ONLINE SOCIAL INTERACTIONS AND FINANCIAL MARKETS

MUHAMMET SERGEN AKARSU

BOĞAZİÇİ UNIVERSITY

2025

ONLINE SOCIAL INTERACTIONS AND FINANCIAL MARKETS

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Management

by

Muhammet Sergen Akarsu

Boğaziçi University

2025

Online Social Interactions and Financial Markets

The thesis of Muhammet Sergen Akarsu

has been approved by:

Prof. Neslihan Yılmaz
(Thesis Advisor) _____

Prof. Ahmet Vedat Akgiray _____

Prof. Arzucan Özgür Türkmen _____

Prof. Oktay Taş
(External Member) _____

Assoc. Prof. Cumhur Enis Ekinci
(External Member) _____

May 2025

DECLARATION OF ORIGINALITY

I, Muhammet Sergen Akarsu, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by the

Signature……....

Date: …….....

ABSTRACT

Online Social Interactions and Financial Markets

This dissertation examines the relationship between online social interactions and financial market outcomes by analyzing user discussions from Reddit. It consists of three empirical studies that investigate how sentiment dissemination, belief disagreement, and thematic attention shifts emerge in online communities and influence key market variables. The first study shows that a small number of highly active users drive sentiment dissemination within Reddit's stock market discussions. It also finds that network-level sentiment has short-term predictive power over stock returns. The second study focuses on belief disagreement and compares its effects across the stock and Bitcoin markets. Results indicate that disagreement is strongly associated with increases in volatility, trading volume, and downside risk in the stock market, but has a weaker and more limited effect in the Bitcoin market. The third study analyzes investor attention by classifying Reddit discussions into stock-level, market-wide, and inattentive themes. It finds that shifts to stock-level attention reduce return comovement, while shifts to market-wide or inattentive themes increase it. Collectively, the findings show that online social interactions influence how investors form beliefs, process information, and affect markets.

ÖZET

Çevrimiçi Sosyal Etkileşimler ve Finansal Piyasalar

Bu tez, çevrim içi sosyal etkileşimlerin finansal piyasalar üzerindeki etkisini, Reddit'teki kullanıcı tartışmalarını analiz ederek incelemektedir. Tez, yatırımcı duyarlılığının yayılımı, görüş farklılıkları ve tematik yatırımcı dikkati kaymaları gibi olguların çevrim içi topluluklar içerisinde nasıl ortaya çıktığını ve bunların temel piyasa değişkenleri üzerindeki etkilerini araştıran üç ampirik çalışmadan oluşmaktadır. İlk çalışma, Reddit'in hisse senetleriyle ilgili bir topluluğunda az sayıdaki yüksek derecede aktif kullanıcının yatırımcı duyarlılığının yayılımını yönlendirdiğini göstermektedir. Ayrıca, ağ düzeyinde duyarlılığın hisse senedi getirileri üzerinde kısa vadeli öngörü gücüne sahip olduğunu ortaya koymaktadır. İkinci çalışma, kullanıcılar arası görüş farklılıklarını incelemekte ve bunların hisse senedi ile Bitcoin piyasaları üzerindeki etkilerini karşılaştırmaktadır. Bulgular, hisse senedi piyasasında görüş farklılıklarının oynaklık, işlem hacmi ve aşağı yönlü risk artışı ile güçlü şekilde ilişkili olduğunu, ancak Bitcoin piyasasında bu etkinin daha zayıf ve sınırlı olduğunu göstermektedir. Üçüncü çalışma ise yatırımcı dikkatini, Reddit tartışmalarını hisse senedi odaklı, piyasa geneli odaklı ve ilgisiz temalar şeklinde sınıflandırarak analiz etmektedir. Çalışma, yatırımcı dikkati, hisse senedi odaklı temalara kaydığında hisse getirilerinin birlikte hareket etme eğiliminin azaldığını; piyasa geneli veya ilgisiz temalara kaydığında ise arttığını göstermektedir. Üç çalışma birlikte ele alındığında, çevrim içi sosyal etkileşimlerin yatırımcıların görüşlerini nasıl oluşturduklarını, bilgiyi nasıl işlediklerini ve piyasaları nasıl etkilediklerini anlamada önemli bir rol oynadığı ortaya konmaktadır.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Neslihan Yılmaz, for her invaluable guidance, support, and patience throughout the course of this dissertation. Her guidance and support greatly shaped this work.

I am also thankful to the members of my dissertation committee, Prof. Arzucan Özgür and Prof. Oktay Taş, for their constructive comments and valuable suggestions. Their perspectives helped improve this research significantly.

Special thanks to Assoc. Prof. Oğuzhan Bahadır and Dilara Gökçeoğlu for their academic support, thoughtful discussions, and the sense of community they provided.

On a personal note, I am profoundly grateful to my family and friends for their unwavering support and understanding during the challenging phases of this journey. Their encouragement kept me grounded and motivated. I would especially like to thank my wife, Sena Beşli Akarsu, and my sister, İlkay Gülsüm Dede, for their constant support throughout this journey.

Finally, I want to thank everyone who believed in me when I doubted myself. This dissertation would not have been possible without the support and kindness of those around me.

This dissertation was partially funded by Boğaziçi University Research Fund, Grant Number 19825.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: THE DYNAMICS OF ONLINE SOCIAL INTERACTIONS AND IMPLICATIONS ON STOCK MARKET RETURNS 5

2.1 Introduction 5

2.2 Related literature 8

2.3 Hypotheses development 13

2.4 Data and methodology 17

2.5 Empirical findings 30

2.6 Conclusion and discussion 44

CHAPTER 3: SOCIAL MEDIA DISAGREEMENT AND FINANCIAL MARKETS: A COMPARISON OF STOCKS AND BITCOIN 46

3.1 Introduction 46

3.2 Literature review 49

3.3 Hypotheses, data, and methodology 53

3.4 Empirical findings 60

3.5 Conclusion 76

CHAPTER 4: THEMATIC ATTENTION SHIFTS IN ONLINE DISCUSSIONS AND THEIR ROLE IN STOCK RETURN COMOVEMENT 79

4.1 Introduction 79

4.2 Literature review and hypotheses development 82

4.3 Data and methodology 88

4.4 Findings 104

4.5 Conclusion 120

CHAPTER 5: CONCLUSION 122

REFERENCES 126

LIST OF TABLES

Table 1. Number of Comments and Submissions by Year 19

Table 2. Descriptive Statistics of Social Media Data 20

Table 3. Average Number of Comments and Users by Centrality Group 26

Table 4. The Effect of High-Centrality Users on the Sentiment of Low-Centrality Users 32

Table 5. The Effect of Disagreement Levels on the Dissemination of Sentiment 34

Table 6. The Effect of Volatility Levels on the Dissemination of Sentiment 35

Table 7. Abnormal Sentiment and Returns of the S&P 500 Index 37

Table 8. The Effect of Before-Market Sentiment on Returns of the S&P 500 Index 38

Table 9. Trading Strategy Performance 40

Table 10. Alternative Estimations for the Effect of High-Centrality Users on Low-Centrality Users 42

Table 11. The Interaction between the Sentiment of High-Centrality Users and Low-Centrality Users 43

Table 12. Alternative Estimations for the Effect of Abnormal Network Sentiment on the Returns of the S&P 500 Index 43

Table 13. Descriptive Statistics 61

Table 14. Disagreement, Abnormal Trading Volume, and Volatility in the r/stocks . 63

Table 15. Disagreement, Abnormal Trading Volume, and Volatility in the r/Bitcoin 64

Table 16. The Effects of Before-Market Disagreement in the r/stocks 69

Table 17. The Effects of Before-Market Disagreement in the r/Bitcoin 70

Table 18. Disagreement and Extreme Returns 71

Table 19. Lagged Effects of Disagreement at the Individual Stock Level 73

Table 20. Disagreement and Extreme Returns at the Individual Stock Level 75

Table 21. Topics Identified by Latent Dirichlet Allocation (LDA) 90

Table 22. Descriptive Statistics of Stock Ticker Mentions 100

Table 23. The Effect of Market factors on Attention Shifts in Online Interactions 107

Table 24. Attention Shifts in Online Interactions and Return Comovement 109

Table 25. Thematic Attention Shifts in Online Interactions and Return Comovement 111

Table 26. Stock Ticker Mentions and Return Comovement 113

Table 27. Attention Categories, Ticker Mentions, and Return Comovement 115

Table 28. Attention Shifts from Alternative Classification Methods 117

Table 29. Probabilistic Attention Shifts 118

Table 30. Alternative Return Comovement Measures 119

LIST OF FIGURES

Figure 1. Interaction structure of a post on Reddit 18

Figure 2. Distribution of comments within a day 21

Figure 3. Distribution of comments by weekdays 21

Figure 4. An example of a daily network 22

Figure 5. Number of nodes over time 23

Figure 6. Number of edges over time 24

Figure 7. Network density over time 24

Figure 8. Distribution of topics over time 92

Figure 9. Attention to economic issues and the FED policy rate changes 93

Figure 10. Attention to global issues and significant geopolitical events 94

CHAPTER 1

INTRODUCTION

The increasing integration of social media into daily life has significantly affected how individuals acquire, evaluate, and act upon information. This shift is particularly evident in the context of financial markets, where Reddit, Twitter, and StockTwits have emerged as active platforms of financial discussion. These platforms enable millions of investors to interact, exchange opinions, and influence each other's behavior in real time. Therefore, opinion dynamics and crowd behavior have become more visible, as investors pay closer attention to peer opinions and shared sentiment on social media. As a result, understanding how online social interactions affect investor beliefs and financial market outcomes has become an essential topic in financial economics.

While the influence of investor sentiment on asset prices has been extensively studied, the unique role of online interactions in shaping sentiment dissemination, belief formation, and trading behavior remains underexplored. In online communities, sentiment can be affected through user interactions, network structures, and discussion dynamics. This dissertation investigates how online social dynamics influence investor beliefs and market behavior by using discussion data from Reddit.

The dissertation consists of three empirical studies that collectively examine different dimensions of online social interactions and their implications for financial markets. Each study is designed as a stand-alone paper but contributes to a unified research program on the informational and behavioral consequences of online investor communities.

The first study, presented in Chapter 2, explores the dissemination of sentiment within online social networks and its predictive value for stock market returns. Using user-level interactions on the r/stocks subreddit, the study demonstrates that active users have a great influence over low-activity users. A small number of users can exert significant influence over others through intensive comment interaction, especially when market uncertainty is high and when sentiment across users is similar. Importantly, the study shows that sentiment disseminated through the network has positive predictive power for future stock market returns over the short term. The chapter also proposes a network-based market timing strategy that reduces drawdowns while improving performance, which shows the practical value of network-informed sentiment measures.

Chapter 3 investigates investor disagreement and compares how it manifests in traditional and digital asset markets. Using data from both r/stocks and r/Bitcoin subreddits, the study constructs disagreement measures based on cross-user sentiment dispersion. The analysis reveals that higher levels of social media disagreement are positively associated with subsequent market volatility, abnormal trading volume, and the likelihood of extreme negative returns in the stock market. In contrast, disagreement in Bitcoin-related discussions has a weaker and more limited effect, influencing volatility only marginally and showing no significant impact on trading volume or extreme price movements. These findings demonstrate the differences in how online communities across asset classes process and react to information and peer effects.

The third study, in Chapter 4, examines how shifts in investor attention between stock-level, market-wide, and inattentive discussions affect stock return co-movement. Using a supervised classification approach, the study tracks attention

themes in Reddit discussions and quantifies the flow of attention across themes on a daily basis. The results show that stock-level attention increases when firms release new information or experience unusual trading activity, while market-wide themes dominate during macroeconomic events. In low-information environments, discussions often shift to off-topic or noisy content. These thematic shifts significantly affect return co-movement: stock-level attention reduces comovement, while market-wide and inattentive attention increase it. The study also finds that ticker mentions reduce comovement only when they occur in a stock-level context. These findings emphasize the importance of attention context, not just attention volume, in shaping information diffusion and market behavior.

Taken together, the three chapters provide a comprehensive view of how online social interactions shape belief dynamics and financial market outcomes**.** By analyzing online interaction networks, the content of these interactions, and how they evolve, the dissertation demonstrates mechanisms that influence sentiment dissemination, trading activity, volatility, and return comovement. Findings of studies in this dissertation have practical implications for asset pricing, portfolio management, and financial regulation. Across chapters, the dissertation shows that online social activity can be used to construct robust market timing signals and identify periods of systemic risk.

The contribution of the dissertation to the literature is multifold. First, it shows that investor sentiment and disagreement are shaped not only by individual perceptions but also by interaction dynamics within online communities. The findings highlight the importance of social influence and user roles in belief formation, which complements and extends existing work on sentiment-driven market behavior. Second, the dissertation also contributes to the literature by

showing that belief dynamics, emerging from social interactions, can serve as early indicators of market movements, especially extreme movements. Third, by comparing the stock and cryptocurrency markets, the dissertation highlights how online interactions function differently across asset classes. Finally, it provides a framework to analyze attention shifts in online discussions by classifying investor focus into thematic categories and examining how these shifts affect stock return co-movement.

The remainder of this dissertation is structured as follows. Chapter 2 presents the first study on sentiment dissemination. Chapter 3 examines online disagreement and compares its effects in stock and cryptocurrency markets. Chapter 4 investigates thematic attention shifts and their role in stock return comovement. Chapter 5 concludes the dissertation by summarizing the findings, discussing their implications, and outlining directions for future research.

CHAPTER 2

THE DYNAMICS OF ONLINE SOCIAL INTERACTIONS AND IMPLICATIONS ON STOCK MARKET RETURNS

This study examines the relationship between online social interactions, sentiment dissemination, and stock market returns using Reddit data. We find that a small number of active users significantly influence others by disseminating sentiment within their networks. Active users have a more pronounced influence on less active users when they share similar beliefs and during periods of increased uncertainty. Moreover, we show that both pre-market and prior-day network sentiment positively affect future stock market returns. Finally, we evaluate a network-based market timing strategy that effectively reduces drawdowns and highlights the practical implications of online interactions.

2.1 Introduction

Individuals derive benefits from their social networks in their decision-making process. For example, by consulting family, friends, or peers, they gain access to new information, acquire alternative perspectives, and enhance social utility through the adoption of socially acceptable choices. The literature shows that social interactions can affect people's decisions on various financial and economic decisions (Brown et al., 2008; Duflo & Saez, 2002; Hong et al., 2004, 2005). In most cases, studies on peer effects argue that the primary mechanism of influence is learning from the behavior of others. In essence, by interacting with others in social situations, people obtain a chance to observe others' behavior, gain new information, and learn about a new topic. Ultimately, they use the information they get from others in their

decision-making and make decisions like those they interact with. While the impact of social interactions has been acknowledged extensively in the literature, relatively few studies directly investigate their impacts on information dissemination and individuals' beliefs and attitudes. A small number of studies show that the interactions affect information dissemination, the enlargement of Ponzi schemes, individuals' degree of risk aversion, and the tendency to exhibit the disposition effect (Ahern et al., 2014; A. Banerjee et al., 2013; Heimer, 2016; Rantala, 2019). While social media data allows us to examine interactions more directly, we see that the literature overlooks such an approach. Studies have uncovered various aspects of social media activity, but little is known about how beliefs spread in online environments. In this study, we aim to fill this gap and contribute to the growing literature on the role of social media in financial markets. Specifically, by focusing on the dynamics of user interactions, we investigate how sentiment disseminates within online social networks and examine whether it affects market returns.

By using an extensive online interaction dataset from Reddit for the period between January 2019 and June 2022, we show that the sentiment of active social media users can predict the sentiment of less-active ones. Receivers fail to adjust to the repetitive signals of active users; therefore, a small number of active users can influence many more. Our results also reveal that the pandemic period, overall market uncertainty, and lower disagreement increase the effect of high-centrality users on low-centrality users. Finally, we show that abnormal network sentiment positively affects the next day's market returns. This effect also exists for comments made just before the market opens.

To test the practical implications of our results, we examine the performance of a market timing strategy. We show that levels of abnormal network sentiment

relate to significantly higher returns than buying and holding the S&P 500 index. Moreover, the strategy protects against market downturns. The maximum drawdown for the stock market exposure reduces to 10%, which represents a significant reduction compared to the 34% of the benchmark portfolio. We show that sentiment-based strategies can be used for hedging, though profitability depends on the ability to trade at low transaction costs. Our results are also crucial as inputs to policy-making decisions. The potential effect of network participants on each other, specifically in periods of high volatility, can bring fast-growing negative consequences. Our findings suggest that regulators and platform administrators could monitor sentiment trends and high-centrality users to detect and mitigate possible market manipulation or misinformation risks.

