tailed statistical significance at the 10, 5, and 1% levels, respectively.

[16] We obtained similar results for various time windows within two months after the announcement.

We conducted an analysis to examine how linguistic factors in MD&A sections affect price reactions to management forecast news. For each quarter, we categorised forecasts into five groups based on the management forecast surprise and a further five based on MD&A textual variables, resulting in 25 distinct groups. We calculated the average CAR for the most positive (Q5) and negative (Q1) management forecast surprises in each MD&A decile in the announcement and post-announcement periods.

Table 2.7 presents the three-day CAR (CAR[−1,1]) for good (Q1) and bad (Q5) management forecasts, categorised by Uncertainty, Fog index, Word count, and Tone change across five quintiles of MD&A factors. We found that investors' three-day announcement reaction to the management forecast was significant across all quintiles of MD&A factors. Additionally, the magnitude of the spread between extremely good and bad news was slightly greater for Quintile 5 of all MD&A factors. This indicates that investors' reactions are slightly more sensitive and stronger to management forecasts characterised by high uncertainty (Uncertainty rank = 5), low readability (Fog and Word count rank = 5), and significant tone change (Tone change rank = 5) in MD&A disclosures. However, the spread difference between the top and bottom groups of MD&A factors is not statistically significant. The spread in the investors' 3-day announcement reaction across the quintiles of MD&A readability and uncertainty does not exhibit a strictly monotonic pattern.

Table 2.7: Portfolio analysis: double-sorted MD&A characteristics and management forecast surprises (CAR[−1,1])

MD&A Quintile	Uncertainty	Q1	Q5	Q5−Q1	*p*
1	1	−0.051	0.04	0.092	.00
2	2	−0.053	0.043	0.097	.00
3	3	−0.053	0.047	0.096	.00
4	4	−0.054	0.047	0.101	.00
5	5	−0.048	0.051	0.101	.00
Hedge	5−1	0.003	0.012	0.01	
	p	0.553	0.008	0.081	
	Fog index	Q1	Q5	Q5−Q1	*p*
1	1	−0.05	0.04	0.089	.00
2	2	−0.054	0.043	0.098	.00
3	3	−0.056	0.047	0.104	.00
4	4	−0.052	0.05	0.102	.00
5	5	−0.048	0.047	0.098	.00
Hedge	5−1	0	0.008	0.008	
	p	0.992	0.129	0.181	
	Word count	Q1	Q5	Q5−Q1	*p*
1	1	−0.05	0.04	0.09	.00
2	2	−0.053	0.043	0.101	.00
3	3	−0.057	0.047	0.104	.00
4	4	−0.052	0.05	0.099	.00
5	5	−0.047	0.048	0.097	.00
Hedge	5−1	0.004	0.009	0.008	
	p	0.332	0.1	0.215	
	Tone change	Q1	Q5	Q5−Q1	*p*
1	1	−0.05	0.04	0.091	.00
2	2	−0.054	0.045	0.099	.00
3	3	−0.048	0.046	0.096	.00
4	4	−0.056	0.047	0.106	.00
5	5	−0.047	0.047	0.097	.00
Hedge	5−1	0	0.005	0.006	
	p	0.937	0.174	0.256	

Note. We conducted a portfolio analysis to investigate the impact of MD&A linguistic factors on contemporaneous price reactions to management forecast news. For each calendar quarter, we sorted the forecasts into five groups based on the value of the management forecast surprise and independently sorted the forecasts into five groups based on the value of MD&A textual variables, obtaining 25 groups. For each row of the MD&A quintile, we calculated the average three-day cumulative abnormal returns CAR[−1,1] for the most positive (Q5) and the most negative (Q1) management forecast surprise deciles.

We found a positive association between higher uncertainty and complexity in the MD&A section and a stronger post-announcement drift. Table 2.8 demonstrates that the difference in average post-announcement abnormal returns (CAR[2,45]) between good and bad news forecasts is more pronounced for the upper quintiles of MD&A Uncertainty, Fog index and Word count. A significant and substantial spread in post-announcement abnormal returns was observed only for the highest quintiles of MD&A Uncertainty, Fog index and Word count. The spread in post-announcement abnormal returns across the quintiles of MD&A readability and uncertainty did not exhibit a strictly monotonic pattern. This lack of monotonicity could result from variations in firm characteristics, including size and B/M ratio, across the different quintiles of MD&A factors. In the preceding section of 2.4.2 and 2.4.3, we conducted a multivariate regression analysis to consider the impact of firm characteristics and calendar effects, including year and quarter, on the relationship between announcement or post-announcement abnormal returns and MD&A characteristics.

Table 2.8: Portfolio analysis: double-sorted MD&A characteristics and management forecast surprises (CAR[2,45])

MD&A Quintile	Uncertainty	Q1	Q5	Q5−Q1	*p*
1	1	0.005	0.007	−0.003	0.722
2	2	−0.009	0.006	0.015	0.023
3	3	−0.002	0.005	0.007	0.184
4	4	−0.002	0.011	0.014	0.026
5	5	−0.005	0.005	0.011	0.023
Hedge	5−1	−0.017	−0.001	0.019	
	p	0.033	0.858	0.02	
	Fog index	Q1	Q5	Q5−Q1	*p*
1	1	0.004	0.009	0.002	0.865
2	2	−0.005	0.004	0.012	0.045
3	3	0.001	0.009	0.008	0.306
4	4	−0.003	0.009	0.014	0.036
5	5	−0.009	0.001	0.01	0.041
Hedge	5−1	−0.018	−0.009	0.009	

MD&A Quintile	Uncertainty	Q1	Q5	Q5−Q1	*p*
	p	0.028	0.16	0.379	
	Word count	Q1	Q5	Q5−Q1	*p*
1	1	0.004	0.009	0.001	0.912
2	2	−0.005	0.004	0.014	0.022
3	3	−0.001	0.009	0.008	0.288
4	4	−0.002	0.008	0.014	0.04
5	5	−0.008	0.002	0.011	0.037
Hedge	5−1	−0.019	−0.009	0.015	
	p	0.029	0.179	0.118	
	Tone change	Q1	Q5	Q5−Q1	*p*
1	1	0.001	0.01	0.002	0.737
2	2	−0.008	0.002	0.011	0.067
3	3	−0.003	0.007	0.013	0.036
4	4	−0.002	0.015	0.017	0.048
5	5	0.000	0.000	0.003	0.614
Hedge	5−1	−0.01	−0.011	0.001	
	p	0.237	0.096	0.936	

Note. We conducted a portfolio analysis to examine the impact of MD&A linguistic factors on price underreactions to management forecast news. For each calendar quarter, we first sorted the forecasts into five groups based on the value of the management forecast surprise and independently sorted the forecasts into five groups based on the value of MD&A textual variables, obtaining 25 groups. For each row of the MD&A quintile, we calculated the average cumulative abnormal returns (CAR[2,45]) for the most positive (Q5) and most negative (Q1) management forecast surprise deciles.

2.4.5 Additional analysis: information acquisition & MD&A linguistic factors

Managers play a crucial role in reducing information asymmetry between firms and investors or analysts by providing earnings forecasts and enhancing the quality of voluntary disclosures. The disclosure quality can also influence investors' demand for information. Research by Drake et al. (2012) found a significant increase in Google searches before earnings announcements, followed by a surge after the announcement that maintained a high level for a period. Drake et al. (2015) emphasised the implications of acquiring information through EDGAR, as price discovery relies on investors accessing filings containing new information. Lawrence et al. (2012) found that user demand for sell-side analyst research spikes on the issuance of management guidance, earnings announcements and reports from All Star Analysts,

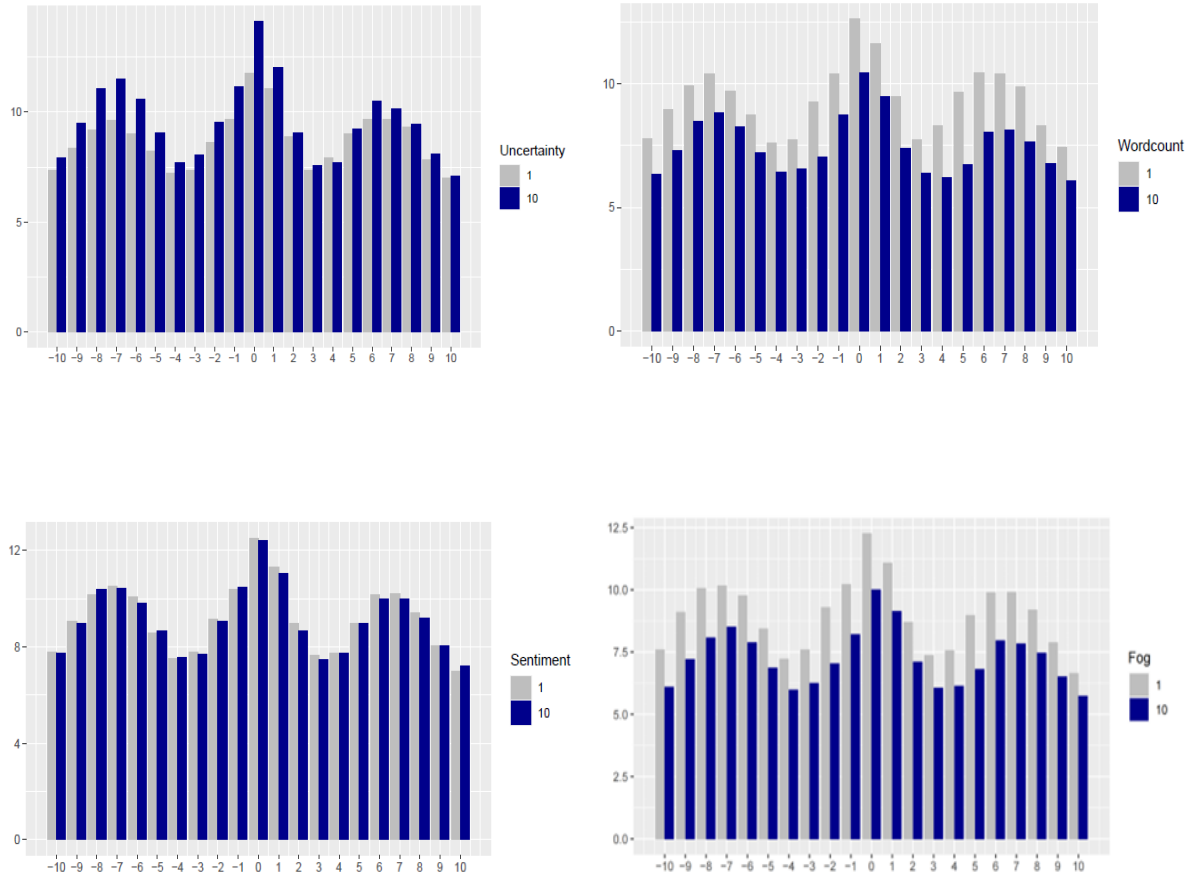
but this was not the case for Form 10-K and 10-Q filings. A number of studies (Bozanic et al., 2017; Drake et al., 2015; Loughran & McDonald, 2017; Ryan, 2017) have used EDGAR log files to investigate the demand for financial reporting information. The EDGAR log file dataset, available from the SEC, offers a direct measure of demand for a company's financial reports, but analysing the data is challenging because of the level of detail. For our analysis, we retained daily non-missing 10-K filings, calculated the daily volume of EDGAR downloads and identified firms with a Central Index Key.[17] We matched accession numbers with the master index file to identify specific forms.

Figure 2.2 shows the information acquisition activity around the management forecast day for the highest and lowest deciles of MD&A factors. The x-axis is centred on the management forecast date, $t = 0$, and spans ten days before and after that date. Panel A shows a spike on Day 0 for both group extremes, indicating that the management forecast announcement is associated with intensive online information-gathering about the firm. Because searching and utilising EDGAR information requires a certain level of financial knowledge, the EDGAR search volume can, to some extent, capture the attention of sophisticated investors. On average, firms with more readable MD&A disclosures (in the bottom decile of Fog index and Word count) in their 10-K reports had higher daily EDGAR search volumes than those with more complex disclosures (in the top decile of Fog index and Word count). Panel B presents the abnormal EDGAR search volume around the forecast day. For Firm *I* on Day *t*, we calculated the abnormal EDGAR search volume (AbEdgar) as the raw EDGAR search volume minus the average raw EDGAR search volume for the same day of

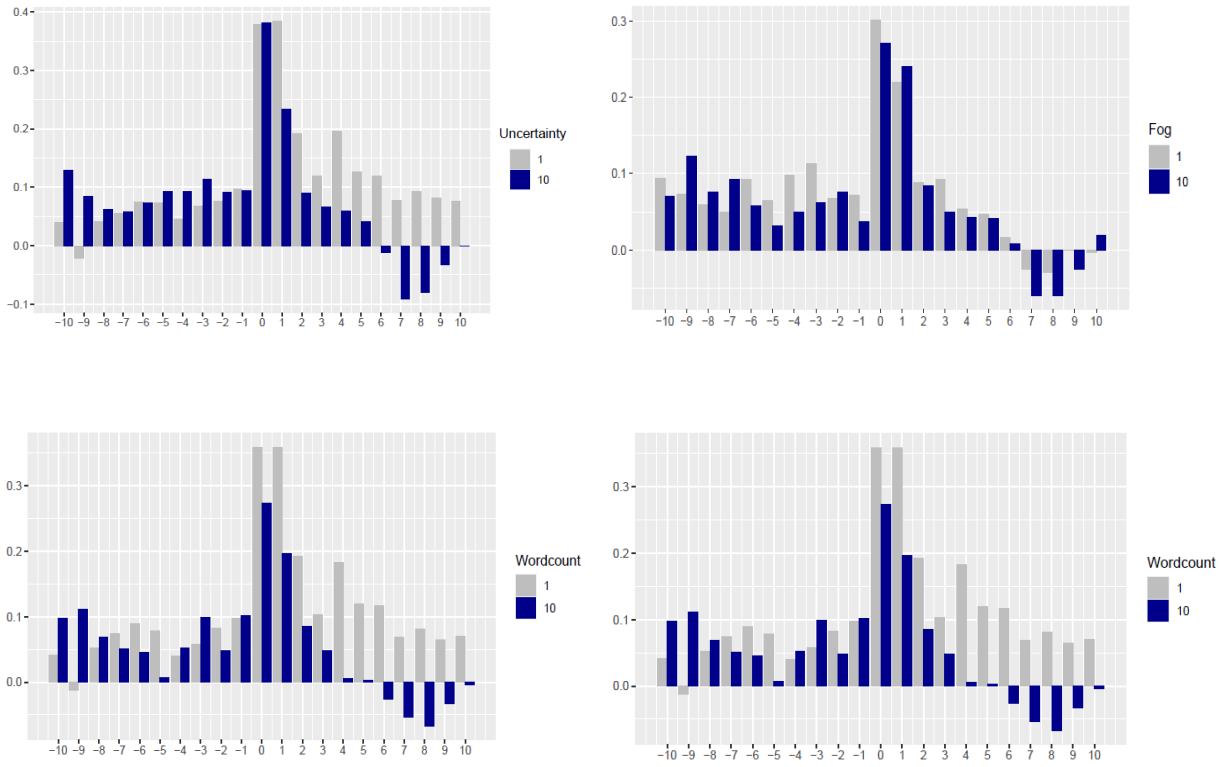
[17] The address ftp://ftp.sec.gov/edgar/full-index provides access to the EDGAR Master Index File.

Week *k* over the previous 10 weeks, scaled by the average raw EDGAR search for the same day of Week *k* over the previous ten weeks. Panel B shows a spike on Day 0 for both extremes of the MD&A group, with a significant difference in AbEdgar between the two extremes of the MD&A Word count group. Firms with shorter MD&A disclosures exhibited more abnormal EDGAR search around the event day; however, the difference between the MD&A groups sorted by Fog index, Uncertainty and Tone change is not obvious. A reverse effect is observed in the abnormal search pattern, where the abnormal EDGAR search volume decreases and becomes negative in the week following the announcement date, meaning that the search volume becomes lower than the average of the previous 10 weeks. This outcome may be partly a result of the unexpectedly high attention given to firms in the days surrounding the announcement. The information acquisition activity or information demand of investors for the same firm appears to decrease after one week.

Figure 2.2: Investor demand for the 10-K report relative to the management forecast date



Panel A: The daily raw EDGAR search index around the management forecast announcement day



Panel B: The daily abnormal EDGAR search volume around the management forecast announcement day

Note. Panel A of Figure 2.2 displays the average daily volume of EDGAR downloads grouped by the two extreme deciles of MD&A linguistic factors around the management forecast announcement day from $t - 10$ to $t + 10$. Panel B presents the abnormal EDGAR search volume around the forecast day. For Firm *I* on Day *t*. The figure is centred around the management forecast date, $t = 0$.

Our study employed cross-sectional regression analysis to investigate the determinants of EDGAR search activities around the management forecast day. Two different measures of EDGAR searches were used as dependent variables, with B/M, market equity (ME), turnover, LogAnalyst, IOR, cumulative return for 12 months (Cret_12m), volatility (Vola), litigation, the Herfindahl–Hirschman index, Altman Z-score and high technology (Hitech) included as additional control variables. The main aim of our analysis was to investigate the association between MD&A linguistic factors and investors' information acquisition behaviour around the management forecast date within a half-month event window, ranging from Day *t* to *t + 15*.

The dependent variables were the logarithm of AbEdgar + 1 (logAbEdgar) and the daily average EDGAR searching IPs (AvEdgar) over the event window. AbEdgar is the average abnormal EDGAR result, calculated as the average of daily abnormal searching Ips over the event window. The daily abnormal search volume was calculated by subtracting the raw EDGAR search volume on Day *t* from the average raw EDGAR search volume for the same day of Week *k* over the preceding 10 weeks. To normalise the distribution and obtain the average of the daily abnormal search volume over the half-month window, we followed Da et al. (2011) by using the logarithm of AbEdgar + 1. The daily abnormal EDGAR volume is useful for identifying unexpected investor attention on the 10-K report, whereas logAbEdgar captures deviations from a firm- and weekday-specific benchmark.

The regression results in Table 2.9 indicate the effects of MD&A factors and management forecast surprises on searches for a company's 10-K SEC filings in EDGAR in the post-announcement period. The table has two main panels. Columns 1 to 5 give the results of models with abnormal EDGAR search volumes (logAbEdgar) as the dependent variable, while Columns 6 to 10 give the results of models with average daily EDGAR search volume (AvEdgar) as the dependent variable. There is a significant negative correlation between D.MF, the decile rank of forecast surprise, and both the average daily EDGAR search and abnormal EDGAR search in the half-month period following the forecast announcement. No significant relationship was observed between MD&A linguistic factors and unexpected attention (logAbEdgar). The results suggest that MD&A linguistic factors in the period prior to the forecast announcement do not account for the unexpected attention. Further, MD&A complexity, as measured by Fog index and Word count, is significantly negatively related to

average daily EDGAR search volume (see Columns 3 and 4). Finally, both extremely bad and extremely good news are significantly positively associated with the average daily EDGAR search (see Column 4).

Table 2.9: Impact of MD&A factors and management forecast news on EDGAR search activities

	MD&A factor									
	logAbEdgar					AvEdgar				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variable	Main	Uncertainty	Fog index	Word count	Tone change	Main	Uncertainty	Fog index	Word count	Tone change
D.MF	−0.00372***	−0.00300	−0.0033*	−0.0036**	−0.0053**	−0.0583***	−0.112***	−0.116***	−0.119***	−0.107***
	(0.000911)	(0.00199)	(0.00178)	(0.00178)	(0.00207)	(0.0142)	(0.0333)	(0.0345)	(0.0345)	(0.0355)
D.MD&A		0.00275	0.000859	0.000606	0.000171		−0.0151	−0.0534*	−0.0531*	−0.0386
		(0.00262)	(0.00215)	(0.00215)	(0.00257)		(0.0300)	(0.0302)	(0.0303)	(0.0316)
D.MF×D.MD&A		−0.000153	−0.0000	−0.0003	0.000314		0.00963**	0.0102**	0.0106**	0.00834*
		(0.000439)	(0.00035)	(0.000354)	(0.000438)		(0.00441)	(0.00466)	(0.00467)	(0.00469)
Controls interacted with MF	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.0412*	0.0219	0.0203	0.0202	0.0377	0.371	−0.711*	−0.620*	−0.631*	−0.338
	(0.0223)	(0.0221)	(0.0222)	(0.0221)	(0.0243)	(0.360)	(0.371)	(0.375)	(0.375)	(0.402)
Observations	38,424	38,424	38,424	38,424	34,669	23,051	23,051	23,051	23,051	21,100
R^2	0.059	0.059	0.059	0.059	0.058	0.513	0.513	0.513	0.513	0.518

Note. Table 2.9 presents the findings regarding the influence of MD&A factors and management forecast news on EDGAR search activities around the management forecast day. The dependent variables examined are the logarithm of AbEdgar + 1 (logAbEdgar) and the daily average of EDGAR searching IPs (AvEdgar) throughout the event window [0,+15]. The daily abnormal EDGAR is useful for identifying unexpected investor attention on the 10-K report, whereas logAbEdgar captures deviations from a firm- and weekday-specific benchmark. Standard errors are given in parentheses and were calculated using two-dimensional clustering by calendar quarter and firm. Statistical significance is given by ***, ** and *, indicating the 0.01, 0.05 and 0.10 levels, respectively.