Our study differs from others in several ways. While some studies have examined sentiment contagion broadly by often focusing on aggregated market sentiment or general social media dynamics (Baker et al., 2012; Cookson & Niessner, 2020), we distinctly analyze interaction-driven sentiment dissemination within social media networks. Specifically, we focus on how individual users influence others in a dynamic, networked environment and emphasize the role of user centrality and repetitive signaling in shaping sentiment transmission. Unlike studies that aggregate sentiment or infer interaction indirectly, our analysis captures the relationships between active and less-active users and shows the strength of a few users in affecting others in a large network.

We also benefit from Reddit's community-based structure. Reddit's topic-specific subreddits provide a unique context for isolating financial conversations and enable us to examine the direct impact of the sentiment in market-related interactions on market outcomes.

We examine these dynamics using a dataset that spans 42 months, including the COVID-19 pandemic period. The pandemic reflects a unique time of elevated uncertainty, structural concerns, and increased online activity. That is why the pandemic period highlights how external shocks strengthen the influence of active users. Our findings suggest that online social interactions become more influential when traditional social channels are constrained.

The remainder of the chapter is structured as follows: We provide a review of the relevant literature in Section 2.2. In Section 2.3, we develop our hypotheses. We explain our data and our methodology in Section 2.4. We report our empirical findings on network dynamics and their implications for the stock market in Section 2.5. Finally, in Section 2.6, we provide concluding remarks.

2.2 Related literature

Social interactions can disseminate information, beliefs, and feelings about markets and affect decision-making. Several theoretical models (Bikhchandani et al., 1992; Ellison & Fudenberg, 1995; Han et al., 2022) show that social interactions can affect behavior through information dissemination and learning. Bikhchandani et al. (1992) argue that people gather information from other people's decisions, and an informational cascade eventually starts. Ellison and Fudenberg (1995) model word-of-mouth communication and suggest that it enables "efficient social learning," meaning people tend to choose superior actions. Han et al. (2022) propose that investors affect others' strategies by transmitting performance-based information, which becomes more influential with stronger social ties.

It is natural to expect social interactions to impact beliefs and decisions in highly uncertain markets and complex decision-making. Prior research across

various contexts confirms social effects in economic and financial behavior. For instance, Duflo and Saez (2002) argue that social interactions increase awareness of the subtle benefits of retirement plans. They examine the decisions of employees of a large university on participation in retirement plans. They find a positive relationship between an employee's probability of participating in a retirement plan and the participation rate in their department. Moreover, they find that the amount people invest tends to increase with the investment of other people in their department. In an experimental context, Duflo and Saez (2003) examine how attending a fair about retirement programs and interacting with those attending the fair affect retirement decisions. They find that social interactions increase the fair attendance rate, and that fair attendance increases participation in retirement programs. They do not find a difference in retirement decisions between those in the same department who attended the fair and those who did not. Duflo and Saez (2003) argue that social effects can explain this finding, as the information given at the fair may have spread throughout the department.

The literature has also examined the role of social interactions in the stock market, particularly their impact on the stock market participation decision. Hong et al. (2004) argue that social interactions can affect stock market participation by allowing people to learn about markets or making interactions about markets more enjoyable. They identify social households as those that interact with their neighbors or attend church. They find that social households in the USA are more likely to participate in the stock market. They also find that the impact of social interactions is more substantial in states with higher stock market participation rates. Brown et al. (2008) find that an individual's probability of investing in stocks increases with the market participation rate of their community and that this community effect becomes

more substantial as the advice-seeking behavior in a community increases. Kaustia and Knüpfer (2012) find that positive peer performance attracts new entrants in Finland's stock market. They also find asymmetric impacts of peer performance. Negative returns do not decrease the number of new entrants, which may suggest that people tend to share their wins but not their losses.

Beyond participation, social interactions can also shape investment behaviors and decisions. Shiller and Pound (1989) survey institutional and individual investors and find that social interactions and word-of-mouth communication significantly affect investment decisions. Ivković and Weisbenner (2007) find that neighbor investment preferences influence U.S. household sector choices, especially in socially active states. Building on the arguments of Han and Hirshleifer (2012), Heimer (2014) examines the effect of sociability on the trading behavior of individual investors. Han and Hirshleifer (2012) argue that active investment strategies are more involved in interactions than passive strategies due to their high variances. Moreover, active strategies are associated with more success because people tend to talk about their accomplishments more than their failures. They argue that this asymmetry leads other investors to adopt more active investment strategies. Consistent with this view, Heimer (2014) finds that social interactions affect investors' trading strategy, and the probability of being an active trader increases with more socialness. Hvide and Östberg (2015) examine the investment decisions of coworkers in Norway. Their results show that a person's monthly trading activity increases if their coworkers' activity increases that month, although without improved performance. Zhang et al. (2018) find that household, workplace, and neighborhood peers affect cash, bond, equity, and property asset allocations. The strength of household peer effects is the highest, followed by workplace and

neighborhood peer effects. They also find that age is a significant factor in how peers affect decisions. For people younger than 20 and older than 60, the household peer effects are more critical. Ouimet and Tate (2020) show that coworker participation influences employees' decisions to enroll and exit stock purchase plans.

While often associated with individual investors, social effects also influence institutional investors. For instance, Hong et al. (2005) examine the effect of word-of-mouth communication among mutual fund managers by comparing the investment decisions of fund managers who live in the same city. They find that a fund manager's level of investment in a company at a given period is significantly affected by the portfolio choices of other fund managers living in the same city. The peer effect also exists in trading decisions. A fund manager is more likely to buy a stock if their peers in the same city increase their holdings of that stock. Cohen et al. (2008) find that educational ties between fund managers and corporate insiders affect investment decisions and performance. They also find that manager changes shift portfolios to reflect new social connections. Pool et al. (2015) examine the portfolio choice of mutual fund managers living close to each other. They argue that living proximity increases the probability of engaging in face-to-face interactions and can better reflect the impact of social interactions on investment decisions. They show that neighbor fund managers hold more similar portfolios, especially among same-ethnicity pairs. Kuchler et al. (2022) argue that social proximity brings investors more awareness of firms in socially connected regions. Accordingly, they find that institutional investors are more likely to invest in firms that are located in regions where they are socially connected. They find that social proximity is more influential for less noticeable firms, decreasing in size and analyst coverage.

Social interactions extend to the real estate market, where people make relatively few decisions involving large sums of money. Burnside et al. (2016) develop a model that studies boom and bust periods in the housing market. They argue that market participants have heterogeneous expectations, and social dynamics cause them to change their expectations. Thus, the fraction of people who believe in a particular view about the future of the real estate market fluctuates over time and either creates boom-bust cycles or protracted booms. Bailey et al. (2018) use Facebook data to examine how social interactions affect house-purchasing decisions and find that people's expectations are affected by the experiences of others. The results show that if a person has a friend who benefited from a more considerable house price increase, that person is more likely to buy a house instead of renting. In a similar setting, Bailey et al. (2019) find that people whose friends experienced a minor growth in the real estate market use more leverage than others. Maturana and Nickerson (2019) find workplace peer effects in mortgage refinancing among Texas teachers, especially younger ones. Bayer et al. (2021) show that neighborhood investment activity predicts real estate participation and performance.

Few studies examine the specific mechanisms by which social interactions exert influence. Banerjee et al. (2013) find that participation status affects microfinance information transmission. Rantala (2019) examines the dissemination of investment ideas by focusing on the network of a Ponzi scheme in Finland. He finds that a small number of influential investors are responsible for the enlargement of the scheme, and it would be much harder to reach the same number of investors without these influential nodes. He also suggests that policies to stop the spread should concentrate on the influential nodes because removing random nodes does not damage the network. Charness et al. (2013) examine whether people change their

attitudes towards ambiguity as a result of interacting with other people. Their experiments show that ambiguity-seeking and ambiguity-incoherent people can change their attitudes if they interact with an ambiguity-neutral person who is incentivized to persuade others to become ambiguity-neutral. Ahern et al. (2014) find that peer effects can positively affect risk aversion in MBA students but do not affect trust. People's degree of risk aversion tends to converge to the average of their peers. The field experiment of Bursztyn et al. (2014) shows that social learning and social utility are significant channels enabling social interactions to influence investment decisions. Heimer (2016) finds that social interactions double a trader's tendency to exhibit the disposition effect. He argues that this increase relates to managing impressions to highlight earnings and hide losses as much as possible. Arrondel et al. (2022) find that social connectedness improves people's perception of past returns and allows them to make better predictions for the future.

This study contributes to the literature by directly studying interacting people on social media platforms and the messages transmitted in these interactions. We specifically focus on the dissemination of investor sentiment in finance-related online communities. Thus, unlike prior studies that focus on financial decisions, we focus on belief content. This way, we can better examine whether social interactions influence people's beliefs, which can ultimately influence decisions.

2.3 Hypotheses development

Building on the theoretical and empirical foundation of studies on peer effects, we now develop our hypotheses regarding sentiment dissemination within online investor networks. Several studies show that people use social media to express themselves and gain information. Whiting and Williams (2013) use in-depth

interviews to explore why people use social media and identify several uses and gratifications, including social interaction, information seeking, expression of opinion, and information sharing. Using online questionnaires, Lin and Lu (2011) find that enjoyment, the number of peers, and usefulness are the most important factors determining the use of social media. Similarly, using questionnaires, Morris et al. (2010) find that people use social media to get information, especially to get recommendations and opinions of others. They also find that infrequent social media users tend to ask questions about uncommon and special events. We argue that people have similar motivations when using social media related to the financial markets. More specifically, people can participate in market-related social media platforms, engage in conversations to share their experiences as well as their expectations, and enjoy the social aspect of investing, as Shiller et al. (1984) mention. In addition, they can also learn from others and form expectations about markets. The source of information in learning can be the news shared on these platforms or, consistent with the information cascade framework of Bikhchandani et al. (1992), it can also be the expectations, beliefs, or actions of others. In any case, the social media community will include users who want to share their beliefs and want to know about the beliefs of others. Some users may be more inclined to talk about their beliefs. The literature suggests active social media users have stronger opinions or more extreme political attitudes (Lee et al., 2018; M.-H. Su et al., 2022). Therefore, the people who actively post on social media platforms may be more likely to have strong opinions they want others to share. Expressing and imposing their beliefs on others would motivate these users to interact online. On the other hand, some users will need more confidence in their views and want to learn from the people they perceive as more experienced.

Considering the nature of the motivation for engaging in online interactions, we argue that people's beliefs are affected by the beliefs of the people they interact with. Once a user expresses their sentiment by posting on social media, the signal will reach others in the community and disseminate throughout the system. Other people will receive this signal and choose to either converge with or diverge from this sentiment. Our main argument is that the receivers will be more likely to converge to this sentiment if senders transmit signals of the same type with a higher frequency. Thus, the beliefs of people who actively participate in the community have a higher chance of being disseminated in the network and affecting others. For instance, a user might have a strongly bullish view of the stock market for the next month. This user may want to express this idea and convince others who think differently or find support from like-minded bullish people. The more this user expresses this idea, the more it will reach others, increasing the probability of bears changing their opinion or other bulls strengthening their view. Frequent transmission of the signal will also facilitate the transmission of information, which, in turn, can contribute to the dissemination of investor sentiment. Cookson and Niessner (2020) argue that information is transmitted from sophisticated investors to unsophisticated investors. Consistent with this view, they find that sophisticated investors' sentiment leads to unsophisticated investors' sentiment. DeMarzo et al. (2003) argue that persuasion bias makes people more likely to believe repeated information. They suggest that central people in a social network can affect other people. Consistently, we expect the sentiment of people who share less on these platforms to follow those who share more. Thus, our first hypothesis is as follows:

Hypothesis 1: The lagged sentiment of high-centrality users positively affects the sentiment of low-centrality users.

The transmission frequency of the signal can also depend on the activity of others, in addition to the activity of the original sender. DeMarzo et al. (2003) suggest that persuasion bias contains network effects, as the repeated information does not have to come from a single source. Users can still fail to consider the repetitions if the information comes from multiple sources that are connected within a network. Turning to our example, when the individuals who actively post on social media share similar bullish views, the receivers are more exposed. In addition to the aggregate increase in the transmission frequency, receiving a similar signal from other users can increase the power of the transmitted sentiment. Suppose on the other hand, there is high disagreement among active social media users. In that case, the receivers are exposed to many conflicting views, which may raise doubts, increase uncertainty, and cause a user to be less influenced by the opinions of other users. Thus, we argue that during periods of high disagreement among active social media users, the dissemination of sentiment within the network is weaker. Less active users can still be affected by the signals of active users, as the dominant sentiment among active users can spread out and become the dominant sentiment among less active users. However, the aggregate effect is weaker. Thus, our second hypothesis is as follows:

Hypothesis 2: The positive effect of the lagged sentiment of high-centrality users on the sentiment of low-centrality users is weaker in periods of high disagreement among high-centrality users.

Uncertainty can also affect the influence of high-centrality users on sentiment dissemination. Due to increased ambiguity and fear of incorrect decision-making, market participants are more likely to seek information and others' opinions when uncertainty is high (Bartoszek et al., 2022; Mirc & Parker, 2020). Additionally, high

volatility can increase reliance on heuristic decision-making and make low-centrality users place greater weight on the opinions of high-centrality users. Under such conditions, high-centrality users who actively interact with others in the network can gain disproportionate attention. These differences can make repeated signals more impactful. Consequently, volatility can enhance the reach of sentiment dissemination and strengthen the influence of dominant voices within the network. Thus, our third hypothesis on sentiment dissemination is as follows:

Hypothesis 3: The positive effect of the lagged sentiment of high-centrality users on the sentiment of low-centrality users is more pronounced in periods of high volatility.

Finally, building on the prior literature on the relationship between online sentiment and stock returns, we argue that as sentiment disseminates through interactions within online social networks, it creates a self-reinforcing mechanism. This mechanism amplifies sentiment-driven behaviors, leading to short-term trading volume and price impacts that align with the prevailing sentiment. Accordingly, we hypothesize that lagged network sentiment will positively influence market returns. Such short-term sentiment-driven impact has also been documented in prior studies (Eierle et al., 2022; Gan et al., 2024; Tan & Tas, 2021; Yang et al., 2015).

Hypothesis 4: The lagged network sentiment positively affects market returns.

2.4 Data and methodology

We first introduce the social media platform we focus on in this study and explain the dynamics of user interactions within it. Then, we describe our methodology for constructing social networks and measuring network sentiment. Finally, we describe

the statistical methods that we use to examine the influence of participants in online social networks on each other and their aggregate effects on the markets.

2.4.1 Social media data

We study the interactions of users on Reddit, a U.S.-based discussion platform with worldwide popularity. Reddit is the 7th most popular website globally and the 3rd most popular in the U.S. as of March 2025, with more than 6 billion monthly visits.[1] Along with its popularity, the structure of Reddit offers several advantages for our study design. Reddit operates as a network of communities where users interact with one another in communities called subreddits. For example, social media users interested in the stock market subscribe to subreddits related to the stock market and interact with others on stock-market-related issues. Therefore, we can link users and their interactions to specific financial markets.

Figure 1 illustrates how interactions can occur on Reddit.

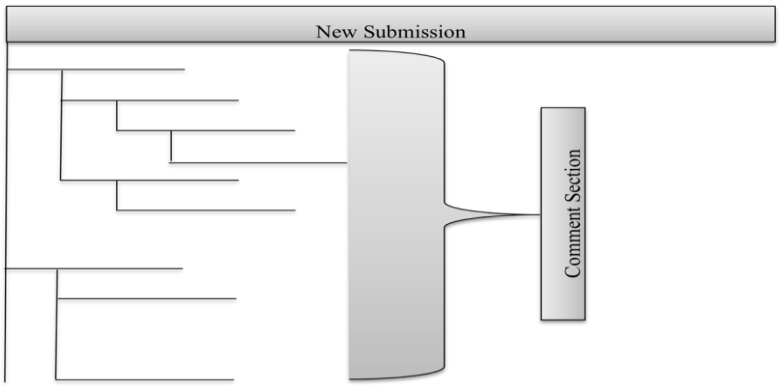

Figure 1. Interaction structure of a post on Reddit
This figure illustrates how people can interact on Reddit. A new submission with its title and content is located at the top. People can comment on this new submission or on other people's comments. Each horizontal line in the comment section represents a separate comment, connected by a vertical line showing the interaction relationship.