Table 2.10 presents the asymmetric effects of good and bad news separately for abnormal EDGAR search volume (logAbEdgar) and average daily EDGAR search volume (AvEdgar). Column 1 shows that bad news is significantly positively associated with the abnormal EDGAR search volume. Good news and bad news are both significantly positively associated with average daily search frequency (see Columns 5 to 10). In contrast to Table 2.9, where a negative relationship was observed between MD&A complexity (Fog index and Word count) and average search volume, Table 2.10 shows a significant positive association. The positive coefficients for Fog index and Word count suggest that as MD&A clarity decreases (that is, as complexity increases), investors engage in more EDGAR searches, probably to gather clarifying information and reduce uncertainty. This increased search behaviour can be a response to the cognitive demands of processing complex information, whereby investors seek to reduce uncertainty by looking for additional explanatory details or external analyses. However, the interaction terms between bad or good news and Fog index or Word count are negative. This indicates that the presence of bad or good forecast news dampens the effect of complexity in increasing the search volume. One interpretation is that when faced with bad news, investors might prioritise news content over a detailed analysis of complex MD&A disclosures. Alternatively, a combination of bad news and high complexity could lead to information overload, causing investors to reduce their search efforts because of the high cognitive cost of processing the information. For good news, the interaction with the Fog index or Word count was also negative (−0.139), meaning that the effect of complexity on search behaviour was reduced. After they receive good news, investors may be less motivated to seek additional information when the news itself is reassuring. Further, good news may reduce the

perceived urgency to analyse complex disclosures thoroughly, leading to a lower search volume than might be expected based on the complexity of the MD&A. Additionally, more optimistic and uncertain MD&A language can mitigate the positive influence of bad or good news on average daily searches. Our findings suggest that MD&A linguistic factors asymmetrically affect the marginal effects of forecast news on search behaviour. Overall, the results confirm that MD&A linguistic factors have direct and indirect effects on post-forecast EDGAR searches, although the precise mechanisms warrant further exploration.

Table 2.10: Impact of MD&A factors and good and bad forecast news on EDGAR search activities

	MD&A factor									
	logAbEdgar					AvEdgar				
Variable	(1) Main	(2) Uncertainty	(3) Fog index	(4) Word count	(5) Tone change	(6) Main	(7) Uncertainty	(8) Fog index	(9) Word count	(10) Tone change
Bad news	0.0266***	0.00857	0.0145	0.0151	0.0143	0.731***	1.859***	1.790***	1.802***	1.780***
	(0.00567)	(0.0117)	(0.0108)	(0.0108)	(0.0123)	(0.104)	(0.231)	(0.238)	(0.238)	(0.244)
Good news	0.00508	−0.0140	−0.00287	−0.00439	−0.0296**	0.366***	1.191***	1.116***	1.113***	1.104***
	(0.00631)	(0.0133)	(0.0119)	(0.0119)	(0.0139)	(0.105)	(0.249)	(0.250)	(0.250)	(0.265)
D.MD&A		−0.000452	−0.00087 9	−0.000841	−0.000868		0.141***	0.0979***	0.101***	0.101***
		(0.00152)	(0.00131)	(0.00131)	(0.00155)		(0.0247)	(0.0241)	(0.0242)	(0.0258)
Bad news ×D.MD&A		0.00416*	0.00262	0.00251	0.00294		−0.206***	−0.193***	−0.195***	−0.191***
		(0.00249)	(0.00210)	(0.00210)	(0.00245)		(0.0323)	(0.0330)	(0.0330)	(0.0341)
Good news ×D.MD&A		0.00451	0.00177	0.00210	0.00739**		−0.153***	−0.139***	−0.139***	−0.135***
		(0.00306)	(0.00244)	(0.00244)	(0.00307)		(0.0338)	(0.0351)	(0.0351)	(0.0355)
Controls interacted with MF	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.0156	0.0219	0.0203	0.0202	0.0377	−0.119	−0.711*	−0.620*	−0.631*	−0.338
	(0.0214)	(0.0221)	(0.0222)	(0.0221)	(0.0243)	(0.351)	(0.371)	(0.375)	(0.375)	(0.402)
Observations	38,424	38,424	38,424	38,424	34,669	23,051	23,051	23,051	23,051	21,100
R^2	0.059	0.059	0.059	0.059	0.058	0.513	0.513	0.513	0.513	0.518

Note. The table presents the findings regarding the influence of MD&A factors and good and bad forecast news on EDGAR search activities surrounding the management forecast day. The dependent variables examined are the logarithm of AbEdgar + 1 (logAbEdgar) and the daily average of EDGAR searching IPs (AvEdgar) during the event window [0,+15]. The daily abnormal EDGAR volume is useful for identifying unexpected investor attention on the 10-K report, whereas logAbEdgar captures deviations from a firm- and weekday-specific benchmark. Standard errors, given in parentheses, were calculated using two-dimensional clustering by calendar quarter and firm. Statistical significance is given by ***, ** and *, indicating the 0.01, 0.05 and 0.10 levels, respectively.

2.4.6 Additional analysis: MD&A linguistic factors, management forecast news & forecast precision

Managers often provide earnings forecasts motivated by a desire to reduce the information asymmetry between themselves and investors or analysts. However, there are instances when managers issue forecasts driven by personal incentives or self-interest. Some literature (Nagar et al., 2003) suggests that different levels of managerial incentives can affect forecasting behaviour. Ettredge and Zhang (2013) studied the influence of financial restatements on managers' subsequent earnings forecasts and found that companies issuing restatements were less likely to provide quarterly earnings forecasts, resulting in less precise and less optimistically biased post-restatement forecasts. Our study contributes to the literature on disclosures by exploring how financial restatements affect companies' voluntary forecast behaviour. Hassell and Hillison (2000) found that management tends to voluntarily provide causal attributions when the forecast news is negative, which indicated a link between voluntary disclosure and management's forecast characteristics. Attributions improve the credibility and precision of management forecasts. Clarkson et al. (1994) examined why some firms voluntarily include directional forecasts in their MD&A disclosures while others do not and found that the likelihood of forecasting increases with financing requirements for good-news firms but decreases with the threat of competitor entry. The annual report, particularly the MD&A section, is a potentially important disclosure channel for forecast information. We have proposed that managers' MD&A language may capture their forecast behaviour. Managers have significant incentives to control the MD&A disclosure quality for self-serving purposes, and this may affect their forecasting behaviour, for example, the forecast precision. We thus

anticipate that firms with shorter, more readable MD&A disclosures are likely to produce more accurate management forecasts in subsequent announcement periods.

We used the model in Equation 22 to examine the effect of management's disclosure quality on the forecast precision.

$$Precision = \beta_0 + \beta_1 D.MF + \beta_2 D.MD\&A + \beta_3 D.MD\&A \times D.MF$$
$$+\beta_4 Bad\ news(Good\ news) + \beta_5 D.MD\&A \times Bad\ news(Good\ news)$$
$$+\delta Control\ variables + \varepsilon \quad (22)$$

Our hypothesis, confirmed in our results, is that MD&A linguistic factors, including Word count and Fog index, are associated with the firm-specific managerial disclosure incentive that may play a role in the decision to issue earnings forecasts. In contrast, MD&A Tone change and Uncertainty were not significantly related to forecast precision. To measure forecast precision, we employed the negative value of its forecast width. The width of a range forecast is the magnitude of the interval, whereas for point forecasts, the forecast width is zero. Using the forecast width provides a quantifiable measure of forecast precision, with a higher negative value indicating a more precise forecast. The additional control variables used included earning persistence, forecast horizon, the B/M ratio, firm size, turnover, LogAnalyst, institutional ownership, momentum, return volatility, litigation dummy, the Herfindahl–Hirschman index, the Altman Z-score and the Hitech indicator. We expected the coefficient of D.MDA with precision to be significant if the MD&A linguistic factors are associated with the manager's forecast accuracy.

Panel A of Table 2.11 shows a statistically significant negative coefficient for the deciles of the fog index. This result implies that firms with higher complexity in their MD&A

disclosures are more prone to issuing inaccurate management forecasts. Panel B reveals a negative coefficient for the deciles of Word count that is significant at the 5% level, suggesting that firms with longer MD&A sections also have inaccurate management forecasts. However, Panels C and D do not show any significant association between the MD&A uncertainty and tone and management forecast precision. Our results emphasise the effects of readability in firm MD&A disclosures in influencing the precision of management forecasts. Specifically, we found that managers tend to provide more precise earnings forecasts in companies that have transparent and easily understandable MD&A disclosures. Specifically, the study indicates that managers tend to provide more precise earnings forecasts for companies that have transparent and easily understandable MD&A disclosures. Given the observed impact of MD&A disclosure quality on forecast accuracy, as highlighted in Panels A and B, it's intriguing to consider the broader strategic decisions made by managers. At first glance, it might seem suboptimal for managers to override private information production. However, high-quality management disclosures can alleviate the reliance on private information. As information providers, managers should be motivated to provide high-quality disclosures, as this aligns with their motivation to provide accurate management forecasts. More succinctly, information providers who prioritise clear disclosures are more likely to issue accurate forecasts.

This study has explored the relationship between forecast news type and precision. The coefficients for negative news are statistically significant and negatively associated with forecast precision in all panels of Table 2.11, indicating reduced precision for forecasts with highly negative news. The coefficients for positive news are statistically significant and negatively associated in all panels as well, indicating lower precision for forecasts involving

extreme positive news. These findings slightly diverge from Baginski and Hassell's (1997) study, which reported no discernible difference in precision between good and bad news forecasts. Our findings were that forecasts involving extremely positive or negative news had lower precision than neutral news forecasts. Previous studies have shown that accurate forecasts are associated with a significant decline in share prices. Therefore, if managers aim to minimise the adverse effects of bad news on prices, issuing inaccurate news forecasts becomes an attractive option. Our results suggest that managers may be motivated to issue imprecise good news forecasts to maximise their positive impact on stock prices. This becomes evident from the negative and statistically significant coefficients for good news. In Panels A and B, the interaction term for good news with readability deciles is significantly negative. These findings suggest that the marginally negative impact of good news on management forecast precision is influenced by the readability of the MD&A section. The tendency of managers to issue imprecise good news forecasts aligns with their inclination to produce less readable MD&A disclosures. Consistent with this interpretation, the findings in Panels A and B demonstrate a significant relationship between MD&A readability and the precision of management forecasts. The marginally negative impact of bad news on management forecast precision is affected by the uncertainty and tone of the MD&A section, as indicated by the significant negative interaction term in panel C and panel D.

To conclude, while managers have considerable discretion to shape management forecasts in the MD&A section, we lack a comprehensive understanding of how voluntary disclosures in this section influence forecast characteristics. To bridge this gap, our study investigated the relationship between linguistic factors in the MD&A and the precision of

management forecasts, providing evidence that managers strategically modify the forecast precision for self-serving reasons. Future research should further investigate the relationship between MD&A disclosure quality and managerial self-serving behaviors, exploring how these factors impact forecast accuracy and investor trust. Understanding these dynamics is essential for evaluating the credibility of management forecasts and ensuring market efficiency.

Table 2.11: MD&A factors, management forecast news and forecast precision

Panel	Model	Dependent variable: forecast precision							
		D.MD&A	D.MF	Bad news	Good news	D.MF ×D.MDA	Bad news ×D.MDA	Good news ×D.MDA	R
A	Fog index	−0.00802**	−0.00267	−0.0656***		0.00120**	−0.00113		0.057
		(0.00407)	(0.00279)	(0.0152)		(0.000574)	(0.00280)		
	Fog index	−0.0115**	0.0127***		−0.0571***	0.00204**		−0.00636*	0.058
		(0.00542)	(0.00412)		(0.0165)	(0.000873)		(0.00330)	
B	Word count	−0.00871**	−0.00307	−0.0600***		0.00126**	−0.00218		0.057
		(0.00412)	(0.00279)	(0.0156)		(0.000577)	(0.00304)		
	Word count	−0.0139**	0.0107**		−0.0493***	0.00240***		−0.00778**	0.058
		(0.00560)	(0.00416)		(0.0166)	(0.000895)		(0.00334)	
C	Uncertainty	−0.000422	0.00392	−0.0447***		0.0000	−0.00501*		0.057
		(0.00346)	(0.00276)	(0.0145)		(0.0005)	(0.00298)		
	Uncertainty	−0.00771**	0.0170***		−0.0611***	0.00127**		−0.00567**	0.056
		(0.00382)	(0.00414)		(0.0171)	(0.0006)		(0.00279)	
D	Tone change	0.00392	0.00857***	−0.0270*		−0.000652	−0.00697**		0.056
		(0.00375)	(0.00321)	(0.0157)		(0.0005)	(0.00339)		
	Tone change	−0.00501	0.0198***		−0.0674***	0.000740		−0.00423	0.057
		(0.00403)	(0.00439)		(0.0179)	(0.0007)		(0.00289)	

Note. Table 2.11 reports the regression analysis of the impact of MD&A and management forecast news on forecast precision. The table presents the coefficients from the regression of the management forecast precision. In each panel, D.MD&A acts as an independent variable encompassing four different linguistic factors, including fog index, word count, uncertainty, and tone change. Our study controlled for year and industry effects. The industry classification was based on the Fama–French 49-industry framework. Standard errors were adjusted for heteroskedasticity and clustered by the firm and quarter. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

2.5 Robustness checks

To mitigate potential effects on market responses to management forecasts of latent variables linked to managers' overall attitudes or the linguistic style of the management team, we controlled for linguistic factors extracted from two other sections of the 10-K report: 'Quantitative and Qualitative Disclosures about Market Risk' and 'Risk Factors'. 'Quantitative and Qualitative Disclosures about Market Risk', typically part of Item 7A, details the company's exposure to various market risks, such as interest rate, commodity price and foreign currency exchange rate risks. It contains quantitative measures, including sensitivity analyses and market risk measurements, and qualitative information on the company's risk management policies and procedures. 'Risk Factors', part of Item 1A, discusses factors that may affect the company's business, including industry-specific, regulatory and general economic risks. This section aims to provide investors with a comprehensive understanding of the risks involved in investing in the company, thereby enabling well-informed investment decisions. We extracted linguistic factors from these two sections using the same methodology as for the MD&A sections to obtain the linguistic variables: Uncertainty, Fog, Word count and Tone change. The variable D.RF gives the decile ranks of linguistic factors in the 'Risk Factors' section, and the variable D.QQDMR gives those in the 'Quantitative and Qualitative Disclosures about Market Risk' section.

Table 2.12 gives the results of a placebo regression analysis for the short-term market reaction. The interaction terms between D.RF and D.MF, and D.QQDMR and D.MF, are non-significant. However, the coefficient of D.MD&A and D.MF is significant and robust after the introduction of placebo variables. Table 2.13 gives the post-announcement abnormal return and confirms the non-significance of the coefficients of the interaction terms between D.RF and D.MF and D.QQDMR and D.MF. The interaction effect of D.MD&A with abnormal return remained significant. These results

suggest that the effect of MD&A linguistic factors does not arise from the unobserved linguistic styles of managers or the overall linguistic styles of management teams who fill out the 10-K form. If the MD&A linguistic factors were attributable to the overall linguistic style of managers, similar effects would be expected for the linguistic factors extracted from other sections. However, the placebo regressions did not reveal any significant effects of the placebo variables on either short-term market reactions or market underreactions. This placebo regression analysis reinforces the robustness and consistency of our findings.

Table 2.12: Placebo regression: contemporaneous market reactions to management forecasts

	MD&A factor							
	CAR[−1,1]				CAR[−1,1]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Uncertainty	Fog index	Word count	Tone change	Uncertainty	Fog index	Word count	Tone change
D.MF	0.353***	0.373***	0.373***	0.323***	0.363***	0.345***	0.351***	0.315***
	(0.1080)	(0.1080)	(0.1070)	(0.1150)	(0.1070)	(0.1080)	(0.1070)	(0.1150)
D.RF	0.00129	0.00343	−0.00854	0.0111				
	(0.0294)	(0.0296)	(0.0310)	(0.0341)				
D.RF×D.MF (placebo 1)	0.000825	−0.00377	−0.000998	−0.00246				
	(0.0046)	(0.0046)	(0.0047)	(0.0053)				
D.QQDMR					0.0143	−0.0369	−0.0506*	0.0194
					(0.0305)	(0.0277)	(0.0281)	(0.0365)
D.QQDMR					−0.00232	0.00511	0.00652	0.000626
×D.MF (placebo 2)								
					(0.0047)	(0.0044)	(0.0044)	(0.0057)
D.MD&A	−0.155***	−0.146***	−0.131***	−0.165***	−0.156***	−0.139***	−0.123***	−0.165***
	(0.0381)	(0.0358)	(0.0360)	(0.0388)	(0.0381)	(0.0361)	(0.0363)	(0.0388)
D.MD&A*D.MF	0.0340***	0.0289***	0.0273***	0.0322***	0.0342***	0.0277***	0.0259***	0.0321***
	(0.0059)	(0.0056)	(0.0056)	(0.0060)	(0.0059)	(0.0056)	(0.0057)	(0.0060)
Controls interacted with MF	YES	YES	YES	YES	YES	YES	YES	YES
Constant	−5.576	−5.496	−5.539	−5.770***	−5.64	−5.284	−5.316	−5.822***
	(3.5850)	(3.6230)	(3.6380)	(0.8260)	(3.5830)	(3.6320)	(3.6450)	(0.8230)
Observations	42,929	42,929	42,929	38,552	42,929	42,929	42,929	38,552
R^2	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201

Note. We used the linguistic factors extracted from two other sections of the 10-K report: 'Quantitative and Qualitative Disclosures about Market Risk' and 'Risk Factors', in the regression analyses. We used the same methodology as for the MD&A section to construct the linguistic variables, Uncertainty, Fog index, Word count and Tone change. D.RF represents the decile ranks of linguistic factors in the 'Risk Factors' section, while D.QQDMR represents the decile ranks of linguistic factors in the 'Quantitative and Qualitative Disclosures about Market Risk' section. The values of CAR[−1,1] are given as percentages. Standard errors were adjusted for heteroskedasticity and clustered by firm and quarter. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 2.13: Placebo regression: management forecasts and post-announcement returns

	MD&A factor							
	CAR[2,45]				CAR[2,45]			
Variable	(1) Uncertainty	(2) Fog index	(3) Word count	(4) Tone change	(5) Uncertainty	(6) Fog index	(7) Word count	(8) Tone change
D.MF	0.298*	0.308*	0.305*	0.323*	0.336*	0.316*	0.319*	0.349*
	(0.1790)	(0.1810)	(0.1800)	(0.1900)	(0.1770)	(0.1830)	(0.1820)	(0.1900)
D.RF	−0.0868*	−0.0539	−0.0674	−0.0457				
	(0.0466)	(0.0476)	(0.0488)	(0.0532)				
D.RF×D.MF (placebo 1)	0.0108	0.00669	0.0086	0.00782				
	(0.0074)	(0.0075)	(0.0076)	(0.0084)				
D.QQDMR					0.00576	0.00216	0.0102	0.0393
					(0.0476)	(0.0456)	(0.0454)	(0.0539)
D.QQDMR ×D.MF (placebo 2)					−0.000155	0.00277	0.00129	−0.00112
					(0.0076)	(0.0073)	(0.0073)	(0.0086)
D.MD&A	−0.138**	−0.162***	−0.163***	−0.0395	−0.142**	−0.170***	−0.177***	−0.0428
	(0.0588)	(0.0564)	(0.0571)	(0.0602)	(0.0587)	(0.0566)	(0.0572)	(0.0603)
D.MD&A*D.MF	0.0226**	0.0217**	0.0219**	0.00493	0.0232**	0.0222**	0.0233***	0.00518
	(0.0092)	(0.0089)	(0.0090)	(0.0094)	(0.0092)	(0.0089)	(0.0090)	(0.0094)
Controls interacted with MF	YES	YES	YES	YES	YES	YES	YES	YES
Constant	−13.36	−13.26	−13.27	−5.052***	−13.63	−13.35	−13.38	−5.325***
	(8.692)	(8.71)	(8.699)	(1.397)	(8.688)	(8.709)	(8.688)	(1.401)
Observations	42,924	42,924	42,924	38,539	42,924	42,924	42,924	38,539
R^2	0.017	0.017	0.017	0.018	0.016	0.017	0.017	0.018

Note. We used the linguistic factors extracted from two other sections of the 10-K report: 'Quantitative and Qualitative Disclosures about Market Risk' and 'Risk Factors', in the regression analyses. We used the same methodology as for the MD&A section to construct the linguistic variables, Uncertainty, Fog index, Word count and Tone change. D.RF represents the decile ranks of linguistic factors in the 'Risk Factors' section, while D.QQDMR represents the decile ranks of linguistic factors in the 'Quantitative and Qualitative Disclosures about Market Risk' section. The values of CAR[−1,1] are given as percentages. Standard errors were adjusted for heteroskedasticity and clustered by firm and quarter. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

2.6 Conclusion

The objective of this study was to explore the influence of linguistic factors in the MD&A on investors' responses to management forecasts. We have analysed how these linguistic factors affect the manager–investor relationship and the subsequent market response. In this research, MD&A was deemed informative when its qualitative content affected the market's reaction to forecast news. In most prior studies, the quality of MD&A disclosures (or other financial disclosures like SEC filings or press releases) is analyzed in terms of how it affects the price reaction to the information conveyed in the disclosure itself. In my study, however, the focus is different. I investigate the price reaction to management forecasts, which are issued separately and later than the MD&A disclosures. This creates a unique dynamic where the MD&A disclosures serve as background or contextual information that investors might use to interpret the subsequent management forecasts.

While MD&A disclosures and management forecasts are different types of disclosures, they are related in that both provide insights into managers' views on the company's performance. The MD&A disclosure, issued earlier, often sets the stage for the subsequent management forecast by outlining risks, operational challenges, and strategic direction. If the MD&A is unclear or uncertain, investors may struggle to form accurate expectations about future performance, making the management forecast, which comes later, more surprising and impactful. This dynamic helps explain the more pronounced price reactions observed in my results.

The study found that firms with more uncertain, less readable and more optimistic MD&A disclosures tended to elicit a stronger contemporaneous market reaction to management

forecasts. When MD&A disclosures are unclear or filled with uncertainty, investors are left with unresolved questions about the firm's prospects. They may find it challenging to extract clear insights from the MD&A, leading to greater uncertainty about the firm's near-term outlook. Consequently, when a management forecast is later issued, it serves to resolve some of this uncertainty and provide new information that was not apparent from the MD&A. The less readable and more uncertain the MD&A disclosure, the greater the level of uncertainty among investors, and therefore, the more surprising and impactful the management forecast is when it arrives. This explains the more pronounced immediate price reactions observed for companies with less readable and more uncertain MD&A disclosures. In essence, the forecast acts as a clarifying event for investors, and its impact is magnified when the earlier MD&A disclosure failed to provide clear guidance. This behavior is consistent with the concept of information asymmetry, where the less information investors have leading up to a forecast, the more sensitive they are to new information that fills those gaps.