[1] https://www.semrush.com/website/reddit.com/overview/

Within subreddits, people can post a new submission and appear on the subreddit's main page or comment on other people's new submissions. Additionally, people can maintain communication and interaction by commenting on what others comment on in new submissions. Thus, we have the opportunity to learn a lot about the nature of interactions on financial markets through the comments of interacting users.

We collect all comments and new submissions posted on the r/stocks subreddit throughout the sample period using the Pushshift Reddit API, which is designed to capture Reddit comments and submissions conveniently.[2] Our dataset contains information on usernames, user sign-up dates, ID numbers, dates, and the complete body of new submissions and comments, along with ID numbers of interacted comments. Table 1 presents the number of new submissions and comments.

Table 1. Number of Comments and Submissions by Year

	Submissions	Comments
2019	10,303	200,746
2020	41,116	1,071,959
2021	29,149	1,361,363
2022	9,941	794,752
Total	90,509	3,428,820

This table provides information on the number of new submissions and comments posted on the r/stocks subreddit. The number of new submissions and comments in 2022 includes only those made until the end of June. The r/stocks subreddit is a community related to the stock market.

[2] https://github.com/pushshift/api

Table 2 provides descriptive statistics of the daily social media data. We see that, on average, users make 2732 comments per day in the r/stocks subreddit, of which 63% are made to another user's comment. Also, users start an average of 72 new submissions per day. During the sample period, subscribers to the r/stocks subreddit grew by an average of 0.19% per day. This growth rate provides a substantial increase during the sampling period. At the start of January 2019, the r/stocks subreddit had approximately 300,000 subscribers and reached more than 4 million subscribers by the end of June 2022.

Table 2. Descriptive Statistics of Social Media Data

	Comments	To comments (%)	Submissions	Subscriber Growth (%)
Mean	2732	62.7	72	0.19
St. Dev	2074	4.6	55	0.66
Q1	691	59.7	34	0.1
Median	2736	62.8	55	0.12
Q3	3853	65.7	90	0.17
Min.	42	46.4	1	0.03
Max.	21688	79.9	519	14.9

This table reports descriptive statistics of the daily social media data. To comments (%) shows the ratio of comments made to another comment in all comments. In other words, it represents the degree of interaction within the subreddit. Q1 and Q3 represent the 25th percentile and the 75th percentile, respectively.

Figure 2 illustrates the frequency of online activity within a day. The activity starts to increase in the morning, accelerates with the start of trading hours, and declines after noon until the following day. We see the least activity at night. Figure 3 shows the online activity for the days of the week. The community exhibits the highest activity on Wednesdays and Thursdays and the least on weekends. Previous studies also show that stock market trading volume is higher in the morning and on

Wednesdays and Thursdays (Foster & Viswanathan, 1993; Jain & Joh, 1988; Kiymaz & Berument, 2003). Overall, the similarity between online and trading activities implies that there may be an association between market activity and social media activity.

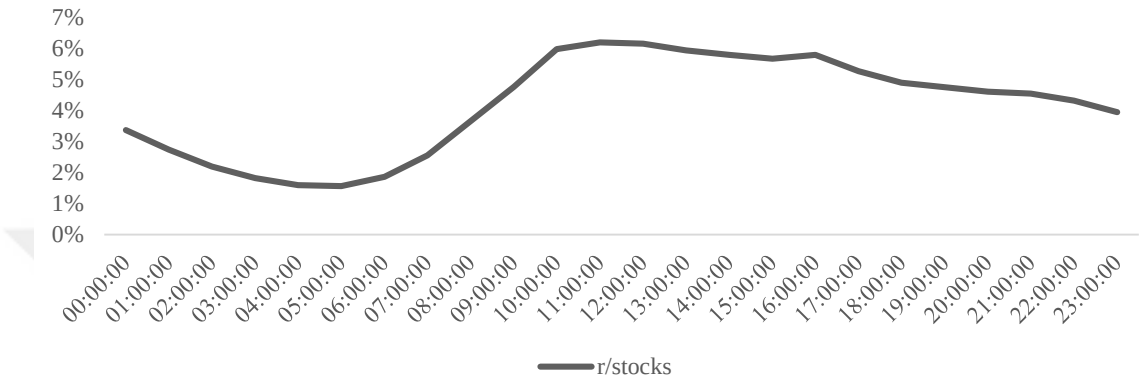

Figure 2. Distribution of comments within a day

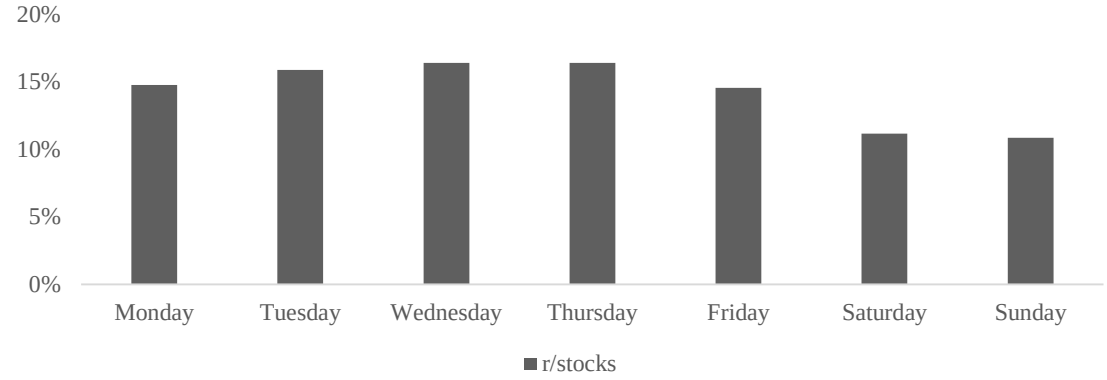

Figure 3. Distribution of comments by weekdays

2.4.2 Social network and network centrality

In order to understand the dynamics within online social networks, we model the networks of users based on their interaction with one another. We define interaction as directly commenting on other users' comments. Therefore, we consider two users to have interacted if one of them comments on the other one's comment. We only

focus on the interactions in the comment section and do not link the users who start a new submission with those who comment on them. Doing so may overweigh the influence of new submissions and the users who created them. However, we also construct interaction networks containing new submissions to examine their nature separately. Users who comment but do not interact with other users' comments constitute the isolated nodes in the network. We include isolated nodes in our study but exclude them in robustness checks to see whether they have any impact. Figure 4 illustrates an example graph based on the activity in the comments section. We can see relatively few nodes at the center and many isolated nodes at the periphery that have not interacted with anyone else.

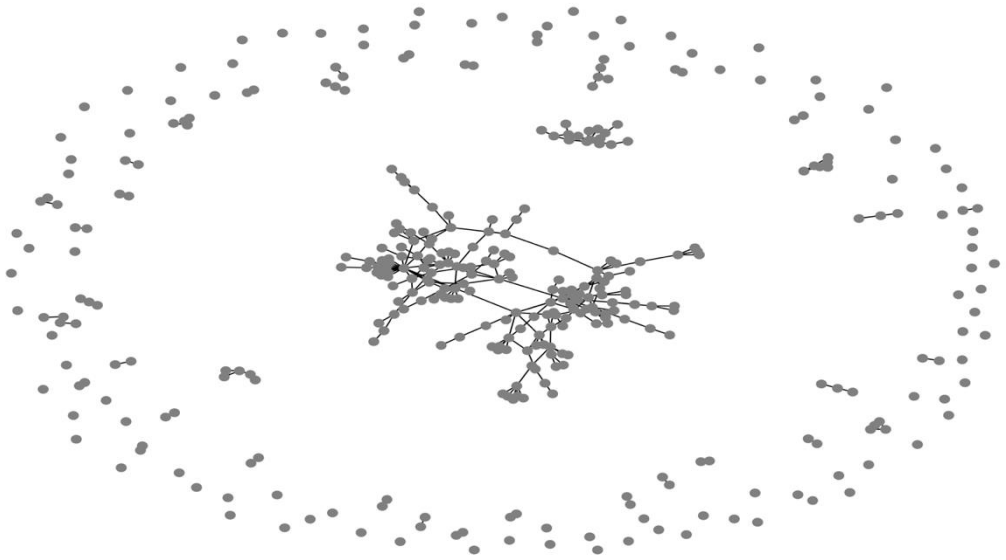

Figure 4. An example of a daily network

Figures 5, 6, and 7 illustrate how the number of nodes, the number of edges, and network densities evolve over time, respectively. The number of nodes reflects the number of unique users who comment on a given day, the number of edges

shows the frequency of comment-to-comment interaction, and network density shows the ratio of the number of actual connections to all possible connections. In Figures 5 and 6, we see a noticeable increase in the number of nodes and edges following the COVID-19 pandemic, with a sharp increase starting in March 2020. Although Figures 5 and 6 show an upward trend in online activity from January 2019 to June 2022, Figure 7 shows that interaction frequency tends to decline throughout our sample period. We see that more users subscribe to the subreddits and make more comments, but comment-to-comment interactions do not increase at this rate of increase. Network density decreased until March 2020, then stabilized after March 2021. The correlation between the number of nodes and network densities is high and negative, with a coefficient of -0.78. Overall, we see that online activity experiences significant changes in our sampling period.

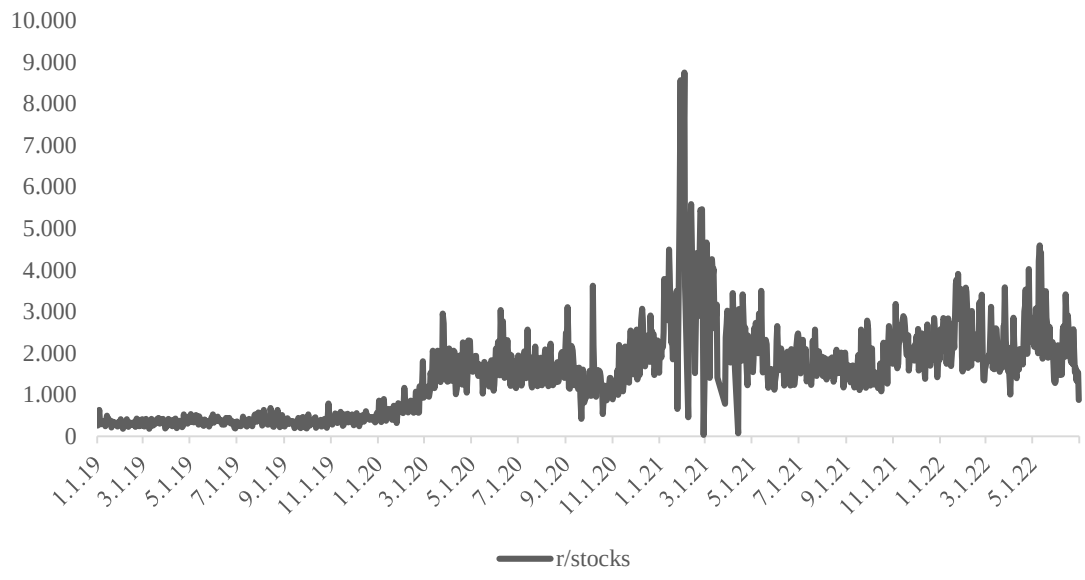

Figure 5. Number of nodes over time

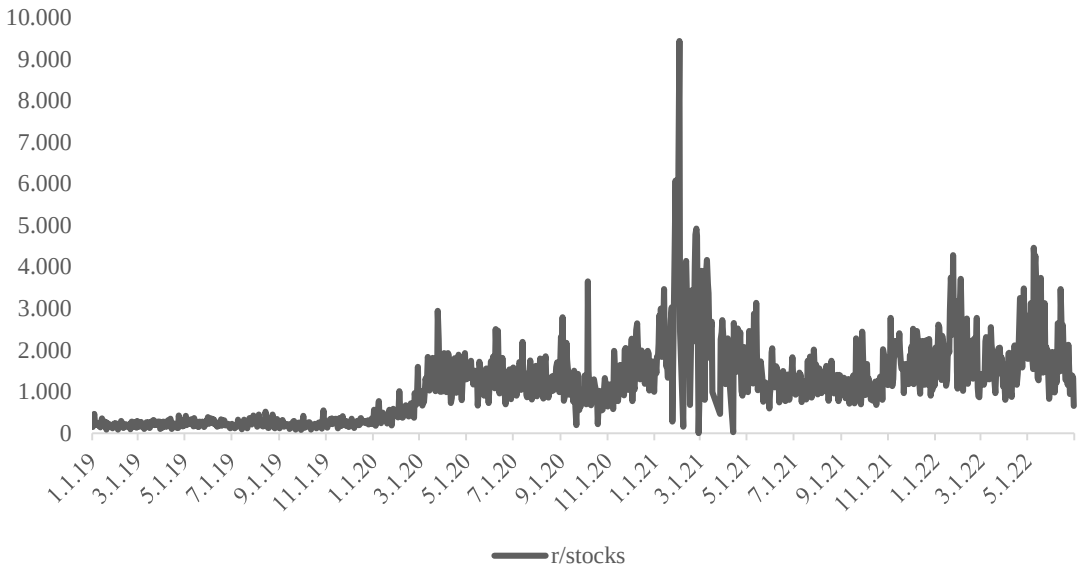

Figure 6. Number of edges over time

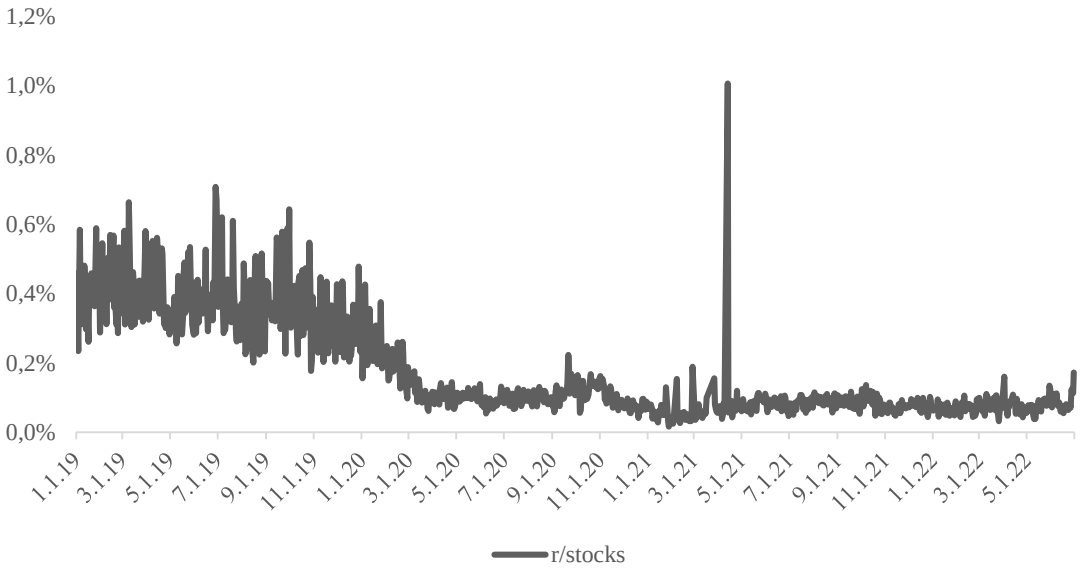

Figure 7. Network density over time

To determine the frequency required to model the networks, we first map the comments that interacted within a year and calculate the time elapsed between them. The average time between interacting comments is 11.10 hours, and 97% are less than 24 hours. Considering the fast-paced nature of social media platforms, we

construct a new network for each day, as the interactions occur mainly within a day. Nonetheless, we also construct weekly and monthly networks to understand whether network construction frequency affects our results.

The primary measure of centrality we use is the eigenvector centrality of Bonacich (1972). It is based on the number and popularity of connections a node has. The links to more popular nodes contribute more to the centrality score than those to less popular nodes. The weight of popular nodes makes eigenvector centrality particularly suitable for our study as it captures not only direct influence but also the broader reach of sentiment dissemination through indirect connections. Nodes may be more influential on other nodes if a connection to an already popular node gives them more visibility. Nonetheless, as a robustness check, we also use the degree centrality measure, which does not consider node popularity and counts only the number of connections a node has. The calculation of eigenvector centrality is as follows:

$$c_i^{EIG} = (1/\lambda)\sum a_{ij} c_j^{EIG} \tag{2.1}$$

where c is the centrality score, a_{ij} are the elements of the adjacency matrix and take a value of 1 if a node i is connected to node j and 0 otherwise, and λ is a constant.

We sort comments into five quintiles according to the centrality scores of the users who post them. We refer to the users in the top centrality group as high-centrality users or active users and the bottom centrality group as low-centrality users. Table 3 reports the average number of comments and users in each quintile. On average, the lowest centrality group has 612 comments per day made by 490 users. The highest centrality group, on the other hand, has 521 comments made by 107 users. Thus, we examine how the sentiment of a small number of users impacts the sentiment of the other users in the network.

Table 3. Average Number of Comments and Users by Centrality Group

	Lowest	Q2	Q3	Q4	Highest
Comments	612	416	450	450	521
Users	490	234	208	151	107

The mean daily count of comments and distinct nodes across centrality groups. The number of comments for the year 2022 is restricted to those until the end of June.

2.4.3 Investor sentiment

We measure the sentiment of each comment in our dataset by creating a binary sentiment variable that equals one if the sentiment in the comment is positive and zero otherwise. We use two rule-based models to evaluate whether a comment is positive or negative. First is the VADER-Sentiment-Analysis tool of Hutto and Gilbert (2014). Hutto and Gilbert (2014) created VADER to measure the sentiment of social media content; thus, its design is highly compatible with our data. The authors first develop and validate a sentiment lexicon in which each word has sentiment intensity scores ranging from -4 to +4. Then, they identify several generalizable heuristics that are used in sentiment classification. Using these heuristics, they build a model that gives an overall sentiment score of the input text ranging from -1 to +1. The second model we use to increase the reliability of the sentiment measurement is Textblob (Loria, 2020). Textblob assigns sentiment scores to words and calculates their weight in a sentence. Like VADER, it gives a sentiment polarity score ranging from -1 to +1. In a relevant study to ours, Cookson and Niessner (2023) use maximum entropy classification, Stocktwits' own sentiment measure, and VADER to construct binary sentiment indices and examine the effect of network disagreement on trading volume. They obtain similar results when they use alternative sentiment measures, showing that the sentiment measurement method has negligible impact. Nonetheless, we average the sentiment polarity scores from

both models and use this average value to form the binary sentiment index to smooth out the differences between the two models and reduce the possible noise that sentiment measurement methods can cause. Additional analyses, which we omit for brevity, show no significant differences when using VADER or Textblob separately. Similar to Cookson and Niessner (2020, 2023), we remove comments that both sentiment models give a neutral classification for.

After we classify each comment by sentiment, we compute a sentiment score for each centrality group by averaging the sentiments of the comments in each group. We calculate the sentiment score of a centrality group as follows:

$$sent_{g,t} = \frac{1}{n}\sum_{c=1}^{n} sent_{c,t,g} \tag{2.2}$$

where $sent_{g,t}$ represents the sentiment of the group g on day t, $sent_{c,t,g}$ represents the sentiment of comments made by the users of group g on day t, n represents the number of comments made by the users of group g on day t. We also average the sentiments of all comments to obtain the sentiment for the entire network. As a robustness check, we weigh each comment's sentiment by its author's centrality score and calculate a weighted sentiment index for each group.

2.4.4 Empirical models

We regress the sentiment of low-centrality nodes on the sentiment of high-centrality nodes to analyze sentiment dissemination. Given that subreddits are theme-based discussion networks, the most common characteristic of users who subscribe to and interact within the r/stocks subreddit is their shared interest in the stock market. That is why we control for past returns, abnormal trading volume, and volatility in our regression framework to account for possible correlated effects. Following Cookson

and Niessner (2020), we measure abnormal trading volume by comparing the trading volume on day t with the average trading volume of the previous 140 days, leaving out the most recent 20 days. To measure volatility, we calculate the range between the highest and lowest intraday prices. Thus, our base model is as follows:

$$sent_low_t = \alpha + \beta_1 sent_high_{t-1} + \beta_2 r_{t-1} + \beta_3 avol_{t-1} + \beta_4 prange_{t-1} + \beta_5 sent_low_{t-1} + \varepsilon_t \quad (2.3)$$

where $sent_low$ and $sent_high$ represent the sentiment of the lowest and highest centrality nodes, r represents returns of the S&P500 index, $avol$ represents abnormal trading volume, $prange$ represents the natural logarithm of the range between highest and lowest intraday prices, and t represents day. To ensure the robustness of our findings, we complement the regression analysis with vector autoregression (VAR) models and Granger causality tests. These additional models also provide information on the directionality of sentiment dissemination.

To further investigate the dynamics of sentiment dissemination and explore potential moderating factors, we extend our analysis by incorporating additional conditions. First, we create a COVID-19 dummy variable and introduce an interaction term in the regression model to assess whether the pandemic period strengthens the effect of high-centrality nodes on low-centrality nodes' sentiment. Second, we test whether sentiment dissemination weakens under conditions of high disagreement by splitting the sample into periods of high and low disagreement among active users and rerunning Equation 2.3 for each subsample. Finally, to examine the role of market uncertainty, we divide the sample into periods of high and low volatility and estimate the regression model separately for each period.

After analyzing the dynamics of sentiment dissemination and identifying the moderating effects of the COVID-19 period, disagreement, and market volatility, we turn our attention to understanding how network sentiment influences financial markets. Specifically, we test the relationship between abnormal network sentiment and excess returns of the S&P 500 index. We calculate the abnormal network sentiment by dividing the network sentiment on day t by the average sentiment of the prior 20 days. We explore the implications of online network interactions by estimating the following model:

$$r_t = \alpha + \beta_1 abn_sent_{t-1} + \beta_2 r_{t-1} + \beta_3 lnnodes_{t-1} + \beta_4 avol_{t-1} + \beta_5 prange_{t-1} + \varepsilon_t \quad (2.4)$$

where abn_sent represents the abnormal network sentiment, r represents the market returns in excess of the risk-free rate, $lnnodes$ represents the natural logarithm of the number of nodes in social networks, $avol$ represents the abnormal trading volume and $prange$ represents the intraday price range. Data on the S&P 500 index are collected from Datastream. We include $lnnodes$ in the model to investigate the effects of network participation on stock returns. The number of social media users can be considered a proxy for investor attention, which has been shown to be related to trading behavior and future stock returns in previous studies (Barber & Odean, 2008; Da et al., 2011). We standardize the abnormal sentiment and number of nodes so that coefficients show the return response to the one standard deviation increase in network dynamics.

We also investigate the effect of the sentiment in before-the-market interactions, occurring between 00:00 and 09:30, on subsequent daily returns. These early interactions allow us to assess whether sentiment formed before the official market opening has predictive power over the rest of the trading day. We hypothesize

that sentiment expressed during this window could reflect user reactions to overnight developments, global markets, or early news releases, which then propagate through the network and influence market outcomes.

Finally, we test the practical relevance of network-based sentiment measures by evaluating the performance of a trading strategy, which is based on the levels of abnormal sentiment. A buy or cover signal is generated when the sentiment of the prior day is abnormally high, while a sell or short signal is triggered otherwise. The strategy is benchmarked against the S&P 500 index to evaluate its relative performance. By comparing returns and drawdowns, we assess the potential of sentiment-based trading strategies to generate superior returns and provide protection during periods of market uncertainty.

2.5 Empirical findings

In this section, we first report our results on the dissemination of sentiment, followed by the implications of network interactions on the market returns. Next, we examine the profitability of the network-based trading strategy and perform several robustness checks.

2.5.1 Dissemination of sentiment

To analyze sentiment dissemination within online social networks, we examine the relationship between the sentiment of high-centrality nodes and the sentiment of low-centrality nodes. Table 4 reports the regression results. The $covid$ dummy in the model equals 0 before March 11, 2020, and 1 after, based on the World Health Organization's (WHO) declaration of COVID-19 as a pandemic. To account for

correlated effects, we control for market variables, including lagged returns, abnormal trading volume, and intraday price range.

Panel A of Table 4 shows that in the overall sample period without the $covid$ interaction, the sentiment of high-centrality nodes has a significant positive effect on the sentiment of low-centrality nodes, with a coefficient of 0.1198. This result supports our hypothesis that users with higher centrality play a key role in disseminating sentiment to less active users in the network. The coefficient on the lagged sentiment of low-centrality nodes is also positive and highly significant, which indicates persistence in sentiment among less-active users. Panel B reveals significant shifts during the pandemic period. Consistent with increased uncertainty during the pandemic period, the coefficient of the $covid$ dummy is negative and significant at the 1% level. After the onset of the pandemic, the sentiment of low-centrality nodes significantly decreases. Additionally, we see that the coefficient of $sent_high_{t-1}$, which measures the effect of high-centrality nodes' sentiment in the pre-pandemic period, becomes statistically insignificant.

Table 4. The Effect of High-Centrality Users on the Sentiment of Low-Centrality Users

	Panel A: Without $covid$	Panel B: With $covid$
$sent_high_{t-1}$	0.1198***	-0.0193
	(0.0402)	(0.0530)
r_{t-1}	0.3393*	0.3594**
	(0.1820)	(0.1769)
$avol_{t-1}$	-0.0289**	-0.0315**
	(0.0132)	(0.0128)
$prange_{t-1}$	-1.3491***	-1.0995**
	(0.4420)	(0.4331)
$sent_low_{t-1}$	0.3379***	0.2892***
	(0.0431)	(0.0441)
$covid$		-0.1092***
		(0.0235)
$sent_high_{t-1} \times covid$		0.3064***
		(0.0631)
Constant	0.2631***	0.3353***
	(0.0249)	(0.0322)
Observations	587	587
R^2	0.4058	0.4353

The relationship between the lagged sentiment of high-centrality nodes ($sent_high_{t-1}$) and the sentiment of low-centrality nodes ($sent_low_t$) with and without the $covid$ dummy. *r*, *avol*, and $prange$ refer to the returns in excess of the risk-free rate, abnormal trading volume, and intraday price range of the S&P 500 index. $sent_low_{t-1}$ is the lagged sentiment of low-centrality nodes. $covid$ is a dummy variable that equals 0 before March 11, 2020, and 1 after. $sent_high_{t-1} \times covid$ is the interaction term of $sent_high_{t-1}$ and $covid$. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

Nevertheless, the coefficient of the interaction term is 0.3064 and statistically significant. This result suggests that the influence of high-centrality nodes on low-centrality nodes' sentiment did not exist before the pandemic but emerged strongly after its onset. This change may be attributed to the influx of many new users, possibly novice traders, joining the network and amplifying the effect of signal

senders on receivers. The increase in online activity and uncertainty during the COVID-19 period likely enhanced the reach and impact of high-centrality users in the network.

Following the role of the pandemic, we examine the potential impact of belief heterogeneity by focusing on periods of abnormally high disagreement among high-centrality users. Following Cookson and Niessner (2020), we measure disagreement as the standard deviation of sentiment scores within a day across comments. Abnormally low disagreement refers to the disagreement below the mean of the prior 20 days. As a robustness check, we also consider the mean of the prior 60 days and 120 days and find similar results, though we omit them for brevity. Table 5 reports our findings. The impact of high-centrality nodes is approximately 11% higher in low-disagreement periods than in high-disagreement periods. These findings are consistent with the persuasion bias argument: when high-centrality users share similar views, the perceived repetition of signals increases, which leads to stronger dissemination.

Lastly, we turn to market volatility as another moderating factor. Periods of high volatility increase uncertainty, which may lead market participants to seek external opinions and rely more on heuristic decision-making. To investigate the role of volatility, we divide the sample into high-volatility and low-volatility periods based on intraday price ranges. More specifically, we compare the intraday range on day t with the average of the prior 20 days. When volatility is higher than the mean of the prior 20 days, we consider it a high-volatility day. Then, we rerun Equation 2.3 for both subsamples.

Table 5. The Effect of Disagreement Levels on the Dissemination of Sentiment

	Low Disagreement	High Disagreement
$sent_high_{t-1}$	0.216***	0.195**
	(0.072)	(0.079)
r_{t-1}	0.471	0.351
	(0.324)	(0.223)
$avol_{t-1}$	-0.047**	-0.016
	(0.022)	(0.016)
$prange_{t-1}$	-1.03	-1.53***
	(0.76)	(0.54)
$sent_low_{t-1}$	0.261***	0.360***
	(0.065)	(0.062)
Constant	0.263***	0.230***
	(0.045)	(0.032)
Observations	273	314
R^2	0.3270	0.4535

The effect of disagreement levels among high-centrality nodes ($abn_disagt_{high,t-1}$) on the relationship between the lagged sentiment of high-centrality nodes ($sent_high_{t-1}$) and the sentiment of low-centrality nodes ($sent_low_{t}$). The dependent variable is the sentiment of low-centrality nodes. *r*, *avol*, and $prange$ refer to the returns in excess of the risk-free rate, abnormal trading volume, and intraday price range of the S&P 500 index. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

Table 6 shows that the sentiment of high-centrality nodes does not significantly affect the sentiment of low-centrality nodes when volatility is not abnormally high. On the other hand, in high-volatility periods, we see a positive association between the two, statistically significant at the 1% level. One unit increase in the sentiment of high-centrality nodes increases the sentiment of low-centrality nodes by 0.20 units in high-volatility days, which is two-thirds more than the effect we report in our base model. Classifying the high-volatility days using the mean of the prior 60 days and 120 days does not affect our results.

Table 6. The Effect of Volatility Levels on the Dissemination of Sentiment

	Low Volatility	High Volatility
$sent_high_{t-1}$	0.076	0.1996***
	(0.053)	(0.0562)
r_{t-1}	0.264	0.508***
	(0.369)	(0.182)
$avol_{t-1}$	-0.033*	-0.009
	(0.021)	(0.015)
$prange_{t-1}$	-2.719***	-1.267***
	(1.017)	(0.508)
$sent_low_{t-1}$	0.280***	0.378***
	(0.057)	(0.069)
Constant	0.284***	0.196***
	(0.032)	(0.027)
Observations	347	240
R^2	0.308	0.5306

The effect of intraday price range on the relationship between the lagged sentiment of high-centrality nodes ($sent_high_{t-1}$) and the sentiment of low-centrality nodes ($sent_low_t$). The dependent variable is the sentiment of low-centrality nodes. *r*, *avol*, and $prange$ refer to the returns in excess of the risk-free rate, abnormal trading volume, and intraday price range of the S&P 500 index. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

Overall, our results emphasize the role of high-centrality users in disseminating sentiment within online social networks. Their influence becomes more pronounced during periods of increased uncertainty and reduced disagreement. Our findings demonstrate that the dynamics of online social networks are sensitive to broader market conditions and structural shocks, such as the COVID-19 pandemic.

2.5.2 Implications on market returns

Next, we investigate the relationship between network and market dynamics, specifically focusing on the impact of abnormal sentiment on subsequent returns of the S&P500 index. We calculate the abnormal network sentiment by dividing the

network sentiment on day t by the average sentiment of the prior 20 days. We also use 60-day- and 120-day periods for robustness and obtain similar results, though we omit them for brevity.

The results in Table 7 reveal that the effect of the abnormal network sentiment of the stock market community is positive and statistically significant at the 5% level. One standard deviation increase in abnormal sentiment increases the daily return by an additional 0.13%. This finding suggests that participants in financial market-related online networks act on their expressed beliefs, which creates a positive price pressure in the short term. Additionally, optimistic sentiment may reflect favorable market expectations or improved economic outlooks, which can explain the higher returns and provide evidence for information-based network interactions. The positive relationship between network sentiment and stock returns is consistent with prior studies that examine different social media platforms (Chen et al., 2014; Tan & Tas, 2021; Yang et al., 2015).

By contrast, the number of nodes does not exhibit a statistically significant relationship with stock market returns. Merton (1987) argues that stocks with low investor recognition are riskier and require higher returns. On the other hand, Barber and Odean (2008) and Da et al. (2011) argue that attention-induced buying and short-selling constraints create positive price pressure on attention-grabbing stocks. Although these two views have different expectations for the relationship between attention and returns, the fundamental aspect in both is the attention constraint that prevents investors from processing all stock-specific information. The high number of individual stocks can make investor attention influential on the returns of individual stocks, but it can be insignificant at the index level, which constitutes the

most popular stocks. This difference may be the reason why we fail to see a significant effect of attention on returns at the index level.