Our study further finds that firms with more uncertain and less readable MD&A disclosures also experienced stronger post-announcement abnormal returns. The coexistence of both immediate and delayed price reactions can be understood through the lens of information processing costs. When MD&A disclosures are complex and uncertain, investors may not be able to fully interpret or process all the relevant information at the time of the forecast. Although the initial forecast resolves some of the uncertainty, the complexity of the prior MD&A disclosure increases the time and effort required to fully digest the implications of both the MD&A and the forecast. This delayed digestion of information can lead to an underreaction in

the short term, followed by a gradual correction over time, as more investors come to a clearer understanding of the information.

Other studies have indicated that managers strategically choose forecast attributes, including forecast precision, to affect market reactions and boost trading profits. This study has contributed to the understanding of management forecasting by analysing the impacts of managers' disclosure quality on the market's response to their forecasts. We have also explored the relationship between management forecast precision and linguistic factors in the MD&A. Our results indicate a positive association between MD&A readability and management forecast precision. This outcome suggests that managers who provide clear and readable disclosures are more likely to issue accurate forecasts, which positions them as reliable information providers. To express this differently, managers' efforts to provide clearer and more easily understandable management disclosures align with their efforts to enhance forecast accuracy and alleviate the information asymmetry between managers and investors.

Lastly, we have examined the relationship between MD&A disclosure factors, management forecast news and investors' information acquisition activities during the management forecast period. Our findings indicate that the complexity of the MD&A, as measured by the fog index, is significantly negatively associated with the average raw daily EDGAR searches. In parallel, both highly positive and highly negative news are significantly and positively correlated with the average daily EDGAR searches. However, we did not observe a significant association between MD&A linguistic factors and unexpected information-gathering activities (abnormal EDGAR searches). Although we found that the MD&A linguistic factors can influence the marginal impact of forecast news on investors' information

activities, the mediating impact was ambiguous and depended on the type of news (positive v. negative) and the type of attention (unexpected v. raw).

To conclude, this study has contributed to the literature about management forecasts by providing insights into the relationship between managers' disclosure quality, MD&A linguistic factors, management forecast characteristics and market reactions. The voluntary disclosure of information is closely linked to managerial incentive, as research has shown that different incentive can influence managerial forecasting behaviour and market outcomes. Our research has confirmed that the language style in the MDMA can influence investors' perceptions and responses to management forecasts. It will be of interest to investigate further how investors perceive and incorporate MD&A disclosures into their expectations of a firm's future earnings prospects. If this relation is not fully understood and adjusted for by the market, a trading opportunity arises. Further investigation of this question is needed.

Chapter 3: How global sentiment affects A–H cross-listed share mispricing in China: mechanisms of global sentiment contagion in a segmented market

3.1 Introduction

The principle known as the law of one price contends that identical assets should have similar prices across different markets. However, empirical evidence often contradicts this theory, particularly in the context of cross-listed shares. Notably, while such shares generally command a price premium in most markets, they tend to trade at a discount in China. This deviation prompts a reevaluation of the law of one price and has inspired several scholarly inquiries into its underlying causes.

Numerous studies have tackled the price disparity observed in cross-listed shares. Information asymmetry is frequently cited, with works by Brennan and Cao (1997) and Stulz and Wasserfallen (1995) highlighting the significant discrepancies in information access between foreign and domestic investors. These studies propose that foreign investors' lack of information about domestic assets leads them to perceive higher risk, thereby demanding a corresponding premium. Chakravarty et al. (1998) further this discussion by pointing out the challenges foreign investors face due to language barriers and divergent accounting standards, which contribute to the risk premium demanded in Chinese markets. Furthermore, Cheng (2000) underscores the different modes of accessing information: foreign investors typically rely on international news services like Reuters, whereas local investors often use personal networks and informal sources, which impacts the speed and quality of information they receive. This disparity influences the observed differences in market returns and volatility.

Chen et al. (2001) suggest that the price differences between A-shares and B-shares in China can also be attributed to market liquidity, supporting Amihud and Mendelson's (1986) theory that less liquid stocks warrant higher returns due to their increased transaction costs. Further studies by Bailey and Jagtiani (1994) and Ma (1996) explore how differing risk perceptions among investors contribute to price variations, particularly in cross-listed stocks.

Building on this foundational research, our study introduces a global sentiment index to investigate its impact on the pricing discrepancies between A-shares and their H-share counterparts. A key contribution of this study is the examination of the global sentiment transmission mechanism between A and H markets. While earlier studies have identified global sentiment as a potential driver of price dynamics in segmented markets, the specific channels through which global sentiment influences A-H share prices remain underexplored. This research addresses this gap by analyzing how global sentiment affects capital flows between these markets and how intermediary factors, such as relative arbitrage risk, moderate this effect. We examine the role of global sentiment in the context of Chinese policy reforms aimed at liberalizing capital flows, such as the Qualified Foreign Institutional Investor (QFII), Renminbi-Qualified Foreign Institutional Investor (RQFII), and Shanghai-Hong Kong Stock Connect (SHSC) programs. Our research seeks to determine whether shifts in global sentiment directly influence market prices or if their effects are mediated through indirect channels such as policy reforms and investor behavior.

In this study, we construct sentiment indices using Principal Component Analysis (PCA), guided by the methodologies of Baker and Wurgler (2006) and Stambaugh et al. (2015), utilizing sentiment data from six developed countries. Unlike previous studies, we uniquely

develop separate sentiment indices for Hong Kong and Mainland China, which are excluded from the global sentiment index construction. Additionally, we orthogonalize these local sentiment indices against the global sentiment to distinctly isolate and examine the impact of local sentiment dynamics from global influences.

This research also investigates the complex interplay between global sentiment, regulatory policies, and the A–H premiums, focusing particularly on how institutional reforms and market openness influence these dynamics. Employing a robust panel regression framework with fixed effects, we analyze the extent to which global sentiment, mediated by policy changes such as QFII, RQFII, and SHSC, affects the disparities in cross-listed stock prices[18]. The study enhances the model's robustness by using White's heteroskedasticity-robust standard errors to address potential serial correlations and heteroskedasticity within the panel data.

Furthermore, we consider the role of cross-border capital flows as a mechanism through which global sentiment impacts the A–H premium. By examining capital inflow ratios and their interaction with global sentiment, we ascertain the extent to which increased market liberalization influences price disparities. The findings suggest that the RQFII policy significantly moderates the impact of global sentiment on A–H premiums, especially in the post-implementation phase. This analysis confirms the pivotal role of financial reforms in aligning domestic markets with global financial trends and underscores the effectiveness of these policies in fostering market integration and reducing mispricing due to global sentiment.

[18] As Attanasio et al. (2000) have shown, fixed-effect estimators tend to be more accurate than GMM and IV estimators when the number of time-series observations per micro-unit is greater than 30.

Additionally, the study assesses the role of arbitrage risks in amplifying the effects of global sentiment. We consider several factors—dividend yield, bid-ask spread differences, market capitalization, and short-sale eligibility—to determine their impact on the A–H premium. Specifically, Results find that firms with high H-share trading costs and those where A-shares are ineligible for short selling exhibit greater vulnerability to global sentiment shifts. Conversely, larger firms with lower levels of information asymmetry, as indicated by their market capitalization, are less affected by these global sentiment fluctuations.

Lastly, we utilize time-series and Fama–French factor regressions to analyze how global sentiment affects A–H premiums and returns. Our empirical approach, which builds on the methodology of Bekaert et al. (2009), uses pooled regression analysis to examine how sentiment predicts stock returns in markets characterized by distinct regulatory changes and foreign institutional holdings. This dual approach enables a comprehensive analysis of how both global economic trends and specific local market factors contribute to pricing disparities in cross-listed stocks. The inclusion of local and global Fama–French factors allows us to control for varying risk profiles that might affect the stock returns in both Chinese and Hong Kong markets. This methodology is crucial for isolating the effect of global sentiment from other confounding market dynamics, providing a clearer picture of sentiment's direct impact on financial markets. Results reveal that global sentiment significantly predicts H-share excess returns but not A-share returns, indicating market-specific sensitivities. Moreover, the introduction of policy dummies—QFII, RQFII, and SHSC—allows us to assess their moderating effects on these relationships. Notably, the RQFII and SHSC policies amplify the responsiveness of A-share returns to global sentiment, suggesting effective market integration

facilitated by these policies. In contrast, the QFII policy exhibits a negative effect, highlighting the complexities of policy impacts on financial markets.

This chapter contributes to a nuanced understanding of how international sentiments and policy dynamics interact to shape market behaviours, offering valuable insights for policymakers and investors in international finance.

The remaining sections of this chapter are structured as follows. Section 3.2 offers the background information. Section 3.3 provides a literature review. Section 3.4 describes the data, variables and research methodology. Section 3.5 presents the empirical findings. Section 3.6 closes the chapter.

3.2 Background of A–H cross-listed shares

The Shanghai Stock Exchange was inaugurated on 26 November 1990, and the Shenzhen Stock Exchange was established on 11 April 1991. The origins of the Hong Kong Stock Exchange (SEHK) go back to 1891. There are two classes of shares issued by Chinese firms and listed in China. A shares are traded on the Shanghai and Shenzhen stock exchanges in Renminbi and are primarily available to domestic investors, though some foreign access is allowed via the Qualified Foreign Institutional Investor (QFII) programs. B shares, also listed on these exchanges, are traded in foreign currencies (U.S. dollars on the Shanghai Stock Exchange and Hong Kong dollars on the Shenzhen Stock Exchange) and are accessible to both domestic and foreign investors. Chinese companies can cross-list their shares overseas to obtain finance, with SEHK a preferred choice because of its proximity. Cross-listed shares listed in Hong Kong are known as H-shares, and many are in state-owned enterprises. A- and H-share cross-listed Chinese firms enjoy advantages such as synchronised trading hours, consistent

accounting standards and transparent disclosures. Trading time synchronicity reduces opportunities for arbitrage and promotes market efficiency. Consistent accounting standards and transparent disclosures minimise price deviations between the two markets. Consistent accounting standards also increase transparency for investors. These advantages play a vital role in managing external factors that could drive price disparities between the two markets. Cross-listing creates barrier-free investment opportunities in Chinese firms for Hong Kong and foreign investors. SEHK has more rigorous listing and information disclosure requirements compared with mainland Chinese stock exchanges. Over 40% of SEHK trading comes from overseas institutional investors. Capital controls imposed by mainland China generate a segmentation of the Hong Kong and mainland China stock markets, restricting the free movement of capital from both local and foreign investors. The A-share market in mainland China is dominated by domestic retail investors. Foreign investors faced strict restrictions before the launch of the QFII program in 2003. From 2011, the RQFII program allowed direct participation in mainland China's markets for foreign investors. The SHHK Connect program, launched in 2014, further facilitated cross-border investments by enabling stock trading by local brokers and clearing houses in each market.

3.3 Literature review

The literature shows that country sentiment can affect cross-listed share disparity. Arquette et al. (2008) found that changes in exchange rate expectations, market-specific sentiment and company-specific sentiment influenced the discounting of cross-listed shares and American depositary receipts (ADRs) in China. Grossmann et al. (2007) showed that the premiums on ADRs moved in the opposite direction to those on country closed-end funds in

response to consumer sentiment differences between the US and the ADRs' country of origin. Both the ADR price and the underlying asset price were more influenced by US sentiment than sentiment in the ADRs' country of origin. Zweig (1973) suggested that the mispricing of closed-end funds can be explained by differences in sentiment between investors holding the fund's shares and those holding the underlying assets. Bodurtha et al. (1995) found that the premiums of foreign country closed-end funds moved in the same direction as their stock prices and were influenced by the changing sentiments of US investors. An index of country fund premiums reflects the difference in sentiment between US investors and their foreign counterparts.

Cross-border sentiment contagion is a common phenomenon in developed countries. Hwang (2011) developed a measure of country-specific sentiment based on Gallup polls of American attitudes towards foreign countries. Hwang's study found that country-specific sentiment significantly affected discounts and premiums in closed-end country funds and ADRs and fund flows to foreign mutual funds. Sentiment also influenced cross-border merger and acquisition activity, suggesting that attitudes towards a country affected investment decisions abroad. Baker et al. (2012) introduced 'global sentiment' as a contrarian predictor of country-level market returns. The authors hypothesised that the flow of highly liquid capital among developed countries might facilitate the spread of global sentiment. Recently, Yawen and Christopher (2015) investigated the spillover of sentiment across countries. They found that investor sentiment significantly predicted the stock market performance in other countries, with a noticeable impact on international financial markets. In recent decades, market liberalisation has made it possible for investors from developed countries to invest in emerging markets. These markets are characterised by a low degree of capital liberalisation and high levels of

investment restrictions for foreign investors, which sets them apart. Most research has assumed that the capital flowing into emerging stock markets can lower the domestic cost of capital without any adverse effects. However, some work shows that foreign investors can be driven by impulsivity rather than rational thinking; this may represent the dark side of foreign capital flow. Bailey and Jagtiani (1994) studied the effect of barriers to capital flow into emerging markets and identified different risk exposures for local and foreign returns, suggesting capital market segmentation. In this context, local and cross-listed shares may react differently to global risks because of differing risk exposures. Chi et al. (2012) found that investor sentiment, as measured by mutual fund flows, positively affected stock returns in China, a result that differs from findings in developed markets. The discrepancy may be attributable to China's status as an emerging capital market in which sentiment may have a greater influence on stock returns. Han and Li (2017) challenged the idea that sentiment acts primarily as a contrarian predictor of market returns across different time frames. Their study found that local sentiment acted as a momentum signal on a monthly basis and improved portfolio performance in China.

3.4 Data, variable & research method

This study examined the effect of global and local sentiment on the price disparity of A–H cross-listed shares, using China's A–H cross-listed shares as the research target. The sample covered the period from 2000 to 2016 and included 80 matched pairs with A–H listing information for more than 12 months. Data for A–H cross-listed firms and stock information were obtained from CSMAR for A-stocks and Datastream for cross-listed H-stocks. The QFII investment ownership data for individual A-stocks was sourced from CSMAR's institutional

investor database, while A-share short sale eligibility information was retrieved from the Resset database.

3.4.1 Construction of a sentiment index

The global sentiment index was determined as the principal component of aggregated sentiment indices for six developed countries: the US, the UK, Japan, France, Germany and Canada. Subject to data availability, country-specific sentiment indices were constructed using sentiment proxies as defined by Baker and Wurgler (2006), which include four established market sentiment measures and consumer confidence as an additional indicator. To ensure cross-country comparability, we selected appropriate proxies for each country to build a consistent sentiment index. The lnipo proxy is obtained as the natural logarithm of monthly IPOs. Data for the US and Japan were sourced from Baker and Wurgler's website and Kaneko and Pettway's Japanese IPO database, respectively. Data for the UK were sourced from the London Stock Exchange, encompassing the Main Market, the Alternative Investment Market and the International Main Market. IPO data for China were sourced from the CSMAR China Initial Public Offering Research Database, and closed-end fund discount data were from the CSMAR funds database. For the other countries, IPO data were collected from the Securities Data Company (SDC) database. The dp proxy, as defined by Baker and Wurgler (2004a, 2004b), captures the dividend premium by comparing the average market-to-book-value ratios of dividend-paying and non-dividend-paying entities. Data for dividend premium was acquired from Datastream. A third proxy, equityratio, measures the equity share of total equity and debt issuances by all corporations in a country, calculated as the gross equity issuance (including IPOs and seasonal equity issues) divided by the sum of gross equity and gross long-term debt

issuances each month. Data for equityratio was collected from the SDC database. A fourth proxy, the OECD consumer confidence index, was used to measure confidence. A fifth proxy, CEFD, was used for the China sentiment index, depending on data availability. The final proxy, PVOL, represents the volatility premium. It is calculated as the natural logarithm of the month-end ratio of high-volatility firms' value-weighted average market-to-book ratio to that of low-volatility firms. High- and low-volatility groups were determined as the top and bottom three deciles, respectively, based on the variance of the previous month's daily returns. The decile breakpoints were determined on a country-specific basis.[19]

Each of the sentiment proxies was orthogonalised to the macroeconomic variables in the relevant country. The macroeconomic variables included the employment growth rate (Emgrowth),[20] consumption growth rate (Consumgrowth), consumer price index, industrial production growth rate (Industgrowth), three-month treasury rate (TreasuryBill3) and the term premium (TermPrem). The term premium was measured as the difference between the 10-year and three-month treasury bills in each country. Additional variables included the exchange risk (Curisk), measured as the monthly annualised 21-day rolling standard deviation of the US–China exchange rate, and the monthly percentage change in the Hong Kong–China exchange rate (DeltaExHK).[21]

We estimated the regression loadings for the sentiment proxies using the first principal component of each macro-orthogonalised proxy. The resulting total sentiment indices are linear

[19] Please refer to Appendix C for detailed information about the data sources for all sentiment proxies for each country.
[20] Because of a lack of availability of employment data for all countries, to ensure data consistency, we used the unemployment rate to calculate the employment rate for each country.
[21] The data sources used to construct the macro variables are given in Appendix C.

combinations of standardised proxy values for each country, with a mean value of zero. The fraction of variance explained by the first principal component was 43.86%, 42.70%, 36%, 31.55%, 41.90%, 48.61%, 35%, and 37.28% for the US, UK, Japan, France, Canada, Germany, China and Hong Kong.[22]

Country sentiment:

$$Sentiment_{US,t}^{total} = 0.603 * lnipo_t + 0.516 * equityratio_t - 0.427 * DP_t + 0.433 * VOLP_t \quad (23)$$

$$Sentiment_{UK,t}^{total} = 0.612 * lnipo_t + 0.547 * equityratio_t - 0.458 * DP_t - 0.342 * VOLP_t \quad (24)$$

$$Sentiment_{Japan.t}^{total} = 0.558 * lnipo_t + 0.33 * equityratio_t - 0.516 * DP_t + 0.557 * VOLP_t \quad (25)$$

$$Sentiment_{France,t}^{total} = 0.588 * lnipo_t + 0.407 * equityratio_t + 0.175 * DP_t - 0.677 * VOLP_t \quad (26)$$

$$Sentiment_{\text{Canada,t}}^{total} = 0.367 * lnipo_t + 0.635 * equityratio_t + 0.647 * DP_t + 0.204 * VOLP_t \quad (27)$$

$$Sentiment_{Germany,t}^{total} = 0.556 * lnipo_t + 0.07 * equityratio_t - 0.620 * DP_t + 0.546 * VOLP_t \quad (28)$$

$$Sentiment_{China,t}^{total} = 0.584 * lnipo_t - 0.316 * DP_t + 0.381 * VOLP_t - 0.637 * CEFD_t \quad (29)$$

[22] Baker et al. (2012) reported that the first principal component of the total sentiment indices explained 48%, 40%, 37%, 37%, 38% and 42% of the variance in Germany, France, the UK, Japan, Canada and the US, respectively.

$Sentiment_{HK,t}^{total}$

$$= 0.523 * \ lnipo_t \ + 0.411 * equityratio_t \ - 0.539 * DP_t + 0.517 * VOLP_t \qquad (30)$$

The global index is obtained as the first principal component of the six indices.

$$Sentiment_t^{Global} = 0.559 * \ Sentiment_{US,t}^{total} \ + 0.461 * Sentiment_{UK,t}^{total} \ + 0.346 * Sentiment_{JPN,t}^{total} + 0.092 * Sentiment_{Canada,t}^{total} + 0.445 * Sentiment_{France,t}^{total} + 0.388 * Sentiment_{Germany,t}^{total} \qquad (31)$$

This first principal component accounted for 30.22% of the variance.

In this study, we orthogonalized sentiment in China and Hong Kong relative to global sentiment to isolate its unique impact on the A-H share price disparity, thereby obtaining local sentiment for both regions. This technique, widely used in the literature, such as Baker, Wurgler, and Yuan (2012), helps remove overlapping effects and capture the independent influence of global sentiment. We also analyzed unorthogonalized local sentiment for robustness, and the results remained consistent; however, to save space, they are not included.

3.4.2 Firm-level variables

The analysis of A–H cross-listed firms controlled for different firm characteristics. First, the analysis controlled for the firm's annual market capitalisation (mvc), which includes the market value of outstanding stocks on all listed stock exchanges. When a company is listed on multiple stock exchanges, its market capitalisation is determined by aggregating the market values of its outstanding shares across all stock markets. Market capitalisation is a useful proxy to capture the degree of firm information asymmetry. Market capitalisation data were obtained from Datastream. Additionally, the study controlled for the relative transaction costs of dual-listed stocks, measured by the H–A bid–ask spread gap (SPR), defined as the difference

between the monthly bid–ask spreads of H- and A-shares. A positive SPR indicates a higher relative transaction cost for H-stocks. In addition, the analysis considered each firm's relative stock liquidity (Liquidityratio), measured as the ratio of H-stock to cross-listed A-stock turnover. If the liquidity ratio is greater than 1, it indicates that the A-stock trades less frequently and has lower liquidity than the H-stock. Stock turnover was calculated as the trading volume divided by the total outstanding shares in each stock market. The study further considered the A–H volatility ratio (A–Hvolatility), the ratio of the monthly average of 21-day rolling annualised volatility for A- and H-stocks. A higher volatility ratio indicates that the volatility risk of the A-stock is relatively higher than that of the H-stock. The study also considered the relative idiosyncratic risk (Idio), measured as the ratio of A to H idiosyncratic volatility. Idiosyncratic volatility was calculated as the monthly standard deviation of residuals from a time-series regression using the CAPM model. In addition, the analysis also controlled for other firm characteristics: dividend yield (calculated as annual dividends per share divided by stock price), institutional ownership (the fraction of institutional investor ownership of A-shares), short-sale eligibility (indicated by a dummy variable for A-shares' availability for short selling) and the QFII holding ratio (the fraction of QFII ownership of A-shares)

3.4.3 Basic approach

After obtaining the global sentiment index and local sentiment indices for China and Hong Kong, we proceeded to determine whether global sentiment can predict the A–H premium and, if so, the mechanism for the contagion of global sentiment in these two markets. We proposed that the A and H markets would have an asymmetric risk exposure to global sentiment because of differential accessibility to global investors. This could result in variations in the A–

H premium. We further hypothesised that if global sentiment can spread to both markets by cross-border capital flow, the openness of the home market may influence the transmission of global sentiment and its impact on the A–H premium. Arbitrageurs may be unable to fully exploit mispriced cross-listed shares because of market frictions such as transaction costs. This would make the negative marginal effect of global sentiment on the A–H premium more pronounced for shares with higher transaction costs in the Hong Kong market, as it is more expensive for global investors to engage in arbitrage. Baker and Wurgler (2006) have shown that stocks that are difficult to value or arbitrage are more sensitive to sentiment. Therefore, we hypothesised that the impact of global sentiment on the A–H price premium would be asymmetrical across firms with different relative transaction costs. The negative marginal effect of global sentiment on the A–H premium should be stronger for firms with a higher relative H-share arbitrage risk.

The study developed an econometric model to investigate the A–H premium. Panel data techniques are ideal for this investigation as they can deal with factors that change over time and across firms. The main advantage of using panel data is to control for unobservable firm-specific effects.

$$
\begin{aligned}
&AH\ premium_{i,t} \\
&= \alpha_0 + b_1 AH\ premium_{i,t-1} + b_2 Market\ return_{A,t} + b_3 Market\ return_{H,t} + b_4 World\ return_t \\
&\quad + b_5 Globalsent_{t-1} + b_6 Chinalocalsent_{t-1} + b_7 HKlocalsent_{t-1} + b_8 Z_{i,t-1} \\
&\quad + b_9 X_t + C_i + \varepsilon_{i,t}
\end{aligned}
\tag{32}
$$

For each firm *i*, the A–H premium was calculated as the difference between the A-share price and the exchange rate-adjusted H-share price, divided by the A-share price in month *t*. The A–H premium is equivalent to the H–A share discount. $Market\ return_A$ and

$Market\ return_H$ represent the China and Hong Kong market return. $Globalsent$ is the global sentiment. $Chinalocalsent$ and $HKlocalsent$ are China and Hong Kong local sentiment. $World\ return$ is the global market return. We use the MSCI world index return to measure the global market return. The variable $Z_{i,t-1}$ represents firm characteristics with a one-month lag, including the A–H liquidity ratio, A–H volatility ratio, A–H bid–ask spread gap, market capitalisation and dividends. X_t captures country control variables, including the exchange rate, currency risk and interest rate. The variable c_i represents the unobservable firm fixed effect, and $\varepsilon_{i,t}$ is the disturbance term with a mean of 0; it is assumed to be heteroskedastic. To measure the predictive value of the sentiment effect, we used the sentiment index from the previous month.

It is thought that sentiment can spread across international borders by its influence on investment decisions. Accordingly, the degree of market openness and foreign accessibility may affect the impact of global sentiment on A–H mispricing. Bailey and Jagtiani (1994) explored the relationship between the mispricing of cross-listed shares and foreign ownership restrictions. They found that the variability in price spreads correlated with differing risk exposures because of market segmentation. Hence, financial integration in markets mitigates market segmentation, which enhances price discovery efficiency, leading to price equality, and reduces the sensitivity difference between the two markets to global risk. Our study focuses on how foreign ownership restrictions influence the behaviour of A and H cross-listed shares in the Chinese market, considering the different levels of market openness in Hong Kong and mainland China. We used the QFII, RQFII and SHSC programs to examine the relaxation of ownership restrictions in China. These institutional reforms symbolise China's progressive easing of capital ownership and liberalisation of financial markets. A policy reform dummy

variable was assigned a value of 1 when a policy was implemented. We investigated how policy reform affected the marginal effect of global sentiment on the A–H premium by including each policy dummy in the benchmark regression, as shown in Equation 33.

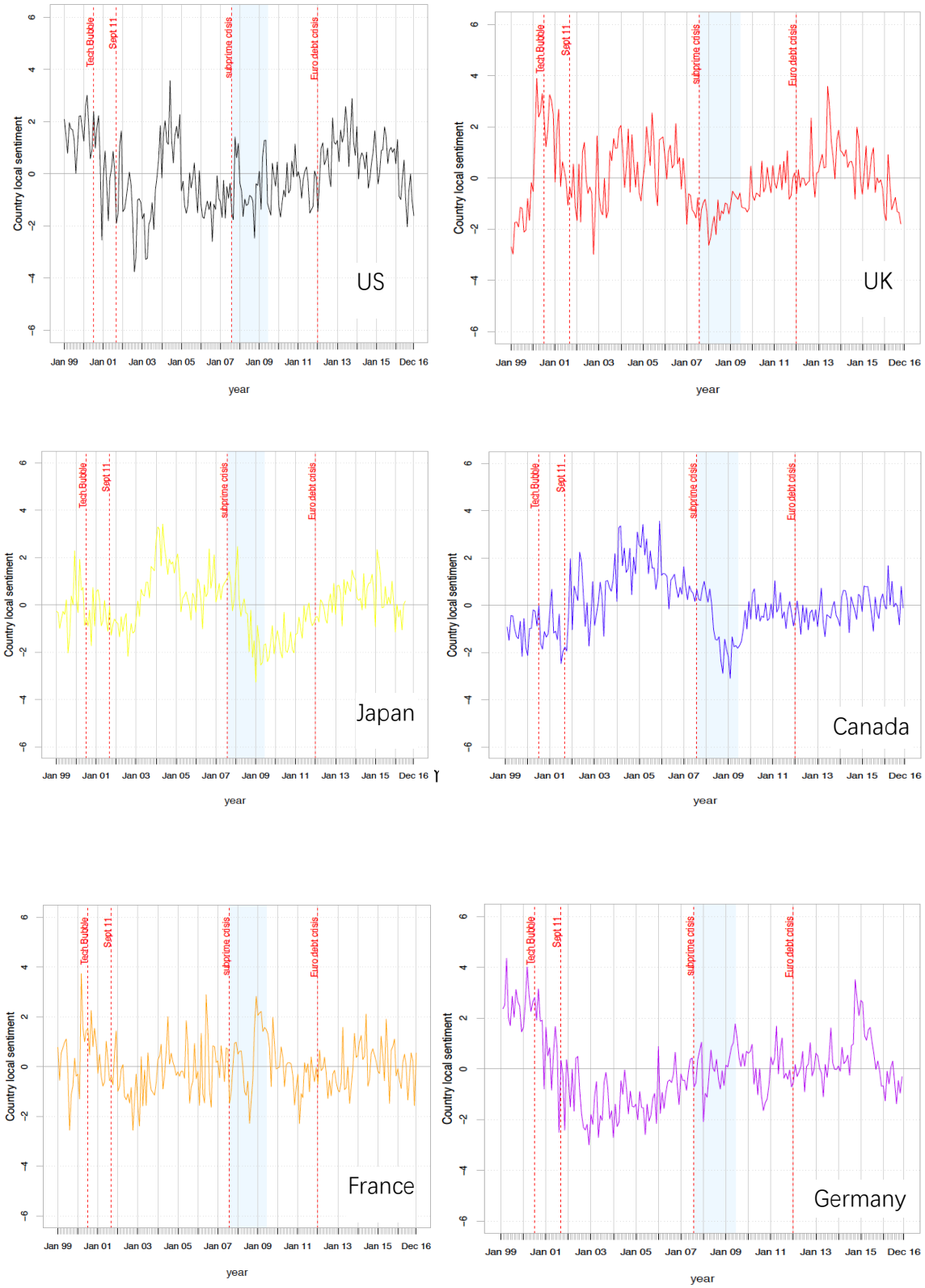
$$\begin{aligned}\text{A}-\text{H premium}_{i,t} &= \alpha_0 + b_1 A-H\ premium_{i,t-1} + b_2 Market\ return_{A,t} + b_3 Market\ return_{H,t} \\ &+ b_4 World\ return_t + b_5 Globalsent_{t-1} + b_6 Chinalocalsent_{t-1} \\ &+ b_7 HKlocalsent_{t-1} + b_8 Policy\ reform_{i,t-1} + b_9 Policy\ reform_{i,t-1} \\ &* Globalsent_{t-1} + b_{10} Policy\ reform_{i,t-1} \\ &* Chinalocalsent_{t-1} + b_{11} Policy\ reform_{i,t-1} * HKlocalsent_{t-1} + b_{12} Z_{i,t-1} \\ &+ b_9 X_t + c_i + \varepsilon_{i,t}\end{aligned} \tag{33}$$

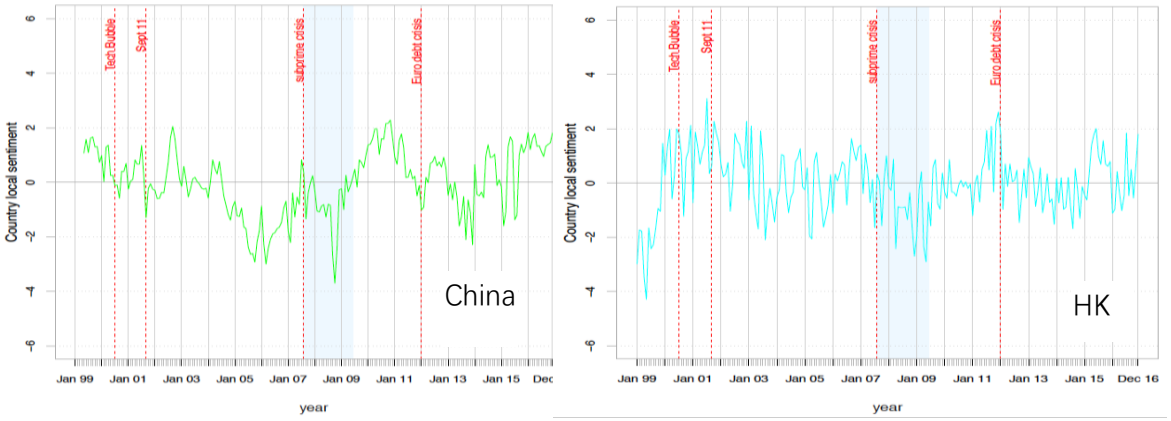
3.5 Empirical results

3.5.1 Statistical analysis

Figure 3.1 shows the sentiment indices of eight countries, standardised to a mean of 0 and a standard deviation of 1. The dashed lines mark financial crisis events, and the shaded area represents the 2007 subprime crisis. This figure covers four significant financial crises: the tech bubble, September 11, the subprime crisis and the Euro debt crisis. At the time of the 2007 subprime crisis, the US experienced a substantial and sudden drop in sentiment. This decline appeared with a slight delay in other countries. In the month of each financial crisis event, there was a noticeable jump in sentiment.

Figure 3.1: Historical sentiment trend



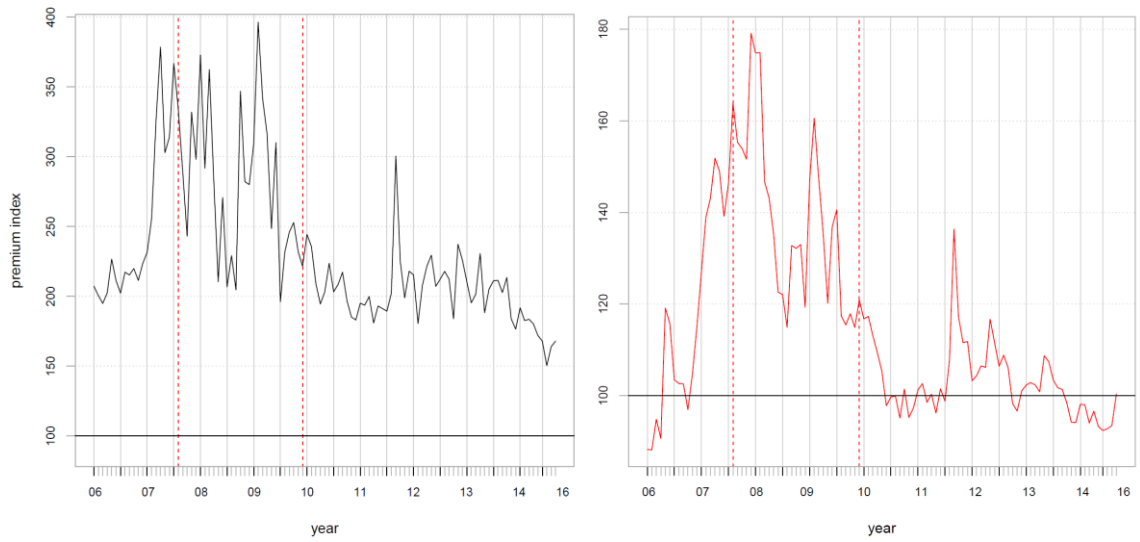


N*ote*. Figure 3.1 shows the historical trend of sentiment for eight countries. All sentiment indices were standardised with a mean value of 0 and a standard deviation of 1. The dashed lines show financial crisis events, and the shaded area shows the 2007 subprime crisis. This figure highlights four significant financial crises: the tech bubble, September 11, the subprime crisis and the Euro debt crisis.

At the end of our sample period, the number of firms dual-listed in the A and H markets had reached 80,[23] representing a larger sample size than in earlier research on A–H cross-listed firms. Our study analysed only the complete sample of matched dual shares with complete stock price information for each month. Two stocks with less than 12 months of matched sample periods were excluded from the regression analysis. Figure 3.2 shows the historical trend of the sample A–H average price ratio, defined as A/H×100, alongside the Hang Seng A–H premium index, which employs a different calculation method. Although the two lines show similar trends, they do not fully coincide. There is a gradual decrease in the A–H premium after 2011. The Hang Seng A–H premium index is based on the most liquid and representative A–H stocks. Its recent downward trend is attributed to improved market integration and effective implementation of financial reform policies in China. A detailed description of the methodology for measuring the Hang Seng A–H premium index is provided in Appendix C.

[23] See Appendix C for more sample details for the cross-listed firms with A–H-shares.

Figure 3.2: Historical trend of the sample average A–H price ratio and the Hang Seng A–H premium index



N*ote*. Figure 3.2 shows the historical trend of the sample average A–H price ratio, defined as A/H×100 (left) and the Hang Seng A–H premium index (right). Sample period ranges from 2006 to 2016. The sample A–H price ratio is based on the average across all A–H-stock pairs in our sample. The Hang Seng A–H premium index is based on the most liquid and representative A–H stocks and uses a different calculation methodology, which is detailed in Appendix C.

3.5.1.1 Summary statistics for firm characteristics

Panel A of Table 3.1 summarises the key statistics for firm characteristics. During the sample period, the average A–H premium reached nearly 0.36. On average, A-shares had a smaller bid–ask spread gap. However, the A–H liquidity ratio was negatively skewed, with the mean ratio exceeding 1 but the median standing at only 0.64. This result suggests that retail investors dominated the A-share market, consistent with the low average institutional ownership of 0.17 in our sample. After the initiation of the QFII program, foreign qualified institutional investor ownership of A-shares remained low, averaging 0.003 with a maximum of 0.02, indicating continued foreign capital controls in China.

Panel B of Table 3.1 presents summary statistics for paired A- and H-shares, revealing several notable points. A-shares showed lower monthly volatility in daily returns, although there was no significant difference in idiosyncratic volatility. The average exchange rate-adjusted price gap between A- and H-shares was 4.12, a considerable difference.

Table 3.1: Summary statistics for firms' characteristics

Panel A: Firms' characteristics

Variable	Min	Max	Median	Mean	SD
A–H premium	−0.639	0.968	0.364	0.358	0.296
MV	2,584.246	891,443.100	37,205.940	127,777.500	228,156.300
SPR	−0.080	0.004	−0.010	−0.016	0.022
Liquidity ratio	0.060	4.754	0.643	1.134	1.269
Volatility ratio	0.429	1.786	0.925	0.977	0.371
Idio ratio	0.260	1.511	0.659	0.727	0.342
Dividend	0	28,031	491	3,293.645	6,945.834
IOR	0.002	0.693	0.082	0.169	0.197
QFII holding ratio	0	0.018	0.001	0.003	0.005

Panel B: Stock return characteristics of A–H-shares

	A-shares					H-shares					A–H pairs	
Variable	Min	Max	Median	Mean	SD	Min	Max	Median	Mean	SD	Diff	T-value
Ret vol	0.16	0.75	0.37	0.40	0.16	0.19	0.83	0.44	0.47	0.18	−0.07	−13.28***
Idio	−0.13	0.16	−0.01	0.00	0.08	−0.21	0.23	−0.02	0.00	0.13	0.00	−0.61
Turnover	0.00	0.03	0.00	0.01	0.01	0.00	0.03	0.01	0.01	0.01	0.00	−7.58***
B/A	0.00	0.02	0.01	0.01	0.00	0.01	0.10	0.02	0.03	0.03	−0.02	−28.31***
Price	0.00	68.18	6.74	9.24	8.31	0.17	65.25	2.48	5.9	9.03	4.12	15.62***

Note. The sample period was from 2000 to 2016. Dividends are in thousands; market capitalisation is in millions. Table 3.1 gives summary statistics of firm characteristics for cross-listed stocks. Panel A summarises the key statistics of the sample firms; Panel B summarises the return characteristics of the cross-listed A–H stocks in the sample. A–H premium is the A–H price premium of the cross-listed shares. MV is market capitalisation, including outstanding stock market value across all listed exchanges. SPR is the ratio of the H-share bid–ask spread to the A-share bid–ask spread. The liquidity ratio is the ratio of H- to A-stock turnover. The volatility ratio is the ratio of the monthly average 21-day return rolling annualised volatility for A- and H-shares. Idio ratio is the ratio of the monthly idiosyncratic volatility for A- and H-shares. Dividend is the annual firm dividend. IOR is the A-share institutional ownership ratio. QFII holding ratio is the qualified foreign institutional investor institutional ownership ratio of A-shares. Ret vol is the monthly stock return volatility. Idio is the idiosyncratic volatility measured as the monthly standard deviation of residuals from the CAPM time-series regression model. Turnover is stock turnover, calculated as trading volume divided by the total shares outstanding in each market. B/A is the bid–ask spread. Price is the stock price. *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

3.5.1.2 The Granger causality

Table 3.2 gives the lead–lag relationship between sentiments across different regions. To ensure stationarity, the Granger causality test was applied to the first derivative of the sentiment factor. Initially, we examined the causality between local sentiment in China and Hong Kong and global sentiment. No predictive relationship was observed in either direction using a three-month lag period selected using the Akaike information criterion. These findings suggest that global sentiment is unlikely to influence local sentiment in either of these markets.

Table 3.2: Granger causality test

Granger causality	F test	*p*
Section 1		
Global sentiment causes Chinese local sentiment	0.887	0.449
Chinese local sentiment causes global sentiment	0.987	0.4
Section 2		
Global sentiment causes Hong Kong (HK) local sentiment	0.333	0.801
HK local sentiment causes global sentiment	0.389	0.761
Section 3		
Chinese local sentiment causes HK local sentiment	0.498	0.737
HK local sentiment causes Chinese local sentiment	0.827	0.51

Note. Table 3.2 shows the results of Granger causality tests used to examine the lead–lag relationship between sentiment indices across regions. The test used the first derivative of the sentiment factors to ensure stationarity. Based on the Akaike information criterion, the optimal lag for our model is three months.

Table 3.3 presents the correlations between sentiment indices in each country. There was a significant positive correlation among the US, the UK and Japan. China's local sentiment showed a slight correlation with the US, but Hong Kong's local sentiment was positively correlated only with the UK. Neither China's nor Hong Kong's local sentiment showed a significant correlation with global sentiment. This result reinforces the need for an independent analysis of local and global sentiment, specifically emphasising the importance of separately analysing China's local sentiment and global sentiment. We also posit the presence of an as-yet-unidentified factor, not accounted for by other risk factors, that could increase the

explanatory capacity of the current benchmark models. This result further reveals that local sentiments in China and Hong Kong are not correlated despite their geographical proximity, suggesting that market segmentation results in distinct investment environments for participants in these two markets.

Table 3.3: Sentiment correlations

Variable	*M*	*SD*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. US	0.00	1.32								
2. UK	0.00	1.31	.22** [.08,.34]							
3. Japan	0.00	1.21	.22** [.08,.34]	.29** [.16,.41]						
4. Canada	−0.00	1.22	−.02 [−.15,.12]	.20** [.07,.33]	.49** [.38,.58]					
5. France	−0.00	1.05	.26** [.13,.38]	.19** [.06,.32]	.06 [−.08,.19]	−.07 [−.20,.07]				
6. Germany	−0.00	1.39	.41** [.29,.52]	.09 [−.04,.22]	−.11 [−.24,.03]	−.38** [−.49, −.26]	.28** [.15,.40]			
7. China	−0.00	1.20	.17* [.04,.30]	−.16* [−.29, −.03]	−.29** [−.41, −.16]	−.24** [−.36, −.11]	−.02 [−.16,.11]	.22** [.09,.35]		
8. HK	0.00	1.22	−.11 [−.24,.03]	.23** [.10,.35]	.09 [−.04,.23]	.02 [−.12,.15]	−.05 [−.18,.08]	−.17* [−.29, −.03]	.08 [−.05,.22]	
9. Global	0.00	1.22	.65** [.56,.72]	.72** [.65,.78]	.73** [.66,.79]	.33** [.21,.45]	.24** [.11,.36]	.16* [.03,.29]	−.13 [−.26,.01]	.11 [−.02,.24]

Note. Table 3.3 reports the correlations between sentiment indices in each country. Means, standard deviations and correlations are listed. Values in square brackets indicate the 95% confidence interval for each correlation. As reported by Cumming (2014), the confidence interval provides a plausible range of population correlations that could have caused the observed sample correlation. * indicates $p < .05$; ** indicates $p < .01$.

3.5.2 Regression analysis

We began our analysis by examining our first hypothesis about the relationship between sentiment and the A–H premium. Using panel data, we employed fixed-effect estimates for our regression models. The dependent variable is the A–H price premium. L.premium denotes the A-premium, L.CNsent represents Chinese local sentiment and L.HKlocalsent signifies Hong Kong local sentiment, all lagged by one month. AretM and HretM are the monthly market returns of China and Hong Kong, respectively. DeltaexHK is the monthly percentage change in the Hong Kong/China exchange rate. CNexvol is the monthly annualised 21-day rolling standard deviation of the US–China exchange rate. CNrf and HKrf denote the risk-free rates in China and Hong Kong, respectively. Lidiosyn is the one month–lagged relative idiosyncratic risk, calculated as the ratio of A and H idiosyncratic volatility. The idiosyncratic volatility is derived from the monthly standard deviation of residuals from the CAPM model. The policy dummy variables QFII, RQFII and SHSC were assigned a value of 1 after policy implementation. Capitalflow is the relative capital inflow, defined as the ratio of portfolio investment in equity and funds into China and Hong Kong and standardised with a mean of 0 and a standard deviation of 1. To conserve space, other control variables, such as the one month–lagged MVC, dividend, SPR and Liquidityratio, are not shown in the table.