Table 7. Abnormal Sentiment and Returns of the S&P 500 Index

	r_t
abn_sent_{t-1}	0.0013**
	(0.0006)
$lnnodes_{t-1}$	0.0003
	(0.0005)
$avol_{t-1}$	0.0005
	(0.0035)
$prange_{t-1}$	-0.0684
	(0.1386)
r_{t-1}	-0.2625***
	(0.0865)
Constant	0.0016
	(0.0017)
Observations	865
R^2	0.0677

The lagged relationship between abnormal network sentiment (abn_sent) and returns of the S&P500 index in excess of the risk-free rate (r). $lnnodes$ is the natural logarithm of the daily number of nodes in social networks, $avol$, and $prange$ is the abnormal trading volume and intraday price range of the S&P500 index. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

We further investigate the implications of social networks on financial markets by evaluating the timing of interactions. We isolate interactions posted from midnight until the markets open as pre-market interactions. Table 8 shows that the sentiment in pre-market interactions can also predict subsequent daily returns significantly. One standard deviation increase in the pre-market sentiment increases market returns by 0.1%. Our findings illustrate the importance of the timing of the

interactions, as focusing only on the pre-market interactions also reveals significant network influence over market returns.

Table 8. The Effect of Before-Market Sentiment on Returns of the S&P 500 Index

	r_t
$sent_before_t$	0.0010**
	(0.0005)
$lnnodes_{t-1}$	0.0002
	(0.0004)
r_{t-1}	-0.2522***
	(0.0873)
$prange_{t-1}$	-0.0800
	(0.1415)
$avol_{t-1}$	0.0010
	(0.0035)
Constant	0.0006
	(0.0034)
Observations	857
R^2	0.0645

The effect of before-market sentiment on subsequent returns of the S&P500 index. $sent_before$ represents the sentiment of comments made between 00:00 and 09:30 on day t. $lnnodes$ represents the natural logarithm of the daily number of nodes. *r*, *avol*, and $prange$ refer to the returns in excess of the risk-free rate, abnormal trading volume, and intraday price range of the S&P 500 index. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

For further robustness, we explore alternative pre-market windows, including the market closing on day $t-1$ to the opening on day t, as well as sentiment expressed two, three, and four hours before the market opens. In all cases, we find consistent evidence that pre-market sentiment positively predicts subsequent daily returns. For the sake of brevity, we do not report the results for the alternative pre-market windows here; however, they are available upon request.

2.5.3 The evaluation of a network-based trading strategy

Considering our findings on network sentiment's impact on stock returns, we explore the performance of a market timing strategy. The strategy is based solely on the level of abnormal sentiment, and the benchmark is the S&P 500 index.

The strategy generates a buy/cover signal when the sentiment of the prior day is abnormally high, and a sell/short signal is triggered otherwise. Table 9 reports the key performance metrics. We see that the long-only portfolio trades on 386 out of 845 days, while the long/short portfolio maintains continuous exposure per the strategy rule. Long positions are held for a maximum of 9 days, while short positions extend up to 13 days. Taking only long positions has a higher hit rate than the benchmark, with 58% compared to 56%. Although the strategy slightly reduces the positive returns on average, it ensures that the portfolio suffers significantly less loss. The average negative return improves to 0.75% from -0.97%. This downside protection leads to a higher total return and a smaller drawdown over the sample period. The long-only strategy gains 75% with a maximum drawdown of 10% versus the 42% return and 34% maximum drawdown of the benchmark. The strategy helps protect portfolios from the collapse caused by the pandemic. The long/short portfolio has a lower hit rate of 51% but still outperforms the benchmark, with total returns more than doubling. However, taking short positions reduces downside protection with a maximum drawdown of 17.6%.

To account for trading costs, we assume a 0.1% commission per trade. After including commissions, the total return for the long-only strategy falls to 27% and for the long/short strategy to 11%. Evidently, constantly switching between long and short positions destroys profitability through commissions, causing the strategy to underperform the benchmark. Nevertheless, the long-only strategy remains attractive

by maintaining a 15% maximum drawdown. Although commissions destroy most of the profitability, which is common in daily strategies, a strategy based on abnormal sentiment can still provide essential and reliable protection signals and offer reliable downside protection. Importantly, these strategies can be more practical and profitable with the rise of commission-free trading platforms.

Table 9. Trading Strategy Performance

	Benchmark	Long Only	Long/Short
Days	845	845	845
Open Days	-	386	845
Max. Long Days	-	9	9
Max. Short Days	-	-	13
Max. No-Trade	-	13	-
Hit Rate	%56	58%	51%
Average Gain	0.87%	0.82%	1.00%
Average Loss	-0.97%	-0.75%	-0.83%
Average Return	0.05%	0.07%	0.10%
Std. Deviation	1.45%	0.84%	1.48%
Profit/Loss Ratio	0.89	1.09	1.19
Total Return	42%	75%	116%
Max. Drawdown	-33.9%	-10.4%	-17.6%

Performance evaluation of a trading strategy based on levels of abnormal network sentiment. Max. Long Days and Max. Short Days show the maximum number of days a long or short position is kept open. Max. No-Trade shows the maximum number of consecutive days without an open position. Benchmark is buying and holding the S&P500 index.

2.5.4 Robustness tests

We perform several robustness checks based on alternative specifications in daily networks. First, we reconstruct networks by excluding isolated nodes and including new submissions. Then, we use degree centrality and sentiment indices weighted by the centrality scores of users as alternative measures. Finally, we employ a vector

autoregression (VAR) model to test the robustness of sentiment dissemination between high-centrality and low-centrality nodes.

We report our findings in Table 10, which show that the dissemination of sentiment is robust to using alternative specifications. We obtain similar results when we exclude isolated nodes or include new submissions within the social networks. The two differences we identify are that excluding isolated nodes removes the moderating effect of high volatility and, including new submissions alters the role of disagreement. When the sentiment is more dispersed among new submissions, the impact of high-centrality nodes on low-centrality nodes is stronger. However, the role of uncertainty stays the same when new submissions are included. Finally, we show that alternative measurement methods exhibit minimal significance for sentiment dissemination. Using degree centrality confirms all our observations. High-centrality nodes can positively affect the sentiment of low-centrality nodes, and their impact is more pronounced during the pandemic and when disagreement is low. Similarly, when volatility is not abnormally high, we observe no significant association. Nonetheless, weighting the sentiment index only confirms the role of the pandemic in increasing the influence of high-centrality nodes.

Table 10. Alternative Estimations for the Effect of High-Centrality Users on Low-Centrality Users

	Entire Sample	During Pandemic	Low Disagt.	High Disagt.	Low Volatility	High Volatility
Baseline model	0.1198***	0.3064***	0.216***	0.195**	0.033	0.1996***
Without isolated nodes	0.0647**	0.2699***	0.185***	0.052	0.064	0.054
Networks with new submissions	0.1346***	0.3008***	0.1466***	0.1886***	0.105***	0.203***
Degree centrality measure	0.1327**	0.1610*	0.2690***	0.2066*	0.112	0.135***
Weighted sentiment index	0.0228	0.4583***	0.1740	0.1472	0.017	0.063

The effect of the one-day lagged sentiment of the high-centrality nodes on the sentiment of low-centrality nodes, estimated from alternative specifications. The entire sample period is from the beginning of January 2019 to the end of June 2022. "During Pandemic" represents the period after March 11, 2020. "Low Disagt." and "High Disagt." represent abnormally low and high disagreement periods, whereas "Low Volatility" and "High Volatility" represent abnormally low and high volatility periods. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

Table 11 reports the VAR results. The lagged sentiment of high-centrality nodes significantly predicts the sentiment of low-centrality nodes with a coefficient of 0.1232. Similarly, the sentiment of low-centrality nodes also affects the sentiment of high-centrality nodes, albeit with a delay, as indicated by the significant coefficient on the third lag. The Granger causality tests confirm a bidirectional relationship between the two groups. These findings support the robustness of our main results and demonstrate that sentiment dissemination is not unidirectional but involves mutual feedback between central and less-central users in the network.

Table 11. The Interaction between the Sentiment of High-Centrality Users and Low-Centrality Users

	$sent_low_t$	$sent_high_t$
$sent_high_{t-1}$	0.1232***	0.1796***
$sent_high_{t-2}$	0.0167	0.0836
$sent_high_{t-3}$	0.0610	0.0176
$sent_low_{t-1}$	0.2960***	0.1340*
$sent_low_{t-2}$	0.0824	0.0292
$sent_low_{t-3}$	0.2251***	0.1691**
Constant	0.1037***	0.1285***
Observations	327	327
Granger	12.58***	13.08***

The estimation results of the VAR models between the sentiment of high-centrality nodes ($sent_high$) and the sentiment of low-centrality nodes ($sent_low$). Granger is the χ^2 values from the Granger causality tests. t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

Lastly, in Table 12, we see that alternative network construction methods provide similar findings on market implications. The abnormal sentiment of networks without isolated nodes or networks with new submissions can still positively and significantly affect market returns.

Table 12. Alternative Estimations for the Effect of Abnormal Network Sentiment on the Returns of the S&P 500 Index

	$abn_sent_{t-1} - r_t$
Baseline model	0.0013**
Without isolated nodes	0.0014**
Networks with new submissions	0.0013**
Weighted sentiment index	0.0007

The effect of abnormal sentiment on the returns of the S&P500 index, estimated from alternative specifications. r, *avol*, and $prange$ refer to the returns in excess of the risk-free rate, abnormal trading volume, and intraday price range of the S&P 500 index. abn_sent represents abnormal network sentiment, and t represents day. ***, ** and * indicates that the coefficient is significant at 1%, 5% and 10%, respectively.

2.6 Conclusion and discussion

Social interactions facilitate the dissemination of information, beliefs, and feelings, which may affect individual decision-making and have broader implications for the markets. In this study, we examine the relationship between online social interactions and financial markets, specifically focusing on the sentiment dissemination within a stock market community on Reddit. Our findings demonstrate that highly active users can spread their sentiments to a broader audience. This influence becomes more pronounced under two conditions: (1) when there is reduced disagreement among high-centrality nodes and (2) during periods of high market volatility. Additionally, we find that the COVID-19 pandemic represents a structural shift that enhances the influence of high-centrality nodes as increased online activity and uncertainty reshaped network dynamics. While we emphasize the role of high-centrality users in leading sentiment dissemination, feedback effects from low-centrality users to high-centrality users cannot be overlooked. Even if individual interactions are limited, a large number of users can send repetitive signals when they have low disagreement among themselves. Thus, even less active users can influence dominant voices, reinforcing online networks' complexity. Furthermore, we show that abnormal network sentiment can predict the next day's market returns. Our findings suggest that participants in financial market-related networks are likely to act on their expressed beliefs and generate short-term price pressures consistent with optimistic or pessimistic sentiment.

The practical implications of our findings are twofold. For market participants, sentiment-based trading strategies can provide increased profitability, especially in commission-free trading environments. In addition, network sentiment can be an effective tool for market timing and creating downside protection. From a

policy standpoint, while online networks can enhance information dissemination, the ability of a few highly active users to stimulate the beliefs of others can be exploited or, to a certain extent, inflate the prices of specific assets in the market; one example being the GameStop short squeeze case.

Nevertheless, our analysis is subject to certain limitations. First, we focus on a single Reddit community, which may provide only a partial view of online social dynamics. Future research could extend this analysis to other social media platforms or other subreddits to cover a broader range of assets. Second, our study does not consider user-specific characteristics when examining network interactions. Notably, influential figures or celebrities can influence others irrespective of their level of engagement on social media platforms. Finally, examining the impact of cultural, regional, or specific information events on sentiment dissemination could enrich our comprehension of the impact of social media on financial decision-making and market dynamics.

CHAPTER 3

SOCIAL MEDIA DISAGREEMENT AND FINANCIAL MARKETS: A COMPARISON OF STOCKS AND BITCOIN

In this study, we examine whether disagreement in social media discussions related to financial markets affects subsequent volatility and abnormal trading volume. We also compare how traditional and digital asset markets differ by comparing stocks and Bitcoin. We show that social media disagreement is positively associated with future market volatility and abnormal trading volume in the stock market. The effect of disagreement is more pronounced at the individual stock level than at the index level. A higher level of social media disagreement also increases the probability of extremely negative stock market returns. In contrast, disagreement in Bitcoin-related social media weakly affects subsequent volatility but does not affect trading volume or extremely negative returns. Our findings also reveal that market activity impacts the disagreement in the stock market and Bitcoin communities differently.

3.1 Introduction

When investors have different beliefs about future market conditions, they can react differently to new information, market events, and changes in economic conditions. Varied expectations, in turn, can result in elevated levels of trading volumes and market volatility. Disagreement on market prospects can also lead to divergent trading strategies and market dynamics. Accordingly, the literature provides evidence of heterogeneous beliefs' volatility and volume-increasing effect in financial markets (Atmaz & Basak, 2018; Banerjee & Kremer, 2010; Carlin et al., 2014;). Most studies use analyst forecasts to focus on disagreement among sophisticated investors. Yet

analyst forecasts fail to provide information on the beliefs of retail investors, especially those interacting on social media platforms. Building upon the growing body of literature using social media content, we investigate how the disagreement among users of social media related to financial markets affects future market activity.

The rise of social media has provided retail investors with platforms like Twitter, Stocktwits, and Reddit to share their views and discuss their expectations on market movements in real time. The literature increasingly focuses on how social media discussions impact financial markets (Chen et al., 2014; Cookson & Niessner, 2020; Tan & Tas, 2021). However, much of the existing research focuses on the direct impact of social media sentiment and paid less attention to the role of disagreement, which refers to the extent to which investors hold opposing views on market direction. Understanding how heterogeneous beliefs impact market activity is important because they can lead to greater trading volume and volatility as investors act on their differing views. This study explores this possibility by analysing comments on Reddit posted between January 2019 and June 2022. Reddit's community-based structure allows us to link discussions to specific assets and facilitate the examination of disagreement's impact on market activity.

We focus on the opinions shared in two important communities about the stock market and Bitcoin. We hypothesise that increased disagreement in these communities will increase trading volume and volatility, as diverse opinions trigger more buying and selling activity. We also expect that disagreement in the stock market community will have a more substantial effect at the individual stock level compared to the index level, where opinions across various stocks can offset each other and reduce the disagreement's overall impact.

We quantify disagreement using the standard deviation of a binary sentiment index across users within a given day. We first average sentiment scores by means of two popular sentiment tools, VADER and TextBlob, and then convert them into a categorical index to capture positive and negative sentiments. The standard deviation of this index shows the belief dispersion. To measure daily volatility, we use the range between the daily highest and the lowest prices. Range-based volatility measures provide stationary and daily volatility measures and have been frequently used in the finance literature (e.g, Blau & Whitby, 2017).

Our findings reveal key differences between the stock market and Bitcoin communities. Disagreement in the stock market community significantly increases the next day's volatility and abnormal trading volume. On the contrary, disagreement in the Bitcoin community does not statistically significantly affect Bitcoin's volatility or abnormal trading volume. However, the influence of disagreement becomes statistically significant for Bitcoin during periods of market drawdowns. We suggest that the Bitcoin community's focus shifts toward immediate market actions when the market is in unfavourable conditions, which amplifies the effect of social media disagreement in such periods. Furthermore, when only posts from midnight until the markets open are considered, disagreement in both communities significantly increases subsequent trading volume and volatility. Since Bitcoin has no official trading hours, its community receives continuous market feedback, allowing before-the-market disagreement to significantly impact subsequent market activity. Considering that we observe statistically insignificant effects in interday models, the Bitcoin community can have a short-term impact on the markets, with possible intraday prediction ability. We also show that investor attention in both communities can affect future volatility. Our results are consistent with the view that online

financial communities facilitate information dissemination, which can induce additional volatility in the near term. We also show that extremely negative returns in the stock market are more likely following increases in social media disagreement. Lastly, focusing on the comments mentioning the tickers of stocks listed in the S&P500 index reveals that the network disagreement effect on volatility and trading volume is more pronounced at the stock level than at the index level.

This study contributes to the literature in several ways. Firstly, we extend previous research on social media and financial markets by shifting the focus from sentiment to disagreement. Secondly, we provide a comparative analysis of the stock market and Bitcoin. By examining two distinct communities on Reddit, we demonstrate the importance of considering asset-specific dynamics in understanding the influence of social media on financial markets. Finally, we examine the discussions on Reddit, which received less attention than Twitter in the literature. Reddit's popularity allows us to reach the beliefs of a wide range of users about the market.