Column 1 and 2 of Table 3.4 shows that global sentiment significantly negatively predicts the A–H premium, but local Chinese sentiment has a positive predictive effect. Intriguingly, Hong Kong local sentiment has similar predictive power to Chinese local investor sentiment for the A–H premium, challenging a literature proposal of opposing effects of local sentiment in these two markets. This outcome could be attributed to geographical proximity

and shared information between investors in both markets. A one standard deviation increase in global sentiment correlated with a 0.004 standard deviation decrease in the A–H premium as a significant inverse relationship. This impact was more pronounced than the effect of local sentiment in China and Hong Kong: a similar increase in local sentiment resulted in an increase in the A–H premium of less than 0.002 standard deviations. The influence of global sentiment on the A–H premium was comparable in magnitude to changes in local economic indicators, such as the local exchange rate percentage change (DeltaexHK) (+0.002) and the Hong Kong risk-free rate (+0.002).

While the estimated coefficient for L.Globalsent appears small at -0.004 controlling the relative idiosyncratic risks , its economic implications are far from negligible, especially when contextualized within the dynamics of the A-H premium's variation. The persistence of the A-H premium, as indicated by its high autoregressive coefficient (approximately 0.93), substantially narrows the scope of unexplained variation, thereby amplifying the impact of new information or external variables like global sentiment. Consequently, a standard deviation increase in L.Globalsent results in a decrement of approximately 0.00488 in the A-H premium, equating to about 5% of the residual standard deviation of the A-H premium post accounting for persistence. This effect is economically significant given the premium's limited residual variance (approximately 0.0188), highlighting a substantial relative change.

Further, it is crucial to consider the broader market context in which these effects occur. The A-H premium, pivotal in arbitrage strategies and market efficiency assessments, reflects deeper market dynamics and investor sentiment. Thus, changes in global sentiment, though seemingly minor, are amplified by the market's characteristics, including high persistence and

investor sensitivity to international cues. Therefore, the economic significance of the global sentiment's influence on the A-H premium extends beyond its immediate impact on price adjustments, influencing broader market behavior and investor strategies.

Table 3.4: Regression of A–H premiums against local and global sentiment

Variable	A–H premium					
	(1)	(2)	(3)	(4)	(5)	(6)
L.premium	0.936***	0.937***	0.933***	0.940***	0.936***	0.926***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
L.CNsent	0.002**	0.002**	0.001	−0.003**	0.002*	0.001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
L.HKlocalsent	0.002***	0.002***	0.015***	0.006***	0.003***	0.0005
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
L.Globalsent	−0.003***	−0.004***	−0.006***	−0.003***	−0.004***	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AretM	0.365***	0.355***	0.360***	0.349***	0.342***	0.377***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
HretM	−0.457***	−0.454***	−0.466***	−0.457***	−0.462***	−0.529***
	(0.025)	(0.025)	(0.026)	(0.025)	(0.026)	(0.028)
DeltaexHK	0.002*	0.002*	0.003**	0.003**	0.002*	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CNexvol	−0.133	−0.178*	−0.030	−0.271**	−0.349***	−0.059
	(0.106)	(0.103)	(0.111)	(0.108)	(0.113)	(0.117)
CNrf	0.001	0.001	0.001	0.003***	0.001	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HKrf	0.003***	0.002***	0.001	−0.0001	0.002***	−0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
L.Idio		0.039***	0.040***	0.039***	0.038***	0.040***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
QFII			−0.010*			
			(0.006)			
RQFII				−0.013***		
				(0.003)		
SHConnect					−0.003	
					(0.003)	
Capitalflow						0.001*
						(0.001)
QFII ×L.Globalsent			0.001			
			(0.002)			
RQFII ×L.Globalsent				0.005***		
				(0.002)		
SHConnect ×L.Globalsent					0.016***	
					(0.003)	
Capitalflow ×L.Globalsent						0.002*
						(0.001)
Controls	YES	YES	YES	YES	YES	YES
DWtest	2.13	2.11	2.1	2.1	2.09	2.12
Adj R	0.9	0.9	0.9	0.9	0.9	0.9
Observations	5,370	5,370	5,370	5,370	5,370	5,370
F-statistic	3,120.41***	3,002.03***	2,413.3***	2,413.3***	2,426.73***	2,423.21***

Note. We estimated these results using panel-data fixed effects. The dependent variable is the A–H price premium. L.premium is the A premium, LCNsent is Chinese local sentiment and L.HKlocalsent is Hong Kong local sentiment, all lagged by one month. AretM is China's monthly market return. HretM is Hong Kong's monthly market return. DeltaexHK is the monthly percentage change in the Hong Kong–China exchange rate. CNexvol is the monthly annualised 21-day rolling standard deviation of the US–China exchange rate. CNrf and HKrf are the risk-free rates of China and Hong Kong, respectively. L,idio is the one month–lagged relative idiosyncratic risk,

measured as the ratio of A and H idiosyncratic volatilities. Idiosyncratic volatility is the monthly standard deviation of residuals from the CAPM time-series regression. QFII, RQFII and SHSC are policy dummy variables taking a value of 1 after implementation. Capitalflow is the relative capital inflow, defined as the ratio of portfolio equity and fund investment into China v. Hong Kong, standardised to a mean of 0 and standard deviation of 1. To save space, we have not shown other control variables, such as the one month–lagged MVC, dividend, SPR and Liquidityratio. DWtest is the Durbin–Watson statistic. The sample period ranges from 2000 to 2016. The Hausman specification test compares fixed and random effects based on the test result to indicate whether random effects could produce biased estimators. Based on the results obtained, the fixed effect was preferred. Robust standard errors clustered by firm are given in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

3.5.3 Mechanism of cross-border contagion for global sentiment

This study further investigated the influence of global sentiment on the price disparity between cross-listed stocks, focusing on mediating factors that may affect the marginal effect of global sentiment on the A–H premium. If global sentiment affects both markets by driving cross-border investment, the contagion effect should differ with the openness and accessibility of each market to foreign investors. In a context of perfect market integration, cross-listed stocks would be expected to have identical risk characteristics and respond similarly to global risk shocks, so their prices should be unaffected by trading location (see Gultekin et al., 1989; Jorion & Schwartz, 1986). In other words, if there are no restrictions on the access of global investors to both markets, global sentiment should have little influence on price disparities. Thus, the ability of a common shock to explain price disparities disappears if it has the same effect on the prices of dual-listed stocks. Research has established a connection between the mispricing of dual-listed stocks and capital ownership restrictions in the domestic market. Such restrictions expose cross-listed stocks to different risks. Capital ownership restrictions in mainland China mean that dual-listed stocks in Hong Kong and mainland China are exposed to different local and global risks because investors in each market are unable to trade freely in the other market.

If global sentiment spills over by means of cross-border investment, the degree of market openness in each market should affect the extent to which spillover occurs. Thus, global sentiment may account for cross-listed price disparities by its differing effects on each stock pair. To examine this, this study examined three important institutional reforms aimed at easing the degree of ownership restrictions in mainland China: QFII, RQFII and SHSC.

Columns 3 to 5 in Table 3.4 show the marginal effects of three institutional reforms on the predictive value of global sentiment for the A–H premium. The regression results show that the QFII and RQFII policies promoted the integration of the Chinese market because these two dummy variables have a significant negative effect on the A–H premium. The coefficient of QFII was estimated with a value of −0.01, significant at the 10% level. The coefficient of RQFII was significant at the 1% level, with a value of −0.013. Our findings further indicate that the implementation of the RQFII and SHSC policies attenuated the negative impact of global sentiment on the A–H premium. The interaction term coefficient for global sentiment with RQFII is 0.005, significant at the 1% level. The interaction coefficient for global sentiment with SHSC is also significant at the 1% level, with a value of 0.016. However, the interaction between QFII and global sentiment is not significant. Our results confirm that financial market integration can affect the spillover effect of global sentiment and suggest that RQFII and SHSC are more effective policies for promoting cross-border sentiment contagion. We established a positive association between the relaxation of foreign ownership restrictions and the degree of global sentiment contagion in the domestic market. These findings show that integration of the financial market in the home market can influence the global sentiment contagion between the two markets.

To further investigate whether institutional reform affects the predictive value of global sentiment for the A–H premium, we examined the impact of global sentiment on the A–H premium for the subperiods before and after policy implementation. The empirical findings for each subperiod, shown in Table 3.5, confirm that the enactment of RQFII moderated the influence of global sentiment on the A–H premium. Column 1 of Table 3.5 presents the panel

regression for the entire sample period, while Columns 2 to 4 and 5 to 7 give the pre- and post-implementation regression results for each of the three policy tools. The difference in sensitivity of the A- and H-stock prices to global sentiment was reduced after RQFII implementation. Before RQFII, the coefficient of global sentiment was −0.003, significant at the 1% level. In the post-RQFII phase, this coefficient lost statistical significance. Similarly, the coefficient of global sentiment only showed significance in the pre-SHSC era. For the period before QFII policy implementation, the global sentiment coefficient did not achieve statistical significance, but this may be due to an insufficient sample size of cross-listed stocks and a constrained time-series range in that subperiod.

Table 3.5: Regression of the A–H premium against local and global sentiment in the subperiods before and after policy implementation

Variable	A–H premium						
	Whole period	Before policy implementation			After policy implementation		
		QFII	RQFII	SHSC	QFII	RQFII	SHSC
L.premium	0.937***	0.980***	0.930***	0.934***	0.923***	0.882***	0.722***
	(0.005)	(0.014)	(0.008)	(0.006)	(0.006)	(0.011)	(0.026)
L.CNsent	−0.0002	0.002	−0.003**	0.001	0.002*	0.004**	−0.004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
L.HKlocalsent	0.003	−0.003	0.005***	0.003***	0.001	−0.001	−0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
L.Globalsent	−0.003***	−0.0001	−0.003***	−0.005***	−0.002	−0.0004	−0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
AretM	(0.011)	0.137***	0.353***	0.343***	0.389	0.369***	0.225***
	−0.453	(0.025)	(0.015)	(0.013)	(0.012)	(0.018)	(0.038)
HretM	(0.024)	−0.065*	−0.339***	−0.413***	−0.532***	−0.576***	0.133
	(0.011)	(0.034)	(0.029)	(0.027)	(0.029)	(0.050)	(0.118)
L.Liquidityratio	0.0001	0.0004	−0.0004	−0.00005	0.0005	0.001	0.002**
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)
L.Volatilityratio	0.008	0.014**	−0.004	0.001	0.010***	0.030***	0.021***
	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.003)	(0.005)
L.MVC	−0.000	−0.00000	0.000	0.000	−0.000	−0.000	−0.00***
	(0.000)	(0.00000)	(0.000)	(0.000)	(0.000)	(0.00000)	(0.00000)
L.SPR	−0.062	−0.121*	−0.069	−0.081**	−0.065**	−0.072*	0.024
	(0.032)	(0.073)	(0.045)	(0.038)	(0.030)	(0.043)	(0.051)
Controls	YES	YES	YES	YES	YES	YES	YES
DWtest	2.16	1.89	2.21	2.22	2.13	2.03	1.88
Observations	5,439	463	3,027	4,441	4,992	2,412	998
R^2	0.9	0.943	0.9	0.9	0.853	0.85	0.61
F-statistic	4,319.5***	682.69***	1,707.68***	2,761.37***	1965.18***	913.92***	110.38***

Note. Table 3.5 gives the effect of global sentiment on the A–H premium in subperiods before and after policy implementation. The dependent variable is the A–H price premium. The Hausman specification test compares fixed and random effects based on the test result to indicate whether random effects could produce biased estimators. Based on the results obtained, the fixed effect was preferred. Ther control variables, including the world return, DeltaexHK and Cnexvol, are not shown in the table to save space. Robust standard errors clustered by firm are given in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

3.5.3.1 Capital flow & sentiment effects

To further confirm that global sentiment exhibits contagion through cross-border capital flow, this study used data for relative capital inflow into China and Hong Kong as the ratio of portfolio investment in equity and funds. This ratio was standardised with a mean of 0 and a standard deviation of 1. Portfolio investment involves buying and selling securities, and the flow of capital into or out of the stock market can be measured by tracking the net purchases of stocks and bonds by foreign investors. Portfolio investment can be useful for measuring capital flows in stock markets, as demonstrated in research such as studies by Baek (2006) and Stulz (1999). Baek (2006) identified push and pull factors that influence foreign portfolio investment in Asian and Latin American economies, finding that risk appetite and external factors drive Asian investment, whereas strong economic growth and foreign financial factors drive Latin American investment. Stulz (1999) concluded that international portfolio flows can decrease the cost of capital in a country without negative effects on its securities markets. In this study, as seen from Column 6 of Table 3.4, the coefficient for the relative capital flow and the A–H premium is statistically significant at the 10% level, with a value of 0.001. The interaction term between relative capital flow and global sentiment is positive, with an estimated coefficient of 0.002, significant at the 10% level. This result suggests that the relative capital inflow into the two markets affects the marginal impact of global sentiment on the price premium and supports the proposal that global sentiment spillover is mediated by cross-border capital flow. Increased capital flow into mainland China, relative to Hong Kong, can temper the adverse incremental effect of global sentiment on the A–H premium. These results are consistent with research emphasising the role of financial institutional reform in promoting market liberalisation and

attracting foreign investment. Ultimately, both financial reform policies and relative increases in capital inflows into China can mitigate the negative marginal effect of global sentiment on the A–H premium.

3.5.3.2 Firm characteristics

The literature suggests that the sensitivity of a stock's mispricing to sentiment relates to that stock's arbitrage risk. Mispricing can arise because of uninformed demand shock and a binding arbitrage constraint. In practical terms, when a widespread change in sentiment occurs, it is expected to have varying effects across different stocks rather than uniformly affecting all stocks. This is because of potential differences in sentiment-based demands and arbitrage constraints among stocks. Interestingly, these two factors often result in similar outcomes because stocks with highly subjective valuations, which are sensitive to speculative demand, also tend to carry higher risks and greater arbitrage costs. Research suggests that investor sentiment will influence stocks of newer and smaller firms, highly volatile firms, distressed firms, firms with significant growth potential and firms without dividends (Baker & Wurgler, 2006). The objective of this section is, thus, to examine the heterogeneous effects of global sentiment on the A–H premium for different firm characteristics.

The first firm-level mediating factor included was the institutional ownership of A-shares. Institutional ownership has a negative correlation with firm noise trader risk, meaning that stocks with higher levels of institutional ownership are less susceptible to sentiment-driven influences. We included foreign institutional investor ownership as a second factor for comparison. The third firm characteristic examined was the relative idiosyncratic risk between China and Hong Kong, utilized as a proxy for arbitrage risk. Following the approach in Gagnon

and Karolyi (2010) and Stambaugh et al. (2015), idiosyncratic volatility has been identified as a robust measure of arbitrage risk. Pontiff (1996) argues that arbitrage opportunities arise from mispricing, which becomes prevalent when the costs associated with arbitrage surpass potential gains. He identifies several factors influencing arbitrage profitability, including the fundamental risk of the security, transaction costs, the security's dividend yield, and prevailing interest rates. The dividend yield can mitigate arbitrage costs by reducing the duration of holding a mispriced security.

In alignment with this theoretical framework, our study incorporated the H–A bid–ask spread gap as an indicator of the relative transaction costs for H-shares, and the dividend yield as a reflection of arbitrage costs. Additionally, short-selling restrictions, which Lamont and Stein (2004) noted as significant barriers to effective arbitrage, were considered by including an A-share short-sale eligibility dummy for individual stocks. This dummy variable takes a value of 1 if the A-share was eligible for short selling, reverting to 0 if removed from the pilot program introduced by the China Securities Regulatory Commission (CSRC) on 31 March 2010. Our sample indicated that 58 of 80 A-stocks were eligible for short selling during the observed period. The market microstructure literature, including works by Glosten & Milgrom (1985) and Kyle (1985), suggests that high levels of information asymmetry, typically reflected in market capitalization, lead to increased trading costs as market makers set higher spreads to hedge against informed trading.

Empirical results, as shown in Table 3.6, indicate that the marginal effects of global sentiment on price deviations are asymmetrical across firms with differing A-share short-sale eligibility, market capitalization, and relative trading costs (as measured by the SPR). We

observed that a higher relative H-share trading cost (SPR) and the ineligibility of A-shares for short selling amplified the negative impact of global sentiment on A–H price deviations. Conversely, firms with larger market capitalizations, which typically exhibit lower levels of information asymmetry, experienced a lesser negative impact from global sentiment.

Significantly, the coefficient for A-share short-sale eligibility was -0.009, with the interaction term between A-share short-sale eligibility and global sentiment also being significant at -0.004. This indicates that when A-shares are short-sale eligible, arbitrage activities are more feasible, allowing investors to exploit and reduce price discrepancies between A and H shares effectively. This enhances market efficiency by correcting mispricings. However, the absence of short-selling options discourages arbitrage, leading to more pronounced sentiment-driven mispricing. Therefore, the interaction between A-share short-sale eligibility and global sentiment is positively significant. This result is consistent with the proposal that short sellers act as stabilising forces in markets in periods of strong sentiment, as they can profit from overpricing by selling short and help correct market imbalances.

Table 3.6: Firms' characteristics mediating the global sentiment effect on the A–H premium

		Dependent variable: A–H premium					
Panel	Model	Control	Globalsent	Globalsent ×Controls	Cnlocalsent ×Controls	Hklocalsent ×Controls	adj. R
A	IOR	−0.013*	−0.004***	0.005	0.009**	−0.003	0.89
		(0.007)	(0.001)	(0.004)	(0.004)	(0.004)	
B	QFIIholding	−0.892***	−0.004***	0.099	−0.022	−0.700***	0.9
		(0.294)	(0.001)	(0.211)	(0.157)	(0.203)	
C	Idio	0.037***	−0.014***	0.016***	0.009***	−0.008***	0.9
		(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	
D	DividendYield	−0.002**	−0.004***	0.0001	0.001	−0.00004	0.9
		(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	
E	SPR	−0.163***	−0.005***	−0.079**	−0.023	0.126***	0.9
		(0.053)	(0.001)	(0.039)	(0.037)	(0.044)	
F	MVC	−0.00*	−0.004***	0.000**	0.000	0.000*	0.9
		(0.00)	(0.001)	(0.000)	(0.000)	(0.000)	
G	Short indicator	-0.009***	-0.004***	0.006***	0.001	0.003	0.9
		(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	

Note. Table 3.6 presents firm characteristics that can affect the marginal effect of global sentiment on the A–H premium. Each panel corresponds to a specific firm characteristic that is controlled for. The regression model includes an interaction term between the firm characteristic, global sentiment and local sentiment. We regressed these models by using panel-data (fixed-effect) estimates. The dependent variable is the A–H premium. The independent variables include L.premium, L.Globalsent, LCNsent, L.HKlocalsent, AretM, HretM, L.Liquidityratio, DeltaexHK and CNexvol. The detailed definition is listed in the appendix. The sample period was from 2000 to 2016. We applied Durbin–Watson statistics. The Hausman specification test compares fixed and random effects based on the test result to indicate whether random effects could produce biased estimators. Based on the results obtained, the fixed effect was preferred. Robust standard errors clustered by firm are given in parentheses.

*, ** and *** show significance at the 10%, 5% and 1% levels, respectively.

To explore how global sentiment affects the A–H premium in stocks with different foreign institutional investor holdings, we divided stocks into two categories based on ownership by QFIIs during the sample period. The QFII-held group included shares owned by QFIIs approved to hold A-shares in this period. The non-QFII-held group comprised stocks for which QFII ownership in A-shares was zero. We performed a panel regression analyses for each group. The results are given in Table 3.7 and reveal that global sentiment did not significantly affect the A–H premium for non-QFII-held stocks. However, for QFII-held stocks, there was a significant negative correlation between global sentiment and the A–H premium. This supports our sentiment contagion hypothesis that A–H-share price premiums are more influenced by global sentiment in pairs where the corresponding A-shares were previously held by qualified institutional investors. For stocks never held by such investors, the predictive power of global sentiment is negligible. This could be due to factors such as higher investment costs, lower investor attraction and less information transparency, making these A-shares firms less attractive to foreign institutional investors. Our findings suggest that the introduction of QFII and RQFII policies had a significant impact in reducing the premium for both groups of stocks, indicating that the relaxation of capital controls is an effective measure in promoting market integration. The interaction between global sentiment and the RQFII and SHSC policies showed a significantly positive effect for QFII-held shares. This result suggests that the A–H premium's responsiveness to global sentiment was influenced by the implementation of the RQFII and SHSC policies. This result aligns with our hypothesis that the spillover effects of global sentiment in the two markets depend on the level of home market openness. In scenarios where foreign investors are not directly involved in A-share investments, the influence of global

sentiment on the A–H premium becomes negligible, and the incremental impact of policy implementation decreases. This finding could be attributed to an inherent lack of attractiveness of these A–H cross-listed Chinese companies to foreign investors. Such an outcome would allow us to dismiss the likelihood of word-of-mouth and emotional contagion as channels for the transmission of global sentiments into these markets.

Table 3.7: Panel regression analysis of QFII-held and non-QFII-held stocks

Variable	A–H premium							
	Non-QFII-held stocks				QFII-held stocks			
L.premium	0.904***	0.893***	0.902***	0.903***	0.939***	0.927***	0.935***	0.939***
	(0.018)	(0.019)	(0.018)	(0.019)	(0.005)	(0.006)	(0.005)	(0.005)
L.CNsent	−0.001	−0.001	−0.001	−0.001	0.00004	−0.0001	0.001	0.0001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
L.HKlocalsent	0.002	0.002	0.002	0.002	0.003***	0.002**	0.003***	0.003***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
L.Globalsent	−0.003	−0.002	−0.001	−0.005*	−0.003***	−0.002**	−0.003***	−0.005***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
QFII		−0.018**				−0.017***		
		(0.009)				(0.003)		
L.Globasent ×QFII		−0.002				−0.002		
		(0.004)				(0.001)		
RQFII			−0.014**				−0.010***	
			(0.006)				(0.002)	
L.Globasent ×RQFII			−0.001				0.004**	
			(0.004)				(0.002)	
SHSC				0.001				−0.004
				(0.006)				(0.003)
L.Globasent ×SHSC				0.005				0.012***
				(0.005)				(0.002)
Controls	YES	YES	YES	YES	YES	YES	YES	
DWtest	2.18	2.16	2.17	2.18	2.05	2.04	2.06	2.04
Observations	4,738	4,738	4,738	4,738	717	717	717	717
R^2	0.9	0.9	0.9	0.9	0.83	0.83	0.83	0.83
F-statistic	3,990.9***	3,390.38***	3,392.65***	3,398.56***	299.28***	253.7***	254.92***	253.26***

Note. Table 3.7 gives the results of panel regression analyses conducted on two distinct groups of stocks. We classified the stocks into two groups based on their ownership by Qualified Foreign Institutional Investors (QFIIs) during the historical sample period. The QFII-held group consists of shares that were held by QFIIs who were approved to hold A-shares during the sample period. The non-QFII-held group includes stocks not held by any QFIIs during our sample period. DWtest gives the Durbin–Watson statistics. Robust standard errors clustered by firm are given in parentheses. *, ** and *** show significance at the 10%, 5% and 1% levels, respectively.

3.5.3.3 Global sentiment & relative idiosyncratic risk

Gao et al. (2012) found that investor sentiment can help correct the negative relationship between idiosyncratic volatility and expected returns. We investigated the influence of global investor sentiment on the relationship between idiosyncratic risk and the price premium of cross-listed shares. Our work contributes to the literature by using investor sentiment to address the puzzle of idiosyncratic volatility. By examining the relationship between global sentiment, cross-listed share mispricing and idiosyncratic risk in two markets, we can shed light on the impact of global sentiment on the relative idiosyncratic risk of cross-listed stocks.

Table 3.8 reveals the impact of global sentiment on the relative idiosyncratic risk of cross-listed A–H-shares (see Columns 1 and 2). The results indicate that global sentiment markedly increased relative idiosyncratic volatility, with a coefficient of 0.05 significant at the 1% level. Both mainland Chinese and Hong Kong local sentiment had a significant negative impact on A–H relative idiosyncratic volatility. However, Columns 3 and 4 show a lack of relationship between global sentiment and A–H-stock volatility. A comparison with Columns 1 and 2 of Table 3.4 reveals that the net impact of global sentiment was greater after we accounted for relative idiosyncratic risk because relative idiosyncratic risk significantly and positively affected the A–H premium. Our main empirical finding is that the negative influence of global sentiment on the A–H premium cannot be attributed to its effect on relative idiosyncratic risk, as the two factors operate in opposite directions.

Table 3.8: Panel regression of A–H idiosyncratic risk and volatility against global sentiment

	Dependent variable			
	A–H idiosyncratic volatility		A–H volatility	
Variable	(1)	(2)	(3)	(4)
AretM	0.278***	0.311***	(0.066)	(0.004)
	(0.055)	(0.056)	(0.060)	(0.061)
HretM	−0.800***	−0.757***	−0.505***	−0.411***
	(0.110)	(0.111)	(0.120)	(0.121)
Worldreturn	1.177***	1.134***	1.035***	0.951***
	(0.119)	(0.120)	(0.131)	(0.131)
CNsent	−0.032***	−0.034***	−0.027***	−0.030***
	(0.004)	(0.004)	(0.004)	(0.004)
HKlocalsent	−0.015***	−0.016***	−0.015***	−0.015***
	(0.004)	(0.004)	(0.004)	(0.004)
Globalsent	0.025***	0.023***	0.004	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Liquidityratio	0.073***	0.068	0.099***	0.091***
	(0.005)	(0.005)	(0.005)	(0.005)
DeltaexHK		−0.029***		−0.026***
		(0.006)		(0.007)
CNexvol		1.729***		3.551***
		(0.511)		(0.559)
SPR	0.025	0.092	−0.434*	−0.309
	(0.238)	(238.000)	(0.261)	(0.260)
MV	0.0000***	0.0000***	0.0000***	0.0000***
DWtest	1.58	1.58	1.26	1.27
Observations	5,374	5,374	5,374	5,374
R^2	0.112	0.117	0.110	0.119
Adjusted R^2	0.097	0.102	0.095	0.104
F-statistic	74.132***	63.868***	72.576***	64.731***

Note. Table 3.8 reports the effect of investor sentiment on A–H idiosyncratic risk and volatility. We regressed our model using panel-data (fixed-effect) estimates. The dependent variables are A–H idiosyncratic volatility and A–H volatility. A–H idiosyncratic volatility is the relative idiosyncratic risk, measured as the ratio of A and H idiosyncratic volatility. The idiosyncratic volatility is the monthly standard deviation of residuals from the time-series regression of the CAPM model. A–H volatility is the ratio of the A and H monthly average of 21-day return rolling annualised volatility. The sample period was from 2000 to 2016. DWtest is the Durbin–Watson statistic. Robust standard errors clustered by firm are given in parentheses. Worldreturn is the global market return. We used the Morgan Stanley Capital International (MSCI) world index return to measure the global market return. *, **, and *** show significance at the 10%, 5% and 1% levels, respectively.

To investigate the impact of global sentiment on the sensitivity of A–H mispricing to relative idiosyncratic risk, we introduced an interaction term for global sentiment with idiosyncratic volatility ratio into our regression. Panel D in Table 3.6 shows a significant positive interaction between global sentiment and the A–H idiosyncratic ratio. This finding

indicates that global investor sentiment can strengthen the relationship between relative idiosyncratic volatility and the A–H premium. It is noteworthy that this result contradicts the concept that global sentiment's explanatory power for the A–H premium derives solely from its indirect effect on the A–H idiosyncratic risk level. Rather, our findings suggest that global sentiment can enhance the positive relation between the A–H relative idiosyncratic volatility and the A–H premium. Although global sentiment does affect the relative unsystematic risk between the two markets, its effect on the A–H price disparity cannot be fully explained by the idiosyncratic component.

3.5.4 Time-series regression of the A–H premium against global sentiment

This research aimed to expand the understanding of the linkage between the A–H premium and global sentiment through a thorough examination of time series data for the A–H premium. According to our model, global sentiment should explain the time-series variations of the A–H premium. We conducted time-series regressions of the A–H premium against global sentiment and the sentiment difference between China and Hong Kong, incorporating indicator variables for foreign ownership restriction and other macro variables. Given a high level of autocorrelation in the series, we included the lagged A–H premium as a control variable. The dependent variable was the monthly average A–H premium of all cross-listed stocks in our sample. L.premium is the one month–lagged average A–H premium. Based on our model, the global sentiment should negatively affect the A–H premium. We found that the coefficients of global sentiment aligned with our theoretical predictions and were statistically significant at the 1% level (see Column 1 of Table 3.9).

Surprisingly, after we controlled for the indicator variable QFII, the coefficient of global sentiment was not significant (see Column 2 of Table 3.9). The implementation of the QFII policy narrowed the difference between A- and H-share prices. This result suggests that a pent-up demand for H-shares by global investors was met in the months after the program started and that this more than offset a pent-up demand for A-shares by global investors. The interaction term for QFII and SHSC with global sentiment was not statistically significant, suggesting that these policies, in their current forms, are less effective in shielding the A–H market from fluctuations driven by global sentiment. In contrast, the interaction term for RQFII with global sentiment is positive and significant at the 1% level, confirming that the RQFII policy has mitigated the negative impact of global sentiment on the A–H premium. Among the macro variables, only the coefficient for the consumption growth rate was significantly positive. Due to space limitations, the coefficients for other macro control variables are not shown in Table 3.9. In conclusion, our results suggest that the effects of RQFII policy implementation can override the negative time-series predictive value of global sentiment for the A–H premium.

Table 3.9: Time-series regressions of the average A–H premium against global sentiment

			A–H premium		
Variable	(1)	(2)	(3)	(4)	(5)
L.premium	0.909***	0.814***	0.902***	0.897***	0.897
	(0.034)	(0.041)	(0.035)	(0.035)	(0.035)
L.Globalsent	−0.007**	−0.001	−0.007**	−0.008***	−0.008
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
L.Sentdif	−0.004*	−0.005*	−0.004*	−0.005*	−0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AretM	0.444***	0.433***	0.433***	0.427***	0.427
	(0.059)	(0.057)	(0.059)	(0.060)	(0.060)
HretM	−0.023	−0.043	−0.032	−0.025	−0.025
	(0.050)	(0.048)	(0.050)	(0.050)	(0.050)
QFII		−0.081**			
		(0.020)			
RQFII			−0.031*		
			(0.009)		
SHConnect				−0.043	
				(0.154)	
Capitalflow					0.127**
					(0.051)
QFII		−0.004			
×L.Globalsent		(0.005)			
RQFII			0.016*		
×L.Globalsent			(0.009)		
SHConnect				0.050	
×L.Globalsent				(0.074)	
Capitalflow					0.003
×L.Globalsent					(0.041)
Controls	YES	YES	YES	YES	YES
Constant	−0.231	−0.647	−0.052	−0.302	−0.436
	(0.231)	(0.249)	(0.290)	(0.233)	(0.271)
Observations	178	178	178	178	145
R^2	0.951	0.955	0.953	0.952	0.902

Note. We conducted time-series regressions of the A–H premium against global sentiment and the sentiment difference between China and Hong Kong, incorporating indicator variables for foreign ownership restrictions and other macro variables. Given the high level of autocorrelation in the series, we included the lagged A–H premium as a control variable. The dependent variable is the monthly average A–H premium of all cross-listed stocks in our sample. L.premium is the one month–lagged average A–H premium. The sample period was from 2000 to 2016. *, ** and *** show significance at the 10%, 5% and 1% levels, respectively.

3.5.5 Global sentiment & stock returns

We used pooled regression analysis to assess how global and local sentiment affect A–H-stock returns. Extending Bekaert et al.'s (2009) work on stock return models, we

incorporated these sentiments as variables. Using multivariate regression, we examined whether global sentiment can predict stock excess returns of cross-listed A- and H-shares. We used excess returns as the dependent variable and investor sentiment (in two forms) as the key explanatory variable. We included policy dummy variables to measure the impact of the relaxation of foreign capital restrictions on the predictive value of global sentiment for the A–H premium. The primary model is given in Equation 34:

$$Retx_{i,t}^{A(H)} = \alpha_i + \beta_{i,1}R_{m,t}^{China(HK)} + \beta_{i,1}R_{m,t}^{global} + \beta_{i,2}SMB_t^{China(HK)} + \beta_{i,2}SMB_t^{global} + \beta_{i,3}HML_t^{China(HK)} + \beta_{i,3}HML_t^{global} + \beta_{i4}Sentiment_{t-1}^{local,china} + \beta_{i,5}Sentiment_{t-1}^{local,HK} + \beta_{i,6}Sentiment_{t-1}^{global} + \varepsilon_{i,t}^{A(h)} \quad (34)$$

In our model, we isolated the influence of local risk factors from global factors by orthogonalising local three-factor returns against global three-factor returns. For example, we adjusted the local Chinese SMB (small minus big) factor by removing the influence of the global SMB factor, yielding an orthogonalised local SMB factor in each market. Table 3.10 presents coefficient estimates obtained from pooled regressions of A–H excess returns with the incorporation of measures of investor sentiment and Fama and French's (1993) three-factor model, which includes both global and local factors. T-statistics were calculated using White's (1980) heteroskedasticity-consistent method. The intercept is not reported.

Localsmb, localhml, and localrmf in Table 3.10 represent the local three factors in the China and Hong Kong markets, respectively. Whml, wsmb, and wrmf represent the world three factors. We controlled for policy dummy variables to examine the impact of the relaxation of foreign capital restrictions on the marginal predictive value of global sentiment for the A–H premium. The dependent variable is the A or H monthly stock excess return. The policy

dummies are QFII, RQFII and SHSC. The main effect is the benchmark regression without the inclusion of the policy dummy. Columns 1 and 5 of Table 3.10 give the main effects for the A and H excess returns. The benchmark model in column 1 and column 5 reveals a significant positive relationship between global sentiment and H-stock excess returns, but the coefficient of global sentiment for A-stock excess returns was not significant. The coefficient of the orthogonalised local market excess return for A excess returns was 0.119. The interaction terms for the RQFII and SHSC policies with global sentiment were significantly positive for the regression against A excess returns. However, the interaction of QFII with global sentiment showed a significant negative effect. The impact of global sentiment on excess returns for H-shares in Hong Kong was more pronounced than its effect on the A market. The findings indicate that global sentiment has predictive power for H-share stock returns. We further verified the asymmetric predictive power of global sentiment for A-stock returns under different policy conditions, with its influence being most pronounced following the RQFII policy implementation.

Table 3.10: Global sentiment and A–H-stock returns

	A excess return				H excess return			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy dummy	Main effect	QFII	RQFII	SHSC	Main effect	QFII	RQFII	SHSC
L.Globalsent	0.180	1.256**	−0.193	−0.116	0.767**	2.352***	−0.163	0.953**
	(0.304)	(0.559)	(0.462)	(0.330)	(0.366)	(0.906)	(0.644)	(0.434)
whml	0.233	0.232	0.251	0.260	−0.108	−0.186	0.043	−0.165
	(0.162)	(0.209)	(0.169)	(0.162)	(0.204)	(0.247)	(0.218)	(0.214)
wsmb	0.527**	0.522**	0.534**	0.573**	0.209	0.212	0.09	0.199
	(0.232)	(0.230)	(0.233)	(0.233)	(0.222)	(0.216)	(0.226)	(0.222)
wrmf	0.136	0.185*	0.129	0.108	0.610***	0.710***	0.477***	0.612***
	(0.090)	(0.098)	(0.095)	(0.093)	(0.108)	(0.120)	(0.111)	(0.108)
L.Cnsent	2.542***	2.528***	2.486***	2.140***	2.751***	2.707***	3.122***	2.891***
	(0.328)	(0.340)	(0.337)	(0.372)	(0.319)	(0.328)	(0.329)	(0.353)
L.Hklocalsent	0.908**	1.076***	1.153***	0.952***	4.631***	4.863***	4.270***	4.572***
	(0.362)	(0.365)	(0.376)	(0.361)	(0.497)	(0.501)	(0.483)	(0.493)
localsmb	−0.118	−0.148	−0.103	−0.124	24.839***	26.603***	27.684***	26.212***
	(0.115)	(0.113)	(0.116)	(0.115)	(7.833)	(7.894)	(7.756)	(7.800)
localhml	−0.608***	−0.553***	−0.625***	−0.700***	32.125***	32.783***	30.787***	34.000***
	(0.127)	(0.131)	(0.129)	(0.134)	(9.523)	(9.497)	(9.658)	(9.627)
localrmf	0.119**	0.125**	0.115**	0.100*	−41.208***	−36.670***	−37.947***	−43.809***
	(0.053)	(0.053)	(0.054)	(0.054)	(10.766)	(10.646)	(11.249)	(11.005)
Policy×L.Globalsent		−1.570*	1.638**	2.669**		−2.242**	−0.328	−1.243
		(0.689)	(0.648)	(0.884)		(1.044)	(0.897)	(1.015)
R^2	0.075	0.08	0.076	0.078	0.15	0.16	0.16	0.15

Note. Table 3.10 reports the effects of global and local sentiment on A–H-stock returns, determined using pooled regression models. We applied Bekaert et al. (2009)'s research on world local pricing models to explain stock returns and controlled for sentiment on the global and local levels. Using a multivariate regression approach, we analysed the predictive value of global sentiment for stock excess return, where the dependent variable is the stock excess return and the primary explanatory variables are global and local investor sentiment. The sample period is from 2000 to 2016. Localsmb, Localhml and Localrmf give the local three-factor model for each stock market. *, ** and *** show significance at the 10%, 5% and 1% levels, respectively.

3.5.6 Portfolio analysis across characteristics

We examined whether the negative relationship between global sentiment and the A–H premium persisted with different firm characteristics, aiming to compare the influence of global sentiment on the A–H premium for diverse firm characteristics. We employed a two-dimensional portfolio classification approach based on high- and low-sentiment periods with further categorisation based on various firm characteristics. The classification methodology used the median value as the cut-off point for both sentiment and firm characteristics.

Table 3.11 shows that the differences in A–H-share price premiums between high and low global sentiment levels are statistically significant at the 1% level, based on a two-tailed *t*-test for portfolios sorted by firm characteristics. The global sentiment effect on the A–H premium is found to persist across firm characteristics, demonstrating the robustness of the general findings. The A–H premium was significantly larger in low-sentiment periods. A finding from analysing portfolios sorted by short-sale eligibility is that non-shortable A-shares have a larger A–H premium than shortable shares. Further, the absolute difference in the A–H premium between high- and low-sentiment periods was greater for shortable firms, providing additional insight into the impact of short-selling constraints on the A–H premium. The consistency of these findings with our regression results further reinforces the validity of the results. However, we observed little variation in the A–H premium between high- and low-sentiment periods for portfolios sorted by other characteristics. Overall, we found that the negative effect of global sentiment on the A–H premium persisted across different firm characteristics and was not mitigated or absorbed by the characteristic effect.

Table 3.11: Comparing A–H premium group means based on sentiment and firms' characteristics

Variable	Global.Sentiment	High	Low	*p*
A–H premium group means sorted by sentiment and dividend yield				
High dividend yield	0.231	0.306	−0.075	0.000
Low dividend yield	0.422	0.486	−0.064	0.000
Dif	−0.188	−0.176	0	0.000
p	0.000	0.000	0.000	0.000
A–H premium group means sorted by sentiment and trading cost (SPR)				
High SPR	0.351	0.413	−0.062	0.000
Low SPR	0.27	0.336	−0.066	0.000
Dif	0.081	0.078	0.004	0.000
p	0.000	0.000	0.000	0.000
A–H premium group means sorted by sentiment and market capitalisation				
High MVC	0.212	0.287	−0.075	0.000
Low MVC	0.442	0.505	−0.063	0.000
Dif	−0.23	−0.218	−0.012	0.000
p	0	0	0	0.000
A–H premium group means sorted by sentiment and shortability				
Non-shortable shares	0.424	0.459	−0.035	0.000
Shortable shares	0.197	0.212	−0.015	0.098
Dif	0.228	0.247	−0.019	0.000
p	0.000	0.000	0.000	0.000
A–H premium group means sorted by sentiment and A–H idiosyncratic risk				
High idiosyncratic volatility ratio	0.327	0.391	−0.064	0.000
Low idiosyncratic volatility ratio	0.326	0.401	−0.075	0.000
Dif	0.001	−0.011	0.012	0.000
p	0.948	0.223	0.000	0.000

Note. Table 3.11 reports the portfolio analysis of cross-listed A–H premiums based on a two-dimensional classification approach whereby portfolios are classified based on high- and low-sentiment periods and then categorised further based on firm characteristics. The classification methodology employed the median value as the cut-off point for both sentiment and firm characteristics.

3.6 Conclusion

In this study, we have examined the impact of global sentiment on the A–H premium in China and identified the mechanisms by which global sentiment affects A–H mispricing. We found that global sentiment negatively influenced the A–H premium, with the magnitude of the global sentiment effect (−0.003) being greater than the China and Hong Kong local sentiment effect (+0.002). Global sentiment had an effect of similar size to key financial indicators, such as percentage changes in A–H exchange rates and risk-free rates of local markets, which

emphasises its significance. The result suggests that global sentiment could be as important as fundamental economic factors in driving the A–H-share price dynamics. We further investigated the factors mediating the marginal impact of global sentiment to examine the mechanism. Our analysis indicated that the negative impact of global sentiment on A–H premiums was mitigated by the relaxation of capital ownership controls in the Chinese domestic market. This result implies that the predictive value of global sentiment for the A–H-stock price premium is influenced by home market openness. The different sensitivities of A-shares and H-shares to global sentiment appear to depend on the accessibility of the domestic market to foreign investors.

That the relationship between global sentiment and A–H mispricing is influenced by relative capital flows between the Chinese and Hong Kong markets is an intriguing finding. It appears that global sentiment plays a role in shaping investment decisions and capital allocation by global investors between these two markets. Our finding aligns with the concepts of behavioural finance, whereby non-fundamental factors such as investor sentiment can affect asset prices and market dynamics, including in emerging markets. Our analysis is strengthened by the finding of a correlation between global sentiment and the performance of A- and H-shares before and after the introduction of the QFII, RQFII and SHSC policies. Our study found that global sentiment significantly predicted the excess returns of H- but not A-shares across the sample period. This can be understood in terms of market accessibility and the investor base. The finding that A-share stock returns became more sensitive to global investor sentiment following the introduction of the RQFII and SHSC programs supports the conclusion that these policy initiatives succeeded in increasing the integration of the mainland Chinese and global

stock markets. Increased integration with international capital flows appears to have made A-shares more sensitive to shifts in global investment dynamics. This increased sensitivity is evidence that the RQFII and SHSC programs effectively opened the mainland Chinese market to foreign investors as intended. The empirical results align with expectations that increased foreign access to mainland stock markets would increase the exposure of A-shares to global market forces.

In our examination of the heterogeneous impact of global sentiment on the price premium, we found that the negative effect of global sentiment on A–H premiums was amplified when the relative arbitrage cost of H-shares was higher. This result suggests an asymmetric influence of global sentiment on A–H premiums, particularly for firms with different levels of A–H arbitrage risk. Our findings imply that the spillover effect of global sentiment can be influenced by the relative arbitrage costs of cross-listed shares and align with theoretical expectations: specifically, that higher arbitrage costs discourage investor participation in arbitrage activities, lowering market efficiency and exacerbating mispricing induced by global sentiment. This effect was more pronounced for H-shares, which are more prone to overpricing during periods of more positive global sentiment.

The findings of this study emphasize the importance of understanding the dynamics between global sentiment and the A-H premium, particularly in the context of regulatory frameworks. The introduction of policies such as QFII, RQFII, and the Shanghai-Hong Kong Stock Connect has proven to enhance market integration, which subsequently impacts the sensitivity of A and H shares to global sentiment. This suggests that policymakers should continue to focus on easing capital ownership restrictions to promote greater integration

between mainland China and international markets. Additionally, our results indicate that the varying impacts of global sentiment on A and H shares highlight the need for regulators to consider the differing perceptions and behaviors of domestic versus foreign investors. Understanding these dynamics can inform policy design that seeks to enhance market efficiency and investor protection.