The remainder of this chapter is structured as follows: In Section 3.2, we review related literature. Then, in Section 3.3, we develop hypotheses and explain our methodology. In Section 3.4, we report our empirical findings. In Section 3.5, we give concluding remarks.

3.2 Literature review

Different opinions, sentiments and strategies across investors concerning asset prices can cause market movements. Hong and Stein (2007) argue that disagreement about the future can emerge from two sources. The first is the different information sets investors might have. Due to gradual information flow or investors' limited attention,

some investors can form expectations based on information others do not yet have, resulting in disagreement. The second is the differences in prior beliefs and economic models that allow investors to interpret the same information differently. Different models enable different investors to obtain and pay attention to the same information but reach different conclusions. Both mechanisms can increase disagreement about the future and have significant market implications. Hong and Stein (2007) suggest that for disagreement to create trading volume, investors should agree to disagree, meaning that they should not consider that they may be at an informational disadvantage. Indeed, studies provide evidence of how investors do not constantly update their beliefs based on others' transactions, making disagreement influential in financial markets.

Studies use different measures of disagreement but demonstrate that it significantly affects trading volume and volatility as heterogeneous opinions lead to more active buying and selling behaviours. In early studies, the measure of investor disagreement tends to rely on the dispersion in analyst forecasts and investment newsletter recommendations. These studies frequently explored the relationship between disagreement and trading volume at the stock level. Ajinkya et al. (1991) measure disagreement with dispersion in analyst forecasts and find that it is positively related to trading volume. Similarly, Atiase and Bamber (1994) find that the reaction of trading volume to earnings announcements increases with analyst forecast dispersion. Barron (1995) uses belief revisions and changes in forecast dispersion to measure disagreement. He finds that both disagreement measures are positively related to trading volume. Using analyst forecasts, Bamber et al. (1997) find that disagreement has three aspects that significantly affect trading volume. Their findings show that dispersion in prior beliefs, changes in dispersion, and

different belief revisions across investors positively affect trading volume. Li and Li (2021) examine household belief dispersion in the US and show that disagreement increases trading volume at the market level as well. These findings emphasise the role of disagreement in driving market activity, particularly trading volume.

In addition to trading volume, disagreement is also linked to volatility and volatility-related phenomena. Graham and Harvey (1996) use dispersion in investment newsletters to capture disagreement and find that higher levels of disagreement predict increased future volatility at the market level. Banerjee (2011) shows that disagreement measured by analyst dispersion is positively related to stock volatility. Wang et al. (2023) use the dispersion in analysts' long-term growth forecasts as a measure of disagreement and find that it is significantly related to the idiosyncratic volatility puzzle. These studies demonstrate that market-wide disagreement can have a destabilising effect and increase price fluctuations.

With the advent and widespread use of social media, there has been a noticeable shift in the way studies measure investor disagreement. The practice of people sharing their expectations about financial markets on social media has led to the quantification of investor sentiment in these posts. The extent of dispersion in the opinions of many users has also become observable. The dispersion of sentiment refers to the variability in sentiment scores among different users and serves as a measure of disagreement. A significant part of the literature investigates the market implications of disagreement among online social network interactions. Although the method of measuring disagreement has changed in recent years, its impact on trading volume has consistently been found to be positive. Antweiler and Frank (2004) examine messages posted on Yahoo! Finance and Raging Bull and find that social media disagreement is positively associated with stock-level trading volume.

Similarly, Sprenger et al. (2014) use the dispersion in the sentiment of tweets about the stock market and find that disagreement increases trading volume at the stock level. Using tweets from StockTwits, Al-Nasseri and Menla Ali (2018) also show that disagreement increases stock-level trading volume. Although they follow different approaches to measure social media disagreement, both Giannini et al. (2019) and Cookson and Niessner (2020) also find that disagreement leads to higher stock trading volume.

Some studies investigate the impact of social media disagreement on volatility, in addition to the trading volume. Siganos et al. (2017) use the distance between positive and negative sentiment on Facebook's status updates to measure disagreement and find that it is positively associated with both trading volume and volatility at the market level. Similarly, Li et al. (2018) also find that higher disagreement leads to higher trading volume and volatility by examining Tweets related to the S&P 100 stocks.

Though disagreement has been widely explored in stock markets, its impact on Bitcoin has seen far less attention. In one of the few studies conducted in this area, Ahn and Kim (2020) find that social media disagreement increases Bitcoin's price volatility. Kantorovitch and Heineken (2021) document that high disagreement in online discussions is positively associated with the volatility of Bitcoin. In a related study, Kim and Ahn (2023) find that heterogeneous emotional feedback increases Bitcoin's intraday volatility.

Previous findings provide significant support for the claim that disagreement inflates trading volume and volatility. It appears that when people's beliefs differ, there are more trades in the market, resulting in increased volatility and trading volume. However, the majority of the findings concern the impact of stock-specific

disagreement. The impact of market-wide disagreement or disagreement related to other assets receives little attention in the literature.

3.3 Hypotheses, data, and methodology

In this study, we extend earlier studies by examining two different markets. We examine the effects of social media disagreement on the stock market and Bitcoin. For the stock market, we examine the effects of disagreement both at the index level and the individual stock level. We measure and examine how disagreement in Reddit communities about the stock market and Bitcoin affects abnormal trading volume and volatility.

3.3.1 Hypotheses

We do not expect differences in how disagreement affects trading volume and volatility; rather, we expect the disagreement to have a stronger effect at the individual stock level than at the index level because, at the index level, some effects can neutralise each other. Thus, our hypotheses are as follows:

Hypothesis 5: Social media disagreement positively affects the next day's trading volume and volatility at both the index level and stock level in the stock market.

Hypothesis 6: Social media disagreement positively affects the next day's trading volume and volatility of Bitcoin.

Hypothesis 7: The positive effect of social media disagreement on the next day's trading volume and volatility is more pronounced at the stock level than at the index level in the stock market.

We investigate the effects of social media disagreement on both trading volume and volatility to provide a more comprehensive analysis. Although some studies have examined these effects separately, the literature shows that disagreement often impacts both trading volume and volatility.

3.3.2 Social media disagreement

We examine Reddit, a popular discussion platform, to explore the effects of social media disagreement on financial markets. Reddit is a platform allowing people to interact in theme-based online communities called subreddits. We focus on interactions on two subreddits, r/stocks and r/Bitcoin. These subreddits are selected based on their high user activity and relevance to our research topic. We focus exclusively on Bitcoin and do not include other crypto assets due to Bitcoin's dominance in the cryptocurrency market. Bitcoin has significantly higher liquidity, larger trading volumes, and a more active online community than most other cryptocurrencies. In addition, Bitcoin serves as a benchmark for the cryptocurrency market due to its being the first and most widely recognised cryptocurrency, and its dynamics are often representative of the general behavior of other cryptocurrencies.

We use comments posted in these subreddits using Pushshift Reddit API[3]. Our sample includes approximately 3.5 million comments in the r/stocks and 5 million in the r/Bitcoin posted between January 2019 and June 2022. Both subreddits become more popular in time, with subscribers growing by an average of 0.19% daily for the r/stocks subreddit and 0.11% for the Bitcoin subreddit. The main reason for this difference is the number of subscribers at the starting point. At the start of January 2019, the r/stocks subreddit had approximately 300,000 subscribers, while

[3] https://github.com/pushshift/api

the r/Bitcoin subreddit had 1,000,000. Online activity for Bitcoin was high from the beginning, while it gained significant importance for stocks during the sampling period.

To calculate social media disagreement, we follow Cookson and Niessner (2020) and use the standard deviation of sentiment scores. Though the literature proposes several proxies to measure the disagreement among investors, such as dispersion in analyst forecasts and forecast revisions, Cookson and Niessner (2020) argue that their measure of disagreement is more practical because it directly captures the distribution of sentiment, does not rely on third-party opinions such as analyst expectations, and is available at a daily frequency. Since our study shares the same features and can benefit from these advantages, we follow this approach. We first determine each post's sentiment in the sample and calculate the standard deviation of sentiment scores across comments within a day.

To calculate the sentiment of comments, we use sentiment analysis tools of VADER and Textblob (Hutto & Gilbert, 2014; Loria, 2020). Both tools are rule-based models and provide a sentiment score ranging from -1 to +1. To reduce the noise that may arise from the differences between the two models and sentiment scores ranging from -1 to +1, we first average the sentiment scores given by both models. Then, we create a sentiment dummy variable that equals 1 if the average is greater than zero and -1 if it is smaller. If both models describe a comment as neutral, we remove that comment from the sample. We remove them because they do not reflect strong opinions that could drive market behaviour. Including them may weaken the impact of disagreement. Nevertheless, we repeat our analysis with neutral comments as a robustness check.

Finally, we measure the daily disagreement based on the standard deviation of the binary sentiment index across users within a day. We first determine the sentiment of each post as either -1 or 1, then aggregate these to the daily level by calculating $sent_t$, the average sentiment of comments posted on day t. Since the sentiment of each comment is either 1 or -1, our disagreement measure is:

$$disagt_t = \sqrt{1 - sent_t^2} \tag{3.1}$$

where $disagt_t$ represents the disagreement on day t and $sent_t$ is the average sentiment on day t.

3.3.3 Data and variables

To investigate how social media disagreement affects the stock market, we first examine the aggregate impact by focusing on the S&P 500 index. Then, we investigate the relationships at the individual stock level. To do this, we filter comments mentioning the tickers of stocks included in the S&P500 at any time during the sample period, excluding those with tickers with generic meanings, such as "T" or "BALL". We only consider comments that mention the ticker of only one stock. If tickers of multiple stocks are mentioned in the same comment, we remove that comment from our sample. To ensure this exclusion does not affect our results, we conduct a robustness check, where we retain comments mentioning multiple stocks and assign the sentiment score to all referenced stocks.

After this elimination, our sample consists of 478 individual stocks that were part of the S&P 500 index during the sample period. After our initial filtering, we begin with 284,575 comments that mention tickers of these stocks. Eliminating comments with neutral sentiment reduces our sample to 233,652 comments. Additionally, we remove 74,368 comments that mention multiple stock tickers to

avoid ambiguity in sentiment assignment. These procedures leave us a total of 159,284 comments for our analysis. On average, 127 comments per day mention the ticker of a single stock, with the number of comments ranging from 1 to 513. Per stock, we observe an average of 4 comments per day, ranging from a minimum of 1 to a maximum of 202. The high range between the minimum and maximum number of observations per stock per day reflects the variability in public interest and online discussions for different stocks. To further ensure reliability, we conduct an additional robustness check by removing stocks with only one comment daily. To study how the Bitcoin community affects Bitcoin's trading volume and volatility, we obtain data from CoinMarketCap, a leading data provider on cryptocurrencies.

The calculations of the market variables are as follows:

$$avol_t = \frac{tvol_t}{\sum_{d=t-140}^{t-21} \frac{tvol_d}{120}} - 1 \tag{3.2}$$

$$prange_t = ln\left(\frac{high_t}{low_t}\right) \tag{3.3}$$

where $avol$ represents abnormal trading volume and $prange$ represents the range between the intraday highest ($high$) and lowest (low) prices within day t. Following Cookson and Niessner (2020), we measure abnormal trading volume by comparing the trading volume ($tvol$) on day t with the average trading volume of the previous 140 days, leaving out the most recent 20 days.

3.3.4 Empirical models

To test our hypotheses, we run the Vector Autoregression (VAR) models. In the first set of equations, we examine the impact of social media disagreement on abnormal trading volume:

$$avol_t = \alpha + \beta_1 disagt_{t-1} + \beta_2 avol_{t-1} + \beta_3 lnusers_{t-1} + \beta_4\, r_{t-1} + \beta_5 prange_{t-1} + \varepsilon_t \quad (3.4)$$

$$disagt_t = \alpha + \beta_1 disagt_{t-1} + \beta_2 avol_{t-1} + \beta_3 lnusers_{t-1} + \beta_4\, r_{t-1} + \beta_5 prange_{t-1} + \varepsilon_t \quad (3.5)$$

In the second set of equations, we shift focus to the relationship between social media disagreement and intraday price range:

$$prange_t = \alpha + \beta_1 disagt_{t-1} + \beta_2 prange_{t-1} + \beta_3 lnusers_{t-1} + \beta_4\, r_{t-1} + \beta_5 avol_{t-1} + \varepsilon_t \quad (3.6)$$

$$disagt_t = \alpha + \beta_1 disagt_{t-1} + \beta_2 prange_{t-1} + \beta_3 lnusers_{t-1} + \beta_4\, r_{t-1} + \beta_5 avol_{t-1} + \varepsilon_t \quad (3.7)$$

where $disagt_t$ represents the disagreement in the entire subreddit on day t, $avol$ represents abnormal trading volume, $prange$ represents intraday price range, $lnusers$ represents the natural logarithm of the number of users who comment in subreddits, r represents market returns in excess of the risk-free rate for the stock market and raw returns for Bitcoin, and t represents day. We control for several variables to adjust for the potential effects of investor attention, market activity, and volatility. We include lagged values of abnormal trading volume, price range, user activity, and market returns. We control for the number of users who comment to examine the possible effect of investor attention. Because we use the same controls in both VAR models, Equation 3.5 and Equation 3.7 are identical but belong to different systems of equations. In line with the usual presentation of VAR models, we include $disagt_t$ as the dependent variable in both models, even though the equation structure is the same. We standardize the social media variables so that coefficients show the effect of one standard deviation increase (decrease) in

disagreement. The VAR models also provide evidence of how market activity influences future social media disagreement.

To examine the stock-level relationships, we estimate fixed-effects regressions with heteroskedasticity consistent standard errors. We also control for some stock characteristics that can affect trading volume and volatility, such as market value, book-to-market ratio, and relative spread between bid and ask prices. We estimate the following models:

$$\begin{aligned} avol_{i,t} = \alpha &+ \beta_1 disagt_{i,t-1} + \beta_2 r_{i,t-1} + \beta_3 prange_{i,t-1} \\ &+ \beta_4 avol_{i,t-1} + \beta_5 lnmv_{i,t-1} + \beta_6 lnbm_{i,t-1} \\ &+ \beta_7 spread_{i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{3.8}$$

$$\begin{aligned} prange_{i,t} = \alpha &+ \beta_1 disagt_{i,t-1} + \beta_2 r_{i,t-1} + \beta_3 prange_{i,t-1} \\ &+ \beta_4 avol_{i,t-1} + \beta_5 lnmv_{i,t-1} + \beta_6 lnbm_{i,t-1} \\ &+ \beta_7 spread_{i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{3.9}$$

where $disagt$ represents the level of social media disagreement; $avol$ and $prange$ represent abnormal trading volume and intraday price range; $lnmv$, $lnbm$, $spread$, and r represent the natural logarithm of market value, the natural logarithm of book-to-market ratio, relative bid-ask spreads, and stock returns in excess of the risk-free rate; i and t represent stock and day.

We also investigate whether social media disagreement can predict extreme market movements. To do this, we separately examine extremely negative and extremely positive return days. We define an extremely positive (negative) day as one where the daily return is above (below) the two standard deviations of the prior 120 days. Based on these definitions, we define two categorical variables: ext_pos and ext_neg that take the value of 1 for days with respectively extremely positive and negative return, and 0 otherwise. We test the impact of social media

disagreement with logistic regressions in which the dependent variable is either ext_pos or ext_neg. The logistic regression model is expressed as follows:

$$logit\big(P(ext_event_t = 1)\big) = \\ = \alpha + \beta_1 disagt_{t-1} + \beta_2 avol_{t-1} + \beta_3 prange_{t-1} \\ + \beta_4 ext_event_{t-1} + \varepsilon_t \tag{3.10}$$

where ext_event represents the extreme event and is either ext_pos or ext_neg. $disagt$ is the social media disagreement, avol is abnormal trading volume, prange is the range between high and low prices, and t indicates day. We include the one-day lag of the extreme event to capture potential persistence or autocorrelation in extreme market movements.

3.4 Empirical findings

Before reporting our findings, in Table 13 we first present summary statistics for the key variables used in this study.

Table 13. Descriptive Statistics

r/stocks and S&P500	Obs.	Mean	Std. dev.	Min.	Max.
Sentiment	866	0.3586	0.0752	0.1324	0.6279
Disagreement	866	0.9300	0.0288	0.7783	0.9912
Number of Comments	866	2469	1855	43	19285
Number of Commenting Users	866	1585	1083	71	8741
Return	866	0.0005	0.0144	-0.1198	0.0938
Abnormal Trading Volume	866	0.0233	0.3339	-0.6820	1.8289
Price Range	866	0.0128	0.0102	0.0017	0.0842
r/Bitcoin and Bitcoin	Obs.	Mean	Std. dev.	Min	Max
Sentiment	1255	0.3121	0.0441	0.1269	0.4410
Disagreement	1255	0.9489	0.0142	0.8975	0.9919
Number of Comments	1255	3070	1906	16	18745
Number of Commenting Users	1255	1869	1075	20	8220
Return	1255	0.0019	0.0384	-0.3717	0.1875
Abnormal Trading Volume	1255	0.1774	0.5743	-0.6861	6.6917
Price Range	1255	0.0484	0.0380	0.0046	0.4894

We see several differences between the r/stocks and r/Bitcoin subreddits. First, the average sentiment in the stock market community is higher compared to the Bitcoin community, but the latter has a smaller range. In addition, disagreement in the Bitcoin community is slightly higher, with less dispersion and a smaller range. Overall, the Bitcoin community seems to be less optimistic but also has a persistently high level of disagreement. The number of comments and commenting users is also higher in the Bitcoin community. The average number of comments per day in the Bitcoin community is 3070, while the stock market community averages 2469 comments. Likewise, the Bitcoin community sees an average of 1869 users commenting per day, compared to 1585 in the stock market community.

Bitcoin exhibits significantly higher volatility and abnormal trading volume than the S&P 500. Bitcoin's average price range is 0.0484, with a standard deviation

of 0.0380, whereas the S&P 500's average price range is much lower at 0.0128. Similarly, Bitcoin's abnormal trading volume is also higher than that of the S&P 500. Summary statistics show the significantly higher volatility observed in the cryptocurrency market.

We run VAR models with one lag to examine whether social media disagreement in both communities affects future trading volume and volatility. The results in Table 14 show strong support for Hypothesis 5, which expects social media disagreement to positively affect the trading volume and volatility of the stock market. In contrast, Table 15 shows that the evidence for Hypothesis 6, which suggests that social media disagreement positively affects the trading volume and volatility of Bitcoin, is weaker.

The results for the r/stocks subreddit, reported in Table 14, show that social media disagreement significantly increases the subsequent abnormal trading volume and volatility of the stock market. More specifically, one standard deviation increase in disagreement increases the next day’s abnormal trading volume by 1.68% and volatility by five basis points. The volatility response roughly equals 4% of the sample mean volatility of 1.3%. In Table 15, we see that, unlike our expectations, disagreement in the Bitcoin community is not significantly associated with the next day’s trading volume. Nonetheless, it positively affects the next day’s volatility, which is weakly significant at the 10% significance level. The disagreement’s effect on volatility is slightly more pronounced for Bitcoin. One standard deviation increase in disagreement increases volatility by 16 basis points.

Table 14. Disagreement, Abnormal Trading Volume, and Volatility in the r/stocks

	Dependent Variable			
	Model 1: $avol$		Model 2: $prange$	
	$avol_t$	$disagt_t$	$prange_t$	$disagt_t$
$disagt_{t-1}$	0.0168**	0.3758***	0.0005**	0.3758***
	(0.0082)	(0.0321)	(0.0002)	(0.0321)
$lnnodes_{t-1}$	-0.0036	0.0950***	0.0004*	0.0950***
	(0.0074)	(0.0288)	(0.0002)	(0.0288)
$avol_{t-1}$	0.6665***	0.2898**	0.0057***	0.2898**
	(0.0305)	(0.1196)	(0.0009)	(0.1196)
r_{t-1}	0.1740	-2.1262	-0.0268*	-2.1262
	(0.4796)	(1.8807)	(0.0144)	(1.8807)
$prange_{t-1}$	4.7474***	16.1685***	0.6132***	16.1685***
	(1.0691)	(4.1920)	(0.0321)	(4.1920)
Constant	0.2806***	-0.4944***	-0.0008	-0.4944***
	(0.0253)	(0.0992)	(0.0008)	(0.0992)
Observations	858	858	858	858
Granger	4.20**	5.87**	4.62**	14.88***

The Granger row reports the χ^2 values from the Granger causality tests. Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

Table 15. Disagreement, Abnormal Trading Volume, and Volatility in the r/Bitcoin

	Dependent Variable			
	Model 1: $avol$		Model 2: $prange$	
	$avol_t$	$disagt_t$	$prange_t$	$disagt_t$
$disagt_{t-1}$	-0.0038	0.4396***	0.0016*	0.4396***
	(0.0107)	(0.0260)	(0.0010)	(0.0260)
$lnnodes_{t-1}$	-0.0130	-0.0587*	0.0086***	-0.0587*
	(0.0126)	(0.0304)	(0.0011)	(0.0304)
$avol_{t-1}$	0.7800***	-0.1288***	0.0024	-0.1288***
	(0.0200)	(0.0484)	(0.0018)	(0.0484)
r_{t-1}	0.1528	-0.2036	-0.0573**	-0.2036
	(0.2749)	(0.6663)	(0.0246)	(0.6663)
$prange_{t-1}$	-0.6350*	1.7411**	0.3350***	1.7411**
	(0.3488)	(0.8454)	(0.0312)	(0.8454)
Constant	0.0692***	-0.0611	0.0318***	-0.0611
	(0.0188)	(0.0456)	(0.0017)	(0.0456)
Observations	1247	1247	1247	1247
Granger	-0.0038	0.4396***	0.0016*	0.4396***

The Granger row reports the χ^2 values from the Granger causality tests. Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

Differences in the asset-specific characteristics and dynamics of online communities can help explain the differences in the impact of social media disagreement on the stock market and Bitcoin. Bitcoin exhibits higher volatility and greater return dispersion compared to the stock market. External factors such as regulatory changes can cause significant price movements in cryptocurrency markets and may weaken the effect of social media disagreement. While interactions within the Bitcoin community can be related to immediate price movements, they often involve long-term ideological debates, which may not necessarily correlate with immediate market activity. As a result, opinions in the Bitcoin community may become more static and entrenched. The Bitcoin community has many participants who follow an ideological perspective that values long-term trust in Bitcoin's

technology over short-term market conditions (Knittel et al., 2019). Therefore, participants in the Bitcoin community may be less likely to adjust their trading behavior based on social media discussions.

Indeed, our data shows that sentiment levels are more stable in the Bitcoin community than in the stock market community, with significantly lower average daily sentiment change. Shifts in sentiment are more closely associated with market activity and volatility of the stock market, while sentiment in the Bitcoin community is less likely to drive immediate trading behavior. Vlahavas and Vakali (2024) suggest that discussions in online cryptocurrency communities concentrate more on immediate market actions and focus less on regulatory issues during market downturns. Considering this finding, we further test whether the disagreement in the Bitcoin community is related to future market activity during unfavorable market conditions. To do this, we first calculate drawdowns in Bitcoin prices. Then, we run a regression model with an interaction term between social media disagreement and drawdowns. The estimation results that we do not report for brevity purposes are available upon request. Briefly, we see that disagreement decreases the next day's abnormal trading volume and volatility on days Bitcoin has no drawdowns. However, as we expected, the interaction term is significant and positive. Our results show that social media disagreement increases future abnormal trading volume and volatility as drawdowns increase. The possible shift of focus in discussions within the Bitcoin community toward immediate market actions during market downturns allows social media disagreement to play a more pronounced role in driving market activity in such periods.

In Tables 14 and 15, we see that investor attention also affects market activity. Increases in the number of users in the stock market community, which

serve as a proxy for investor attention, can only predict the increases in volatility, but the statistical significance is weak, with a p-value of 0.07. The effect size is similar to that of disagreement, with volatility increasing by approximately four basis points following a one-standard-deviation increase in the number of users. Furthermore, Table 15 shows that investor attention is more important for Bitcoin. One standard deviation increase in the number of users increases subsequent volatility by 86 basis points. This increase in volatility is consistent with the investor attention literature, which attributes this effect to the increase in information being processed into the prices because of increased information acquisition (Andrei & Hasler, 2015; Aouadi et al., 2013). Andrei and Hasler (2015) argue that information is gradually incorporated into prices when learning is slow, and investors pay little attention to the news, making greater investor attention lead to higher volatility. Social media participation and interactions can represent more attention to the market, which allows people to obtain and process more information and increases short-term volatility. Thus, our results imply that information transfer exists between participants in social media, and that information dissemination becomes easier as participation increases. Nevertheless, the results and our interpretation should be approached with caution because the statistical significance is weak for the stock market community.

We also detect that market dynamics affect the subsequent disagreement in social media. We see that prior volatility is particularly essential for both subreddits. Increases in market volatility lead to greater future disagreement, which, in turn, amplifies the subsequent volatility. Furthermore, prior trading volume and investor attention play significant roles for both communities, although with distinct implications for each. Following an abnormally high volume of transactions,

disagreement significantly increases in the stock market community but decreases in the Bitcoin community. We also observe the same pattern for the number of users who comment, which has a positive effect on the disagreement of the stock market community and a negative impact on the disagreement of the Bitcoin community. However, the latter is statistically significant only at the 0.1 level. These differences may arise from the distinct factors that contribute to the trading volume of each asset. Abnormally high trading volume in the stock market can signal that market participants interpret information differently or have diverse expectations about the future. On the other hand, an abnormally high Bitcoin volume may reflect periods of market stability, since there is significant controversy about its existence and future use. That is why higher volume can indicate widespread adoption, increased confidence in the asset, and reduced uncertainty about its future. Consistent with these differences, Table 14 shows that volatility has a positive effect on the future trading volume of the stock market, but Table 15 shows a negative impact of volatility on Bitcoin's volume. Furthermore, the r/stocks subreddit can have participants interested in different individual stocks or industries, whereas the r/Bitcoin subreddit can be relatively more homogenous. Participants of the Bitcoin community may share a common interest in adopting Bitcoin, which can enable them to converge with similar opinions following high trading volume. The impact of the number of users on disagreement is consistent with this argument. The results show that as more people comment about the stock market community and interact with one another, overall disagreement tends to increase, implying that giving a dissenting opinion and debating may be more common in the stock market community than in the Bitcoin community. On the contrary, disagreement tends to decrease in the Bitcoin community. It appears that participants in the Bitcoin

community have a greater tendency to be in an echo chamber because Bitcoin is more controversial as an asset class than stocks.

Our findings show that social media dynamics affect financial markets to some degree of confidence. The positive effect of disagreement on abnormal trading volume and volatility implies that divergence in people's sentiments in online communities can serve as a proxy for the overall disagreement in the markets. Additionally, the effect of social media participation on volatility suggests that people disseminate information among social media users. More involvement in online financial communities induces additional volatility in the near term, suggesting that people can get more information due to online social interactions. We see that the social media activity of the previous day can influence the markets, but evaluating the timing of interactions can lead to a better understanding of the implications of social media on financial markets. Social media disagreement before trading activity begins may provide additional insights if pre-market discussions have more future-oriented content. To test this possibility, we conduct a separate examination of the period from midnight until the markets open. Although Bitcoin exchanges do not have official trading hours, we apply the same methodology to see whether we still observe differences between the two asset classes. Table 16 reports the results of Hypothesis 5 for the r/stocks subreddit, and Table 17 reports the results of Hypothesis 6 for the r/Bitcoin subreddit.

In Table 16, we see that the effect of before-the-market disagreement on abnormal trading volume and market volatility is similar to that of one-day lagged disagreement. One standard deviation increase in disagreement before the market opens increases the abnormal trading volume by 1.26% and volatility by six basis points.

Table 16. The Effects of Before-Market Disagreement in the r/stocks

	$avol_t$	$prange_t$
$disagt_before_t$	0.0126**	0.0006***
	(0.0064)	(0.0002)
$lnnodes_{t-1}$	-0.0024	0.0004**
	(0.0063)	(0.0002)
r_{t-1}	0.1340	-0.0271
	(0.6137)	(0.0271)
$prange_{t-1}$	5.1443***	0.6204***
	(1.3929)	(0.0417)
$avol_{t-1}$	0.6654***	0.0056***
	(0.0506)	(0.0013)
Constant	0.2771***	-0.0008
	(0.0376)	(0.0011)
Observations	857	857
R^2	0.6414	0.6542

Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

The results of the Bitcoin community reported in Table 17 further demonstrate the importance of the timing of interactions. Although one-day-lagged social media dynamics do not have significant predictive power for Bitcoin, comments made from midnight until the market opening reveal pronounced and considerable effects. One standard deviation increase in the before-the-market disagreement in the Bitcoin community creates a 2.23% increase in abnormal trading volume and a 33-basis-points increase in volatility. These results show that even though there are no official trading hours for Bitcoin, the timing of interactions also matters for it, and pre-market interactions have more substantial effects. The results support our initial premise that the interactions before market opening are more future-oriented. Additionally, the fact that Bitcoin does not have official trading hours, and its community receives price feedback even during pre-market hours

suggests that the Bitcoin community has predictive power over the short term, with intraday implications.

Table 17. The Effects of Before-Market Disagreement in the r/Bitcoin

	$avol_t$	$prange_t$
$disagt_before_t$	0.0232***	0.0033***
	(0.0082)	(0.0012)
$lnnodes_{t-1}$	-0.0103	0.0088***
	(0.0249)	(0.0013)
r_{t-1}	0.1849	-0.0590*
	(0.4280)	(0.0352)
$prange_{t-1}$	-0.7645	0.3280***
	(0.8636)	(0.0517)
$avol_{t-1}$	0.7827***	0.0024
	(0.1010)	(0.0018)
Constant	0.0743**	0.0320***
	(0.0309)	(0.0024)
Observations	1244	1244
R^2	0.5943	0.2638

Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

In addition to investigating the effects of social media disagreement on abnormal trading volume and volatility, we also examine how it affects extreme market movements by isolating negative and positive return days. Further examining extreme movements may be helpful for predictive insights and anomaly detection. Considering this reason, we investigate the predictive power of social media disagreement over extremely negative or positive returns.

We conduct logistic regressions to test whether social media disagreement can successfully categorise extremely positive and negative days. Table 18 reports estimation results.

Table 18. Disagreement and Extreme Returns

	Panel A: r/stocks		Panel B: r/Bitcoin	
	ext_neg_t	ext_pos_t	ext_neg_t	ext_pos_t
$disagt_{t-1}$	0.719***	0.460	0.015	-0.030
	(0.267)	(0.363)	(0.157)	(0.184)
$avol_{t-1}$	0.559	1.744***	0.316	0.252
	(0.678)	(0.485)	(0.220)	(0.187)
$prange_{t-1}$	0.221	0.194	0.434***	0.461***
	(0.235)	(0.198)	(0.150)	(0.122)
ext_neg_{t-1}	-1.479*		0.000	
	(0.894)		(0.000)	
ext_pos_{t-1}		-1.802		-0.063
		(1.553)		(0.685)
Constant	-3.443***	-4.311***	-3.778***	-3.435***
	(0.216)	(0.311)	(0.202)	(0.168)
Observations	865	865	1224	1254

Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

In Panel A, we observe that social media disagreement is positively associated with the probability of an extremely negative return event in the stock market. More specifically, the likelihood of an extremely negative return event doubles following one standard deviation increase in disagreement. Higher levels of disagreement can be a sign of varying quality of information within the community. Increased disagreement among social media users can reflect uncertainty or conflicting interpretations of market conditions, increasing the likelihood of extreme negative events. This finding shows that social media dynamics can be a risk management tool. Disagreement is important when assessing the probability of extreme negative return events, which can prevent investors from suffering high drawdowns. Nevertheless, in Panel B, we see that social media disagreement does not have a statistically significant coefficient, demonstrating that social media dynamics cannot predict extreme return movements of Bitcoin.

We further investigate the relationship between social media disagreement and the stock market at the individual stock level. While investigating the effect of disagreement at the index level provides a broad overview of how social media dynamics impact overall market activity, examining the relationship at the individual stock level provides a detailed perspective and allows us to observe stock-specific effects that are not evident at the index level. Separately examining how individual stocks are affected by social media discussions also enhances the validity of our study and acts as a robustness check. Stocks with an active social media presence may experience more significant effects, as information can disseminate rapidly within their community, influencing market dynamics. Therefore, the impact of social media disagreement may be more substantial at the individual stock level than at the index level. We closely follow our methodology but adjust for additional variables such as market value, book-to-market ratio, and bid-ask spreads to account for stock-specific factors that can confound our results. We include these variables to control for stock-specific factors that may affect trading volume and volatility independently of social media disagreement, as they have been linked to trading volume and volatility due to factors such as visibility, liquidity, and information asymmetry (Aouadi et al., 2013; Ding & Hou, 2015). We are still testing hypothesis 5, investigating the effect of social media disagreement on abnormal trading volume and volatility. The results are reported in Table 19.