In conclusion, our study has contributed valuable insights into the influence of global sentiment on A–H-stock price premiums and the mechanisms through which global sentiment affects A–H mispricing. These findings have important implications for financial practitioners aiming to comprehend the dynamics of the Chinese stock market and the cross-market spillover effects of global sentiment. Future research may address other factors that contribute to the differential sensitivity of A- and H-stock prices to global sentiment.

Bibliography

Abarbanell, J. S., & Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance*, *47*(3), 1181–1207.

Ajinkya, B. B., & Gift, M. J. (1984). Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *Journal of Accounting Research*, 425–444.

Ali, A., & Trombley, M. A. (2006). Short sales constraints and momentum in stock returns. *Journal of Business Finance & Accounting*, *33*(3–4), 587-615.

Amel-Zadeh, A., & Faasse, J. (2016, 27 November). *The information content of 10-K narratives: Comparing MD&A and footnotes disclosures*. SSRN.

Amihud, Y., & Mendelson, H. (1986). Liquidity and stock returns. *Financial Analysts Journal*, *42*(3), 43–48.

Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, *22*(1–3), 3–30.

Andrei, D., & Hasler, M. (2014). Investor attention and stock market volatility. *The Review of Financial Studies*, *28*(1), 33–72.

Arquette, G. C., Brown, W. O., Jr., & Burdekin, R. C. (2008). US ADR and Hong Kong H-share discounts of Shanghai-listed firms. *Journal of Banking & Finance*, *32*(9), 1916–1927.

Atiase, R. K., Li, H., Supattarakul, S., & Tse, S. (2005). Market reaction to multiple contemporaneous earnings signals: Earnings announcements and future earnings guidance. *Review of Accounting Studies*, *10*(4), 497–525.

Attanasio, O. P., Picci, L., & Scorcu, A. E. (2000). Saving, growth, and investment: A macroeconomic analysis using a panel of countries. *Review of Economics and Statistics*, *82*(2), 182–211.

Baek, I. M. (2006). Portfolio investment flows to Asia and Latin America: Pull, push or market sentiment? *Journal of Asian Economics*, *17*(2), 363–373.

Baginski, S. P., Conrad, E. J., & Hassell, J. M. (1993). The effects of management forecast precision on equity pricing and on the assessment of earnings uncertainty. *Accounting Review*, 913–927.

Baginski, S. P., & Hassell, J. M. (1997). Determinants of management forecast precision. *Accounting Review*, 303–312.

Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2004). Why do managers explain their earnings forecasts? *Journal of Accounting Research*, *42*(1), 1–29.

Bailey, W. (1994). Risk and return on China's new stock markets: Some preliminary evidence. *Pacific-Basin Finance Journal*, *2*(2–3), 243–260.

Bailey, W., & Jagtiani, J. (1994). Foreign ownership restrictions and stock prices in the Thai capital market. *Journal of Financial Economics*, *36*(1), 57–87.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross‐section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680.

Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, *104*(2), 272–287.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818.

Barron, O. E., Kile, C. O., & O'Keefe, T. B. (1999). MD&A quality as measured by the SEC and analysts' earnings forecasts. *Contemporary Accounting Research*, *16*(1), 75–109.

Bekaert, G., Hodrick, R. J., & Zhang, X. (2009). International stock return comovements. *The Journal of Finance*, *64*(6), 2591–2626.

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

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*(Supplement), 1–48.

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, *13*(4), 305–340.

Bochkay, K., & Levine, C. (2013). *Using MD&A to improve earnings forecasts*. SSRN.

Bodurtha, J. N., Jr., Kim, D. S., & Lee, C. M. (1995). Closed-end country funds and US market sentiment. *The Review of Financial Studies*, *8*(3), 879–918.

Bozanic, Z., Hoopes, J. L., Thornock, J. R., & Williams, B. M. (2017). IRS attention. *Journal of Accounting Research*, *55*(1), 79–114.

Brav, A., & Heaton, J. B. (2002). Competing theories of financial anomalies. *The Review of Financial Studies*, *15*(2), 575–606.

Brennan, M. J., & Cao, H. H. (1997). International portfolio investment flows. *The Journal of Finance*, *52*(5), 1851–1880.

Brown, S. J., Goetzmann, W. N., Hiraki, T., Shiraishi, N., & Watanabe, M. (2003). *Investor sentiment in Japanese and US daily mutual fund flows* (Working Paper 9470). National Bureau of Economic Research. https://www.nber.org/papers/w9470

Bryan, S. H. (1997). Incremental information content of required disclosures contained in management discussion and analysis. *Accounting Review*, 285–301.

Cao, Z., & Narayanamoorthy, G. (2005). *The effect of litigation risk on management earnings forecasts* (Working paper). Yale School of Management.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Chakravarty, S., Sarkar, A., & Wu, L. (1998). Information asymmetry, market segmentation and the pricing of cross-listed shares: Theory and evidence from Chinese A and B shares. *Journal of International Financial Markets, Institutions and Money*, *8*(3–4), 325–356.

Chan, K., Hameed, A., & Lau, S. T. (2003). What if trading location is different from business location? Evidence from the Jardine Group. *The Journal of Finance*, *58*(3), 1221–1246.

Chan, W. S. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*, *70*(2), 223–260.

Chemmanur, T., & Yan, A. (2009). Product market advertising and new equity issues. *Journal of Financial Economics*, *92*(1), 40–65.

Chen, C. X., & Rhee, S. G. (2010). Short sales and speed of price adjustment: Evidence from the Hong Kong stock market. *Journal of Banking & Finance*, *34*(2), 471–483.

Chen, J. P., Chen, C. S., & Su, X. (2001). Is accounting information value relevant in the emerging Chinese stock market? *Journal of International Accounting Auditing and Taxation*, *10*, 1–22.

Cheng, K. X. (2000). The microstructure of the Chinese stock market. *China Economic Review*, *11*, 79–97.

Chi, L., Zhuang, X., & Song, D. (2012). Investor sentiment in the Chinese stock market: An empirical analysis. *Applied Economics Letters*, *19*(4), 345–348.

Clarkson, P. M., Kao, J. L., & Richardson, G. D. (1999). Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package. *Contemporary Accounting Research*, *16*(1), 111–134.

Courtis, J. K. (1986). An investigation into annual report readability and corporate risk return relationships. *Accounting and Business Research*, 285–294.

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

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, 52(3), 1035-1058.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under‐ and overreactions. *The Journal of Finance*, *53*(6), 1839–1885.

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, *29*(3), 845–868.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738.

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

Demers, E., & Vega, C. (2008). *Soft information in earnings announcements: News or noise?* (International Finance Discussion Papers 951). Board of Governors of the Federal Reserve System (US). https://ideas.repec.org/p/fip/fedgif/951.html

Demers, E., & Vega, C. (2014). *The impact of credibility on the pricing of managerial textual content*. SSRN.

Diamond, D. W., & Verrecchia, R. E. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, *18*(2), 277–311.

Ding, R., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money*, *37*, 12-26.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research*, *50*(4), 1001–1040.

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

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, *9*(3), 167–175.

Elliott, W. B., Rennekamp, K. M., & White, B. J. (2015). Does concrete language in disclosures increase willingness to invest? *Review of Accounting Studies*, *20*, 839–865.

Elrod, G. B. (2009). *Is there predictive value in the words managers use? A key word analysis of The Annual Report's Management Discussion and Analysis* [Doctoral dissertation, University of Texas at Arlington]. UTA Libraries Research Commons. https://rc.library.uta.edu/uta-ir/handle/10106/4854

Engelberg, J. (2008). *Costly information processing: Evidence from earnings announcements*. AFA 2009 San Francisco Meetings Paper.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, *66*(1), 67–97.

Ettredge, M., Huang, Y., & Zhang, W. (2013). Restatement disclosures and management earnings forecasts. *Accounting Horizons*, *27*(2), 347–369.

Eun, C. S., & Janakiramanan, S. (1986). A model of international asset pricing with a constraint on the foreign equity ownership. *The Journal of Finance*, *41*(4), 897–914.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023–2052.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, *15*, 915–953.

Fischer, P. (2001). *Drift as an evolutionary outcome* (Working Paper). Penn State University.

Foster, G., Olsen, C., & Shevlin, T. (1984). Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review*, 59(4), 574-603.

Francis, J., Schipper, K., & Vincent, L. (2002). Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review*, *77*(3), 515–546.

Frazzini, A. (2006). The disposition effect and underreaction to news. *The Journal of Finance*, *61*(4), 2017–2046.

Froot, K. A., & Dabora, E. M. (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics*, *53*(2), 189–216.

Gagnon, L., & Karolyi, G. A. (2010). Multi-market trading and arbitrage. *Journal of Financial Economics*, *97*(1), 53–80.

Gao, X., Yu, J., & Yuan, Y. (2012). *Investor sentiment and idiosyncratic volatility puzzle* (Working Paper).

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, *68*(3), 1267–1300.

Garfinkel, J. A. (2009). Measuring investors' opinion divergence. *Journal of Accounting Research*, *47*(5), 1317–1348.

Gervais, S., Kaniel, R., & Mingelgrin, D. H. (2001). The high-volume return premium. *The Journal of Finance*, *56*(3), 877–919.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1), 71–100.

Grossmann, A., Ozuna, T., & Simpson, M. W. (2007). ADR mispricing: Do costly arbitrage and consumer sentiment explain the price deviation? *Journal of International Financial Markets, Institutions and Money*, *17*(4), 361–371.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408.

Gultekin, M. N., Gultekin, N. B., & Penati, A. (1989). Capital controls and international capital market segmentation: The evidence from the Japanese and American stock markets. *The Journal of Finance*, *44*(4), 849–869.

Gutierrez, R. C., Jr., & Kelley, E. K. (2008). The long-lasting momentum in weekly returns. *The Journal of Finance*, *63*(1), 415–447.

Hales, J., Kuang, X. I., & Venkataraman, S. (2011). Who believes the hype? An experimental examination of how language affects investor judgments. *Journal of Accounting Research*, *49*(1), 223–255.

Hassell, J. M., & Hillison, W. A. (2000). Voluntary causal disclosures: Tendencies and capital market reaction. *Review of Quantitative Finance and Accounting*, *15*(4), 371–389.

Han, X., & Li, Y. (2017). Can investor sentiment be a momentum time-series predictor? Evidence from China. *Journal of Empirical Finance, 42*, 212–239.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973), 45*(4), 363–407.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1-3), 337-386.

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

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies, 1*(1), 35–73.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance, 55*(1), 265–295.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance, 54*(6), 2143–2184.

Hou, K., Xiong, W., & Peng, L. (2009). *A tale of two anomalies: The implications of investor attention for price and earnings momentum*. SSRN.

Hudson, Y., & Green, C. J. (2015). Is investor sentiment contagious? International sentiment and UK equity returns. *Journal of Behavioral and Experimental Finance, 5*, 46–59.

Hutton, A. P., Miller, G. S., & Skinner, D. J. (2003). The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research, 41*(5), 867–890.

Hutton, A., & Stocken, P. (2007). *Effect of reputation on the credibility of management forecasts* (Working Paper). Boston College and Dartmouth College.

Hwang, B. H. (2011). Country-specific sentiment and security prices. *Journal of Financial Economics, 100*(2), 382–401.

Ikenberry, D., Lakonishok, J., & Vermaelen, T. (1995). Market underreaction to open market share repurchases. *Journal of Financial Economics, 39*(2–3), 181–208.

Ikenberry, D. L., & Ramnath, S. (2002). Underreaction to self-selected news events: The case of stock splits. *The Review of Financial Studies*, *15*(2), 489–526.

Jiang, G., Lee, C. M., & Zhang, Y. (2005). Information uncertainty and expected returns. *Review of Accounting Studies, 10*, 185–221.

Jones, M. J., & Shoemaker, P.A. (1994). Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature, 13*, 142.

Jorion, P., & Schwartz, E. (1986). Integration vs. segmentation in the Canadian stock market. *The Journal of Finance*, *41*(3), 603–614.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kang, J. K. (1997). Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics*, *46*(1), 3–28.

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*(1–3), 105–231.

Kothari, S. P., Shu, S., & Wysocki. P. (2005). *Do managers withhold bad news?* (Working Paper). MIT Sloan School of Management.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 1315–1335.

Lamont, O., & Stein, J. C. (2004). Aggregate short interest and market valuations. *American Economic Review*, *94*, 29–32.

Lawrence, A., Ryan, J., & Sun, Y. (2012). *Investors' demand for sell-side research: SEC filings, media coverage, and market factors* (Working Paper). Haas School of Business.

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, *19*(4), 1499–1529.

Lev, B., & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting Researc*h, *31*(2), 190–215.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247.

Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102.

Loh, R. K. (2010). Investor inattention and the underreaction to stock recommendations. *Financial management*, *39*(3), 1223–1252.

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

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), 307–326.

Loughran, T., & McDonald, B. (2017). The use of EDGAR filings by investors. *Journal of Behavioral Finance, 18*(2), 231–248.

Loughran, T., & Ritter, J. R. (1995). The new issues puzzle. *The Journal of Finance, 50*(1), 23–51.

Ma, X. (1996). Capital controls, market segmentation and stock prices: Evidence from the Chinese stock market. *Pacific-Basin Finance Journal, 4*(2–3), 219–239.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance, 42*(3), 483–510.

Michaely, R., Thaler, R. H., & Womack, K. L. (1995). Price reactions to dividend initiations and omissions: Overreaction or drift? *The Journal of Finance, 50*(2), 573–608.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance, 32*(4), 1151–1168.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6), 2107-2143.

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. *The Journal of Finance, 49*(3), 923–950.

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports, 3*, Article 1801.

Morgan, A. R. (2009). *The reliability of forward-looking statements in the MD&As Fortune 500 companies* (UMI No. 3390445) [Doctoral dissertation, Nova Southeastern University]. ProQuest Dissertations and Theses Global.

Nagar, V., Nanda, D., & Wysocki, P. (2003). Discretionary disclosure and stock-based incentives. *Journal of Accounting and Economics*, *34*, 283–309.

Ng, J., Tuna, I., & Verdi, R. (2013). Management forecast credibility and underreaction to news. *Review of Accounting Studies*, *18*(4), 956–986.

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *The Quarterly Journal of Economics*, *111*(4), 1135–1151.

Porta, R. L., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, *106*(6), 1113–1155.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, Article 1684.

Pritamani, M., & Singal, V. (2001). Return predictability following large price changes and information releases. *Journal of Banking & Finance*, *25*(4), 631–656.

Rogers, J. L., & Stocken, P. C. (2005). Credibility of management forecasts. *The Accounting Review*, *80*(4), 1233–1260.

Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance*, *14*(5), 590–610.

Shi, Y., Liu, W. M., & Ho, K. Y. (2016). Public news arrival and the idiosyncratic volatility puzzle. *Journal of Empirical Finance*, *37*, 159–172.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends?

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, *70*(5), 1903–1948.

Stulz, R. M. (1999). *International portfolio flows and security markets* (Working Paper No. 99-3). SSRN.

Stulz, R. M., & Wasserfallen, W. (1995). Foreign equity investment restrictions, capital flight, and shareholder wealth maximization: Theory and evidence. *The Review of Financial Studies*, *8*(4), 1019–1057.

Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *Journal of Business Communication*, *30*, 49–61.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Tetlock, P. C., Saar‐Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437–1467.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In G. Wright (Ed.), *Behavioral decision making* (pp. 25–41). Springer.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In B. Karpak & S. Zionts (Eds.), *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81–126). Springer.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics*, *82*(1), 103–133.

Welagedara, V., Deb, S. S., & Singh, H. (2017). Investor attention, analyst recommendation revisions, and stock prices. *Pacific-Basin Finance Journal*, *45*, 211–223.

Williams, P. A. (1996). The relation between a prior earnings forecast by management and analyst response to a current management forecast. *Accounting Review*, 103–115.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, *51*(1), 137–167.

You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting studies*, 14, 559-586.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, *61*(1), 105–137.

Zweig, M. E. (1973). An investor expectations stock price predictive model using closed-end fund premiums. *The Journal of Finance*, *28*(1), 67–78.

Appendix A

Appendix to Chapter 1

Table A.1

Variable definitions

Variable	Definition
$\mathrm{AGSV}i,t$	Abnormal investor attention as the aggregate Google Search index (SVI) in month *t* for stock *i* minus the median SVI in the previous three months
$\mathrm{Analysts}_{i,t}$	Analyst coverage as the number of analysts active at the previous month's end
$\mathrm{BHAR}_{i,t,N}$	Buy-and-hold abnormal return, obtained by first calculating the buy-and-hold return of stock *i* then subtracting the buy-and-hold return over the same horizon for a matched risk-adjusted portfolio with benchmark returns for firms matched on size, book-to-market ratio and momentum, following DGTW (1997). N represents the holding period following the formation month *t* and has a value between 2 and 12.
$\mathrm{DGTW}_{i,t}$	The stock's benchmark portfolio monthly return, calculated using the characteristic-based benchmarks of Daniel et al. (1997)
$\mathrm{Disp}_{i,t}$	Analyst forecast dispersion for the previous month, scaled by the absolute value of the analysts' mean forecast
$\mathrm{GrProf}_{i,t}$	Revenue minus the cost of goods sold, scaled by assets
$\mathrm{Idio}_{i,t}$	The standard deviation of daily abnormal returns over the 22 calendar days before the month-end day
$\mathrm{IOR}_{i,t}$	The percentage of shares held by institutional holders, as reported quarterly by 13Filings
$\mathrm{LnBM}_{i,t}$	Logarithm of the book-to-market ratio, with the book value taken for June of the previous year
$\mathrm{LnME}_{i,t}$	Logarithm of the market capitalisation of stock *i* in month *t* (in millions of dollars)
$\mathrm{News\_Coverage}_{i,t}$	The volume of public news in month *t*
$\mathrm{News\_Negative}_{i,t}$	The volume of public negative news in month *t* for stock *i*
$\mathrm{News\_Positive}_{i,t}$	The volume of public positive news in month *t* for stock *i*
$\mathrm{News\_Sentiment}_{i,t}$	The average news sentiment score in month *t* for stock *i*
$Sir_{i,t}$	Short interest ratio as shares sold short divided by shares outstanding in month *t*
$\mathrm{Vol}_{i,t}$	Trading volume as the number of shares traded in month *t* for stock *i* (in thousands)

Appendix B

Appendix to Chapter 2

Table B.1

Variable definitions

Variable	Definition
Altman Z-score	Z-score following Altman (1968), calculated for year *t* as 1.2 × working capital / total assets – 1.4 × retained earnings / total assets + 3.3 × operating income / total assets + 0.6 × market value of equity / total liabilities + sales / total assets, with missing values of working capital and retained equity replaced by 0
Avredgar	Average of daily EDGAR searching IPs over the event window [0, +15]
B/M ratio	Book-to-market ratio as the ratio of the book value of equity to the market value at the end of the previous fiscal year *t*
Bad news	Following the methodology proposed by DellaVigna and Pollet (2009), Bad news takes a value of 1 for Quantiles 1 to 3 (lowest three) and 0 otherwise.
D.Fog	Decile rank of the fog index in the MD&A section
D.MF	Decile ranks of the management forecast surprise
D.Tonechg	Decile rank of Tonechg in the MD&A section
D.Uncertainty	Decile ranks of Uncertainty in the MD&A section
D.Wordcount	Decile ranks of Word counts in the MD&A section
Earning persistence	First-order autocorrelation coefficient of quarterly earnings per share (split-adjusted) for the previous four years with a minimum of four observations
Earning surprise	The difference between the actual earnings per share (EPS) and the consensus analyst forecast for EPS before the earnings announcement. This difference is adjusted for the pre-earnings-announcement share price.
Fog index	A measure of text complexity. The fog index gives a proxy for how many years of schooling a person with average intelligence would need to comprehend the text. We used Loughran and McDonald's (2011) dictionary of words relevant in a financial setting.
Forecast precision	Management forecast precision. We employed the negative of the forecast width, whereby the width of a range forecast is the magnitude of the interval. For point forecasts, the forecast width is zero.
Good news	Following the methodology proposed by DellaVigna and Pollet (2009), Good news takes a value of 1 for Quantiles from 9 to 11 (highest three) and 0 otherwise.
Herfindahl–Hirschman index	A measure of market concentration for an industry based on SIC codes
Hitech indicator	A dummy variable taking a value of 1 for a firm in a high-technology industry based on standard industrial classification (SIC) codes of 2833–2836 (drugs), 8731–8734 (R&D services), 7371–7379 (programming), 3570–3577 (computers) or 3600–3674 (electronics) and 0 otherwise

Variable	Definition
Institutional ownership	The percentage of shares owned by institutions at the end of the most recent calendar quarter
Litigation	A dummy variable taking a value of 1 if a firm is in a high-litigation industry based on SIC codes of 2833–2836 (biotechnology), 3570–3577 (computer hardware), 3600–3674 (electronics), 5200–5961 (retail) or 7371–7379 (computer software), and 0 otherwise
logAbEdgar	The natural logarithm of the abnormal search volume. The daily abnormal search volume was calculated by subtracting the raw EDGAR search volume on day *t* from the average raw EDGAR search volume for the same day of Week *k* over the previous 10 weeks. To normalise the distribution and obtain the average of the daily abnormal search volume over a half-month window, we followed Da et al. (2011) and used the natural logarithm of AbEdgar + 1.
LogAnalyst	The logarithm of 1 plus the number of analysts following the firm in the most recent quarter
Loghorizon	The natural logarithm of 90 plus the number of days between the forecast announcement date and the end of the fiscal year
MF	Management forecast surprise
Momentum	The 12-month cumulative raw return ending one month before the month of the management forecast
Ret Volatility	The standard deviation of the daily return from Day *t-255* to *t-5*
Size	The market value of equity at the starting month of the management forecast date (in millions of dollars)
Tonechg	The percentage change in net optimism from the previous to the current year. Net optimism is defined as the difference between the number of positive and negative words in the MD&A section, divided by the total word count. We used Loughran and McDonald's (2011) dictionary of words relevant in a financial setting.
Turnover	The average daily volume from Day *t-255* to *t-5*
Uncertainty	The percentage of uncertain words in the MD&A section. We used Loughran and McDonald's (2011) dictionary of words relevant in a financial setting.
Wordcount	The total number of words in the MD&A section as a measure of text readability. We used Loughran and McDonald's (2011) dictionary of words relevant in a financial setting.

MD&A Overview Samples

Example 1: eBay Inc.

ITEM 7: MANAGEMENT'S DISCUSSION AND ANALYSIS OF FINANCIAL CONDITION AND RESULTS OF OPERATIONS

This Annual Report on Form 10-K contains forward-looking statements within the meaning of Section 27A of the Securities Act of 1933 and Section 21E of the Securities Exchange Act of 1934, including statements that involve expectations, plans or intentions (such as those relating to future business, future results of operations or financial condition, new or planned features or services, or management strategies). You can identify these forward-looking statements by words such as “may,” “will,” “would,” “should,” “could,” “expect,” “anticipate,” “believe,” “estimate,” “intend,” “plan” and other similar expressions. These forward-looking statements involve risks and uncertainties that could cause our actual results to differ materially from those expressed or implied in our forward-looking statements. Such risks and uncertainties include, among others, those discussed in “Item 1A: Risk Factors” of this Annual Report on Form 10-K, as well as in our consolidated financial statements, related notes, and the other information appearing elsewhere in this report and our other filings with the SEC. We do not intend, and undertake no obligation, to update any of our forward-looking statements after the date of this report to reflect actual results or future events or circumstances. Given these risks and uncertainties, readers are cautioned not to place undue reliance on such forward-looking statements. You should read the following Management's Discussion and Analysis of Financial Condition and Results of Operations in conjunction with the consolidated financial statements and the related notes included in this report.

You should read the following "Management's Discussion and Analysis of Financial Condition and Results of Operations" in conjunction with the audited consolidated financial statements and the related notes that appear elsewhere in this report.

Overview

eBay Inc. is a global commerce leader, including our Marketplace, StubHub and Classifieds platforms. Our Marketplace platforms include our online marketplace located at www.ebay.com, its localized counterparts and the eBay mobile apps. Our StubHub platforms include our online ticket platform located at www.stubhub.com and the StubHub mobile apps. Our Classifieds platforms include a collection of brands such as Mobile.de, Kijiji, Gumtree, Marktplaats, eBay Classifieds and others.

On July 17, 2015, we completed the distribution of 100% of the outstanding common stock of PayPal Holdings, Inc. ("PayPal") to our stockholders (the "Distribution"), pursuant to which PayPal became an independent company. Beginning in the third quarter of 2015, PayPal's financial results for periods prior to the Distribution have been reflected in our consolidated statement of income, retrospectively, as discontinued operations. Additionally, the related assets and liabilities associated with the discontinued operations in the prior year consolidated balance sheet are classified as discontinued operations. Pursuant to the terms of the separation and distribution agreement entered into between us and PayPal on June

26, 2015, upon Distribution, assets related to the PayPal business were transferred to, and liabilities related to the PayPal business were retained or assumed by, PayPal.

During the second quarter of 2015, our Board of Directors ("Board") approved a plan to sell the businesses underlying our former Enterprise segment ("Enterprise"). As a result, the Enterprise financial results were reflected in our consolidated statement of income, retrospectively, as discontinued operations beginning in the second quarter of 2015. On July 16, 2015, we signed a definitive agreement to sell Enterprise and on November 2, 2015, the sale closed. As a result, the related assets and liabilities associated with the discontinued operations in the prior year consolidated balance sheet are classified as discontinued operations.

In January 2015, at a regular meeting of our Board, our Board approved a plan to implement a strategic reduction of our existing global workforce. The reduction was completed in the first half of 2015 and is expected to generate annual savings of more than $150 million, primarily impacting sales and marketing and product development expenses. The savings in these line items are expected to be offset by additional expenses as we reinvest back into these areas of the business to drive additional growth.

We have incurred significant costs in connection with the Distribution and sale of Enterprise. These costs relate primarily to third-party advisory and consulting services, retention payments to certain employees, incremental stock-based compensation and other costs directly related to the Distribution and the sale of Enterprise. The majority of these costs are reflected in our consolidated statement of income as discontinued operations for all periods presented. The remaining costs reflected in continuing operations are classified within cost of net revenues, sales and marketing, product development or general and administrative in our consolidated statement of income as applicable.

Example 2: Willbros-Group, INC.

ITEM 7: MANAGEMENT'S DISCUSSION AND ANALYSIS OF FINANCIAL CONDITION AND RESULTS OF OPERATIONS (In thousands, except share and per share amounts or unless otherwise noted)

The following discussion and analysis should be read in conjunction with our consolidated financial statements and the notes thereto included elsewhere in this Form 10-K. Additional sections in this Form 10-K which should be helpful to the reading of our discussion and analysis include the following: (i) a description of our services provided, by segment found in Items 1 and 2 “Business and Properties”—”Services Provided” (ii) a description of our business strategy found in Items 1 and 2 “Business and Properties”—”Our Strategy”; and (iii) a description of risk factors affecting us and our business, found in Item 1A “Risk Factors.”

In as much as the discussion below and the other sections to which we have referred you pertain to management’s comments on financial resources, capital spending, our business strategy and the outlook for our business, such discussions contain forward-looking statements. These forward-looking statements reflect the expectations, beliefs, plans and objectives of management about future financial performance and assumptions underlying management’s judgment concerning the

matters discussed, and accordingly, involve estimates, assumptions, judgments and uncertainties. Our actual results could differ materially from those discussed in the forward-looking statements. Factors that could cause or contribute to any differences include, but are not limited to, those discussed below and elsewhere in this Form 10-K, particularly in Item 1A "Risk Factors" and in "Forward-Looking Statements."

OVERVIEW

Willbros is a global provider of engineering and construction services to the oil, gas, refinery, petrochemical and power industries with a focus on infrastructure such as oil and gas pipeline systems, electric T&D systems and refining and processing plants. Our offerings include engineering, procurement and construction (either individually or as an integrated EPC service offering), turnarounds, maintenance and other specialty services.

2011 Year in Review

During 2011, revenue increased $489,968 to $1,615,040 from $1,125,072 in 2010. The increase in revenue was primarily a result of a full year of our Utility T&D segment operations versus only a half year in 2010. Our Oil & Gas segment, which expanded its regional presence and breadth of services to include oilfield services work and was also successful in converting engineering assignments into full EPC projects, also increased its top line by over 17 percent. The revenue increase was partially offset by a decline in certain activity, primarily related to turnaround project delays. Operating income decreased $163,881 to a loss of $184,722 in 2011 compared to a loss of $20,841 in 2010. The variance is largely explained by two unfavorable non-cash items in 2011 – a $118,575 increase in goodwill impairment charges and a $35,340 decrease in the benefit of reducing the InfrastruX acquisition's contingent earnout liability. We believe we have made substantial progress in improving our results in the seasonally slower fourth and first quarters and our challenge is to bring lagging business units up to the level of performance of our best businesses. We will continue to take actions to improve upon these results. See the following Financial Summary included in this Item 7 for additional details on the 2011 financial performance.

As part of our 2011 debt reduction process, we completed a study to determine the strategic fit and potential future contributions of various business units. We identified certain businesses that were not producing results at a level comparable to our peer group companies' businesses. We are taking actions to improve the performance of these lagging businesses through a combination of initiatives consisting of a realignment of management, reduction of indirect costs (including fleet management) and general and administrative costs, and a more robust sales and marketing program to expand the revenue base. In instances where our evaluation identifies disadvantaged businesses, we are prepared to make divestiture decisions, when our analysis dictates that is the correct business decision. Specific targets of this profitability initiative are our Utility T&D segment's distribution businesses in the Northeast and South Central regions and certain businesses within our Oil & Gas segment which, as previously mentioned, has seen continued delay in maintenance and turnaround awards and a diminishing customer base as the refining industry rationalizes capacity.

Also, as part of this study, we identified several non-strategic businesses and have divested, or are in the process of divesting, these businesses to strengthen the balance sheet and to allow management to better focus on the remaining strategic businesses. In October 2011, we sold InterCon Construction Inc. ("InterCon") (acquired in the InfrastruX transaction). In connection with this sale, we received total compensation of $18,749 in cash and $250 in the form of an escrow deposit from the buyer. We continue to work towards a sale of our Canada cross-country pipeline construction business and are evaluating divestiture of several other businesses.

The integration of InfrastruX is advancing and we have introduced our management systems and processes. We are seeing positive results in Texas transmission construction and expect this trend to continue in 2012. Our Northeast transmission business has successfully executed multiple large capital projects in the Northeast and we expect to leverage this recent

relevant experience as we focus more management resources on this component of the Utility T&D segment. The distribution portion of the Utility T&D segment is still constrained by the dearth of housing starts, but we expect slow improvement throughout 2012, with upside from expanding gas distribution service opportunities.

Some positive highlights for 2011 include:

- We have reduced our $300,000 term loan facility ("Term Loan") (used for the InfrastruX acquisition) by making payments of $123,379 against the existing balance.

- The *Oil & Gas* segment has been able to increase year-over-year revenue by $131,872, or 17.5 percent, by expanding its regional presence in the liquids basins and shale plays complementing our cross-country large diameter pipeline revenue.

- On March 29, 2012, we entered into a settlement agreement with WAPCo to settle the WAGP project litigation. The settlement stops the significant legal spend on the litigation, including any additional expense for trial and likely appeal, creates certainty of the outcome and eliminates trial risk. We believe this is a fair settlement and are pleased to remove the overhang of this litigation.

- Engineering services rebounded, with significant additions of EPC assignments.

- The TransCanada Pipelines, Ltd. ("TransCanada") billing dispute was settled in June 2011 for $61,000, achieving recovery of approximately 90 percent of the claimed amounts.

To sum up, in 2011, we successfully advanced our strategy to bring more recurring services into our revenue base, successfully responded to a shift in our legacy upstream oil and gas markets, identified and elevated the pipeline integrity opportunity, regained traction in our engineering businesses within our *Oil & Gas* segment, booked new meaningful EPC contracts, established a new oil sands focused model in our *Canada* segment with a high graded management team, generated positive project margins in our electric transmission construction business and isolated non-performing and non-strategic businesses for additional management actions.

Appendix C

Appendix to Chapter 3

The Hang Seng A–H premium index measure was determined using Equation C.1:

$$HSAHP\ Index = \frac{\sum[(IS^A \times FAF^A) + (IS^H \times FAF^H)] \times P^A}{\sum[(IS^A \times FAF^A) + (IS^H \times FAF^H)] \times P^H} \times 100 \quad \text{(C.1)}$$

where P^A is the current A-share price (in US dollars), P^H is the current H-share price (in US dollars), IS^A is the number of A-shares issued, IS^H is the number of H-shares issued, FAF^A is the free float–adjusted factor for A-shares (which lies between 0 and 1) and FAF^H is the free float–adjusted factor for H-shares (which lies between 0 and 1).

Table C.1

Cross-listed companies included in the sample period

STKCD	CONME	INDNME
000002	CHINA VANKE CO., LTD	Properties
000039	CHINA INTERNATIONAL MARINE CONTAINERS (GROUP) CO., LTD	Industrials
000063	ZTE CORPORATION	Industrials
000157	ZOOMLION HEAVY INDUSTRY SCIENCE AND TECHNOLOGY CO., LTD.	Industrials
000338	WEICHAI POWER CO., LTD.	Industrials
000488	SHANDONG CHENMING PAPER HOLDINGS LTD.	Industrials
000513	LIVZON PHARMACEUTICAL (GROUP) INC.	Industrials
000585	NORTHEAST ELECTRIC DEVELOPMENT CO., LTD.	Industrials
000666	JINGWEI TEXTILE MACHINERY CO., LTD.	Industrials
000756	SHANDONG XINHUA PHARMACEUTICAL CO., LTD.	Industrials
000898	ANGANG STEEL COMPANY LIMITED	Industrials
000921	HISENSE KELON ELECTRICAL HOLDINGS COMPANY LIMITED	Industrials
002202	XINJIANG GOLDWIND SCIENCE&TECHNOLOGY CO.,LTD	Industrials
002490	SHANDONG MOLONG PETROLEUM MACHINERY CO. LTD.	Industrials
002594	BYD CO., LTD	Industrials
002672	DONGJIANG ENVIRONMENTAL COMPANY LIMITED	Public Utility
002703	ZHEJIANG SHIBAO COMPANY LIMITED	Industrials
600011	HUANENG POWER INTERNATIONAL CO., LTD	Industrials
600012	ANHUI EXPRESSWAY CO., LTD	Public Utility
600016	CHINA MINSHENG BANKING CO., LTD.	Finance
600026	CHINA SHIPPING DEVELOPMENT CO., LTD	Public Utility
600027	HUADIAN POWER INTERNATIONAL CO., LTD.	Industrials
600028	CHINA PETROLEUM & CHEMICAL CORPORATION	Industrials
600029	CHINA SOUTHERN AIRLINES CO., LTD	Public Utility
600030	CITIC SECURITIES CO., LTD	Finance
600036	CHINA MERCHANTS BANK CO., LTD	Finance
600115	CHINA EASTERN AIRLINES CO., LTD.	Public Utility
600188	YANZHOU COAL MINING CO., LTD.	Industrials
600196	SHANGHAI FOSUN PHARMACEUTICAL (GROUP) CO., LTD.	Industrials
600332	GUANGZHOU BAIYUNSHAN PHARMACEUTICAL HOLDINGS COMPANY LIMITED	Industrials
600362	JIANGXI COPPER CO., LTD.	Industrials
600377	JIANGSU EXPRESSWAY CO., LTD	Public Utility
600548	SHENZHEN EXPRESSWAY CO., LTD	Public Utility
600585	ANHUI CONCH CEMENT CO.,LTD	Industrials
600600	TSINGTAO BREWERY CO., LTD.	Industrials
600685	CSSC OFFSHORE & MARINE ENGINEERING (GROUP) COMPANY LIMITED	Industrials
600688	SINOPEC SHANGHAI PETROCHEMICAL CO., LTD.	Industrials
600775	NANJING PANDA ELECTRONICS CO., LTD.	Industrials
600806	SHENJI GROUP KUNMING MACHINE TOOL CO.,LTD	Industrials
600808	MAANSHAN IRON & STEEL CO., LTD.	Industrials

STKCD	CONME	INDNME
600837	HAITONG SECURITIES COMPANY LTD	Finance
600860	Beijing Jingcheng Machinery Electric Company Limited	Industrials
600871	Sinopec Oilfield Service Corporation	Industrials
600874	TIANJIN CAPITAL ENVIRONMENTAL PROTECTIONGROUP COMPANY LIMITED	Public Utility
600875	DONGFANG ELECTRIC CORPORATION LIMITED	Industrials
600876	LUOYANG GLASS CO., LTD.	Industrials
601005	CHONGQING IRON & STEEL COMPANY LIMITED	Industrials
601038	FIRST TRACTOR COMPANY LIMITED	Industrials
601088	CHINA SHENHUA ENERGY COMPANY LIMITED	Industrials
601107	SICHUAN EXPRESSWAY COMPANY LIMITED	Public Utility
601111	AIR CHINA LIMITED	Public Utility
601186	CHINA RAILWAY CONSTRUCTION CORPORATION LIMITED	Properties
601238	GUANGZHOU AUTOMOBILE GROUP CO., LTD.	Industrials
601288	AGRICULTURAL BANK OF CHINA LIMITED	Finance
601299	CHINA CNR CORPORATION LIMITED	Industrials
601318	PING AN INSURANCE (GROUP) COMPANY OF CHINA, LTD.	Finance
601328	BANK OF COMMUNICATIONS CO., LTD.	Finance
601333	GUANGSHEN RAILWAY COMPANY LIMITED	Commerce
601336	NEW CHINA LIFE INSURANCE COMPANY LTD.	Finance
601390	CHINA RAILWAY GROUP LIMITED.	Properties
601398	INDUSTRIAL AND COMMERCIAL BANK OF CHINA LIMITED	Finance
601588	BEIJING NORTH STAR COMPANY LIMITED	Properties
601600	ALUMINUM CORPORATION OF CHINA LIMITED	Industrials
601601	CHINA PACIFIC INSURANCE (GROUP) CO., LTD.	Finance
601607	SHANGHAI PHARMACEUTICALS HOLDING CO.,LTD.	Commerce
601618	METALLURGICAL CORPORATION OF CHINA LTD.	Properties
601628	CHINA LIFE INSURANCE COMPANY LIMITED	Finance
601727	SHANGHAI ELECTRIC GROUP COMPANY LIMITED	Industrials
601766	CRRC Corporation Limited	Industrials
601808	CHINA OILFIELD SERVICES LIMITED	Industrials
601818	CHINA EVERBRIGHT BANK COMPANY LIMITED	Finance
601857	PETROCHINA COMPANY LIMITED	Industrials
601866	CHINA SHIPPING CONTAINER LINES COMPANY LIMITED	Public Utility
601880	DALIAN PORT (PDA) CO., LTD.	Public Utility
601898	CHINA COAL ENERGY COMPANY LIMITED	Industrials
601899	ZIJIN MINING GROUP CO., LTD.	Industrials
601919	CHINA COSCO HOLDINGS COMPANY LIMITED	Public Utility
601939	CHINA CONSTRUCTION BANK CORPORATION	Finance
601988	BANK OF CHINA LIMITED	Finance
601991	DATANG INTERNATIONAL POWER GENERATION CO., LTD.	Industrials
601992	BBMG CORPORATION	Industrials
601998	CHINA CITIC BANK CORPORATION LIMITED	Finance
603993	CHINA MOLYBDENUM CO., LTD.	Industrials
900948	INNER MONGOLIA YITAI COAL CO., LTD	Industrials

Table C.2

Data sources for each proxy used to construct country sentiment

Country	Parameter	Source
US	Number of IPO issues	Baker and Wurgler's website
	Equity issue ratio	SDC
	Dividend premium	Datastream
	Volatility premium	Datastream
UK	Number of IPO issues	London Stock Exchange statistics
	Equity issue ratio	SDC
	Dividend premium	Datastream
	Volatility premium	Datastream
Japan	Number of IPO issues	Kaneko and Pettway's Japanese IPO Database
	Equity issue ratio	SDC
	Dividend premium	Datastream
	Volatility premium	Datastream
Germany	Number of IPO issues	SDC
	Equity issue ratio	SDC
	Dividend premium	Datastream
	Volatility premium	Datastream
Canada	Number of IPO issues	SDC
	Consumer confidence	OECD
	Equity issue ratio	SDC
	Volatility premium	Datastream
France	Number of IPO issues	SDC
	Equity issue ratio	SDC
	Dividend premium	Datastream
China	Number of IPO issues	CSMAR
	Dividend premium	Datastream
	Volatility premium	Datastream
	Closed-end fund discount	CSMAR
Hong Kong	Number of IPO issues	SDC
	Equity issue ratio	SDC
	Dividend premium	Datastream
	Volatility premium	Datastream

Table C.3

Variable description and data source

Variable	Description	Source
A–H premium	The difference between the A-share price and exchange rate-adjusted H-share price, divided by the A-share price, for month *t*	DataStream, CSMAR
AretM	The monthly market return of China	CSMAR
Capital flow	The ratio of portfolio investment in equity and funds in China and Hong Kong. The value was standardised with a mean of 0 and a standard deviation of 1.	OECD
CNrf	Risk free of China, measured as the one-year fixed deposit rate or one-year treasury note issued by the Chinese Government	CSMAR
CNsent	Chinese local sentiment	
Dividend	Firm annual dividend	CSMAR
Dividend yield	Annual dividend payments to shareholders divided by the stock's current price	Datastream
Globalsent	Global sentiment	
HKlocalsent	Hong Kong local sentiment	
HKrf	The risk-free rate in Hong Kong as the Hong Kong Interbank Offered Rates (three months)	Datastream
HretM	The monthly market return of Hong Kong	Datastream
Idio	The relative idiosyncratic risk, measured as the ratio of A and H idiosyncratic volatility. The idiosyncratic volatility is the monthly standard deviation of residuals from the time-series regression of the CAMP model.	
IOR	The percentage of institutional investor ownership of A-shares	CSMAR
Liquidityratio	H–A liquidity ratio as the ratio of cross-listed H- to A-stock turnover	DataStream, CSMAR
MVC	Annual market capitalisation, including the market value of outstanding stocks in all listed stock exchanges	Datastream
QFII	QFII policy reform dummy variable, which takes a value of 1 once the policy is implemented	
QFIIholdingratio	The percentage of QFII institutional investor ownership of A-shares	Resset
RQFII	The RQFII policy reform dummy variable, which takes a value of 1 once the policy is implemented	
Sentdif	The sentiment difference between China and Hong Kong	
Short indicator	Short sale eligibility dummy for A-shares, which takes a value of 1 if the A-share is available for short-selling	
SHSC	Shanghai–Hong Kong Stock Connect program policy dummy, which takes a value of 1 once the policy is implemented	
SPR	H-A monthly bid–ask spread gap as the difference between cross-listed H- and A-share bid–ask spreads	Datastream
Volatilityratio	A–H volatility ratio as the ratio of the A and H monthly averages of 21-day return rolling annualised volatility	Datastream, CSMAR
worldreturn	Global market return measured as the MSCI world index return	Datastream
Macro characteristic	**Description**	**Source**
CNexvol	Monthly annualised 21-day rolling standard volatility of the US–China exchange rate	Datastream
Consumgrowth	Percentage change in consumption	Canada, France, Germany, Japan, UK, US: Federal Reserve Economic Data

Variable	Description	Source
CPI	Consumer price index	China: Oxford Economics HK: Census and Statistics Department, Hong Kong Canada, China, France, Germany, Japan, UK, US: OECD
DeltaexHK	Monthly percentage change of Hong Kong–China exchange rate	HK: Census and Statistics Department, Hong Kong Datastream
Emgrowth	Monthly employment percentage change	Datastream
HKconfidence	Hong Kong consumer confidence index	Chinese University of Hong Kong
Industgrowth	Industry production percentage change rate	OECD
RECM	Economic recession indicator	Federal Reserve Economic Data
Termprem	Term premium as the difference between the 10-year and three-month treasury bill yields in each country	OECD, Thomson Reuters
Trearybill3	Three-month treasury bill interest rate	US: Federal Reserve Canada, China, France, Germany, HK, Japan, UK: Thomson Reuters