Table 19. Lagged Effects of Disagreement at the Individual Stock Level

	$avol_t$	$prange_t$
$disagt_{t-1}$	0.0592***	0.0008***
	(0.0107)	(0.0001)
r_{t-1}	2.4455***	-0.0277***
	(0.8900)	(0.0076)
$prange_{t-1}$	9.0845***	0.5475***
	(1.0071)	(0.0210)
$avol_{t-1}$	0.2152***	0.0000
	(0.0011)	(0.0000)
$lnmv_{t-1}$	-0.1241	-0.0023
	(0.1602)	(0.0020)
$lnbm_{t-1}$	-0.0788	0.0038***
	(0.1243)	(0.0013)
$spread_{t-1}$	0.4572	0.0363**
	(0.3198)	(0.0142)
Constant	1.1214	0.0451**
	(1.6157)	(0.0217)
Observations	30416	30416
R^2	0.8363	0.3927

Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

We see that stock-specific disagreement can significantly predict the next day's abnormal trading volume and volatility. One standard deviation increase in disagreement leads to a subsequent rise in abnormal trading volume of approximately 6%. This effect is much more significant compared to the abnormal trading volume response of 1.68% observed at the index level. As expected, we see more pronounced effects of disagreement on the stock market when we precisely match disagreement with individual stocks. At the index level, some community effects are attenuated, resulting in smaller increases in trading volume. We demonstrate similar findings for the impact of disagreement on stock volatility. One standard deviation increase in stock-specific disagreement induces an additional

eight-basis-points increase in the next day's volatility. This impact on stock volatility is 60% higher than the disagreement's effect on index volatility.

We also investigate whether disagreement's predictive power over extreme return movements exists at the individual stock level. Table 20 reports our results. We see that disagreement is still significantly associated with future extreme movements but provides less precision. Increases in disagreement are positively related to extreme negative returns at the 0.1 level of significance, but they are also positively associated with extreme positive returns at the 0.01 level. In other words, while an increase in disagreement is associated with a higher probability of subsequent extreme movement at the individual stock level, it cannot provide reliable information about its direction as it does at the index level. More specifically, one standard deviation increase in disagreement increases the probability of subsequent extremely negative returns by 7% and extremely positive returns by 12%. Social media disagreement is still a volatility-increasing factor, but its ability to act as a warning signal diminishes at the individual stock level.

Table 20. Disagreement and Extreme Returns at the Individual Stock Level

	ext_neg_t	ext_pos_t
$disagt_{t-1}$	0.0635*	0.1098***
	(0.0338)	(0.0312)
$lnmv_{t-1}$	0.5153***	-0.0411
	(0.0876)	(0.0743)
$lnbm_{t-1}$	0.0931	0.2206**
	(0.0927)	(0.0865)
$avol_{t-1}$	0.0707***	-0.0005
	(0.0185)	(0.0021)
$prange_{t-1}$	0.2192***	0.2186***
	(0.0180)	(0.0158)
ext_neg_{t-1}	0.4034***	
	(0.0956)	
ext_pos_{t-1}		0.0387
		(0.0956)
Observations	27091	27755

Asterisks represent statistical significance. ***<0.01, **<0.05, and *<0.1.

To test the robustness of our results, we conduct additional analyses by including neutral comments in the disagreement calculation. For the sake of brevity, we summarise these results without reporting complete estimations. We obtain results similar to those of our main models in most cases. At the stock market level, the effects of social media disagreement on the next day's volatility and the likelihood of extreme negative returns remain. We also still observe the significant effects of disagreement on both abnormal trading volume and volatility at the individual stock level. However, we observe some slight changes. Specifically, the effect of disagreement on the next day's abnormal trading volume in the stock market becomes statistically insignificant when neutral comments are included. Similarly, for Bitcoin, including neutral comments reveals before-the-market results to become

insignificant. Although the results are similar in most cases, they suggest that opinions without strong sentiment may weaken the measurement of disagreement.

In addition, we run a robustness check to test the impact of excluding comments that mention multiple stock tickers. Instead of removing these comments, we assign their sentiment scores to all referenced stocks. The results remain consistent with our primary findings and show that excluding comments with multiple stock tickers does not affect our overall results. We observe no significant changes in the effects of social media disagreement on abnormal trading volume or volatility.

Moreover, we perform a robustness check for our stock-level results by excluding stocks with only one comment per day to ensure that these minimal observations do not skew the results. The results remain consistent with our primary findings. The positive impact of social media disagreement on abnormal trading volume and volatility is robust to eliminating stocks with only one daily comment.

Finally, to see the impact of social media disagreement over a longer period, we rerun our VAR models using lags of up to five days. The results indicate that only the one-day lag model produces statistically significant effects on abnormal trading volume and volatility. The effects of social media disagreement on markets appear to be short-lived, with no significant results observed for longer lags. These findings demonstrate the short-term nature of the relationship between social media disagreement and market dynamics.

3.5 Conclusion

In this study, we examine how social media disagreement affects future trading volume and volatility in financial markets. We investigate this relationship in two

popular Reddit communities related to the stock market and Bitcoin to compare and examine whether different asset classes exhibit different community effects. Our findings show that social media disagreement is more influential in the stock market than Bitcoin. Specifically, disagreement is positively associated with increased trading volume and volatility in the stock market, whereas for Bitcoin, disagreement only weakly predicts future volatility. We also show that higher levels of social media disagreement increase the likelihood of extreme negative returns in the stock market. Disagreement at the individual stock level also creates more substantial increases in both trading volume and volatility compared to disagreement at the index level. We also demonstrate high autocorrelation in abnormal trading volume and price range, which emphasizes the persistence of these market measures.

Our study highlights the role of disagreement on social media platforms, which has received less attention than sentiment in prior research. We use established methodologies in the literature to explore discussions about different asset classes on social media platforms. Our results suggest that the dispersion of beliefs in online financial communities plays a significant role in market dynamics, particularly in the stock market, where disagreement drives short-term market behavior. We also demonstrate the differences between the stock market and Bitcoin communities. Unlike the stock market community, the Bitcoin community frequently discusses Bitcoin's underlying technology and usefulness and tends to be more ideologically driven. As a result, participants in the Bitcoin community are less likely to react to daily market fluctuations based on social media disagreements. However, our findings suggest that during significant market drawdowns, the impact of disagreement becomes more pronounced. A shift of focus in discussions in the

Bitcoin community from ideological discussions to market activity may explain the stronger influence of disagreement during downturns.

Our study has some limitations that should be acknowledged. First, we only focus on two communities on Reddit. Future studies can expand the scope to include other social media platforms or less popular subreddits and investigate other crypto assets, such as Ethereum or other less liquid altcoins. Moreover, exploring additional characteristics that drive differences between asset classes and communities could further refine our understanding of social media's influence on investor behavior. Lastly, understanding the behavioral or cultural factors that drive disagreement might contribute to a better understanding of the relationship between social media and financial markets.

CHAPTER 4

THEMATIC ATTENTION SHIFTS IN ONLINE DISCUSSIONS AND THEIR ROLE IN STOCK RETURN COMOVEMENT

This paper examines the determinants of thematic shifts in investor attention and their influence on stock return comovement. Using a large dataset of Reddit discussions, we track daily transitions between stock-level, market-wide, and inattentive themes to capture how attention responds to market and information environments. We find that attention shifts toward the stock level with increased firm disclosures and trading activity and toward market-wide themes during macroeconomic announcements. In contrast, online discussions become noisier in low-information, low-activity environments. These attention shifts have significant market effects. Shifts to stock-level attention reduce return comovement, while shifts to market-wide and inattentive discussions increase it. We further show that ticker mentions in discussions reduce comovement when they are embedded in stock-level discussions. Finally, shifts to stock-level attention are less effective at reducing comovement when they originate from market-wide or noisy themes. Together, our results emphasize the importance of the context and the flow of investor attention in affecting how information moves through markets.

4.1 Introduction

With limited attention capacity, investors cannot process all available information simultaneously and must choose where to focus (Barber & Odean, 2008; Peng & Xiong, 2006). This allocation of attention significantly affects how information is incorporated into prices and, in turn, stock return comovement (Guo et al., 2024; Hu

et al., 2021). While limited attention has received growing interest in the literature, most studies focus on its intensity rather than on how it shifts between themes in response to changing market and information environments.

In this paper, we introduce a novel framework to analyze thematic attention shifts in online investor discussions. Using a large dataset of Reddit posts from the r/stocks subreddit, we capture how interactions dynamically move between stock-level, market-wide, and noise/inattentive themes. We treat attention allocation as a dynamic process that evolves with the flow of information and market sentiment. This perspective allows us to examine both the drivers of attention reallocation and market consequences, particularly its effect on stock return comovement.

The motivation for our approach is twofold. First, financial social media platforms have become highly influential for retail investor activity, information exchange, and opinion formation. They also provide a distinct opportunity to inspect the real-time allocation of retail investors' attention. Second, return comovement is a powerful tool to indicate the reliance on firm-specific information versus the shared market-wide signals. Evaluating return comovement can deepen our understanding of how the structure and flow of attention influence information processing.

Our primary empirical framework uses Latent Dirichlet Allocation (LDA) to identify themes in investor discussions and construct attention shift probabilities across themes on a daily basis. We examine how these shifts respond to market activity, firm-specific information releases, and macroeconomic announcements. We then test whether different types of attention shifts are associated with stock return comovement. Finally, we use ticker mentions in discussions as a direct measure of investor attention to investigate whether their thematic context moderates their market effects.

Our findings reveal a systematic structure in attention allocation dynamics. Shifts to stock-level attention increase with firm disclosures and trading activity, while macroeconomic announcements divert attention toward market-wide themes. Inattentive discussions become less prominent in information-rich, high-activity environments. These attention shifts have significant market implications. Shifts to stock-level attention reduce return comovement, while shifts to market-wide and inattentive themes increase it. These results are robust to alternative attention classification strategies and comovement measures. We also find that both the breadth and volume of ticker mentions are negatively associated with return comovement, while coattention, measured by the daily number of stock pairs mentioned together, is positively associated. Nonetheless, we also show that ticker mentions reduce comovement when they occur in stock-level discussions but increase it if they are in market-wide discussions. Finally, we show that shifts into stock-level attention are less effective at reducing comovement when they originate from market-wide or inattentive discussions. Our thematic investigation emphasizes that not all stock-specific attention is equally informative, and the context and the flow of attention allocation also matter for its impact on return comovement.

Our study makes several distinct contributions to the literature on investor attention and stock return comovement. First, while Peng and Xiong (2006) provide a theoretical model showing that limited attention induces investors to adopt a category-level allocation, we apply this idea in an empirical framework by directly measuring how investor attention shifts between stock-level, market-wide, and inattentive themes in real-time online discussions. Second, while prior studies (Ding et al., 2020; Jiang et al., 2019; F. Su & Wang, 2021) emphasize investor attention's impact on return comovement, they do not consider the thematic context in which

this attention emerges. We fill this gap by showing that stock ticker mentions, a direct measure of stock-level attention, decrease return comovement when it occurs within stock-level discussions. Ticker mentions in market-wide discussions, on the other hand, increase return comovement. Third, while several recent studies rely on social media data to measure investor attention (e.g., Chen et al., 2024; Ding et al., 2020; Guo et al., 2024), they typically focus on attention intensity or user networks rather than how attention shifts thematically in response to changing environments. Our study fills this gap by applying topic modeling techniques to capture daily transitions across attention themes, allowing us to observe how investors dynamically reallocate their attention in response to market and information conditions.

The remainder of the paper is structured as follows. Section 4.2 reviews the related literature and develops the study's hypotheses. Section 4.3 describes the data, attention classification methods, and regression models. Section 4.4 presents the empirical findings on the determinants of attention shifts, their implications for return comovement, and robustness checks. Section 4.5 concludes the study.

4.2 Literature review and hypotheses development

People have limited attention as a cognitive resource (Kahneman, 1973). The constraint on attention implies that people cannot process all available information simultaneously. Thus, they must allocate their attention selectively, especially in financial markets, where stock prices, news, earnings reports, and macroeconomic trends generate a continuous influx of information (Peng & Xiong, 2006). This selective allocation can create periods of inattention, in which critical information is overlooked or incorporated into prices with delays. For example, when investors are

inattentive to firm-specific disclosures, stock prices adjust more slowly to new information (Dellavigna & Pollet, 2009; Hirshleifer et al., 2009). Attention constraint also influences investors' information sets (Barber & Odean, 2008), affects stock returns (Da et al., 2011), and plays a significant role in determining liquidity and volatility, as these are directly related to how efficiently information is disseminated in markets (Andrei & Hasler, 2015; Ding & Hou, 2015).

Because attention is limited, it is not randomly distributed but allocated in response to external economic conditions, market-wide developments, and individual preferences. Attention is drawn to salient events and stocks experiencing large price swings (Barber & Odean, 2008; Da et al., 2011). Studies show that investors allocate attention in categorical ways, such as focusing more on market-wide signals rather than firm-specific information during periods of uncertainty (Peng et al., 2007; Peng & Xiong, 2006). This reallocation of attention suggests that investors do not simply maintain fixed attention levels but dynamically shift their focus based on evolving conditions. Additionally, external distractions such as major sporting events or lottery jackpots also affect the level of investor engagement with firm-level information (Ehrmann & Jansen, 2022; Huang et al., 2019; Zhaunerchyk et al., 2020). These findings indicate that stock-level attention is not static and that attention shifts—rather than just attention levels—can meaningfully impact financial markets.

While existing research primarily examines the level of attention, the mechanisms and impacts of attention shifts are less explored. Attention can shift to the stock level with more focus on individual firms or to market-wide with a concentration on broad economic, financial, or geopolitical conditions. Alternatively, attention can shift to noise and off-topic content with limited information value.

Each type of shift can have a distinct influence on financial markets, and social media provides an opportunity to observe attention shifts in real time as user discussions move between firm-specific events, market-level themes, and broader macroeconomic or geopolitical issues. Prior studies establish that financial social media content influences attention-driven information dissemination (Ding et al., 2020; Jiang et al., 2019). Building on this literature, we focus on examining how online discussions evolve thematically in response to external conditions and how attention shifts between information-rich and inattentive content. The thematic flow of online conversations can reveal what factors affect attention allocation.

In particular, we expect attention shifts to be influenced by both macro-level shocks and the availability of firm-level information. Tantaopas et al. (2016) show that investor attention is significantly affected by volatility. Peng and Xiong (2006) suggest that investors focus more on market-wide signals rather than firm-specific information during periods of uncertainty. Market-wide volatility and economic uncertainty can divert attention away from stock-specific topics. When macroeconomic uncertainty increases, systemic risks are more likely to grab investor attention, which can reduce investor capacity for firm-level monitoring. Conversely, firm-level disclosures such as earnings announcements or regulatory filings can re-anchor attention toward specific stocks by introducing new information to process. These disclosures provide new signals that investors debate, which increases the likelihood of attention shifts toward stock-specific discussions. Additionally, periods of high market activity and firm-level disclosures are likely to increase investor engagement. The intensity of attention-grabbing developments can reduce the space for off-topic or inattentive discussions. In other words, when relevant information is actively entering the market, it competes for attention and diminishes the relative

presence of noise. Thus, taken together, we argue that attention shifts are not random but systematically shaped by information environments. We expect that shifts toward stock-level and inattentive discussions vary with both macro-level volatility and the arrival of firm-specific news. This leads to our first set of hypotheses:

Hypothesis 8a: Shifts to stock-level discussions decrease on days with elevated market volatility and major economic announcements.

Hypothesis 8b: Shifts to stock-level discussions increase on days with higher volumes of firm-level information releases.

Hypothesis 8c: Shifts to inattentive discussions decline during periods of increased market activity and firm-level information flow.

Beyond the factors that determine how attention is allocated, an important question is how these shifts influence market outcomes. A key mechanism through which attention allocation matters is return comovement. Prior research suggests that return comovement increases when investors rely on market-wide signals and decreases when they focus on firm-specific information (Peng et al., 2007). When attention shifts toward stock-level themes, investors are more likely to process idiosyncratic news, which reduces synchronicity in stock returns. This mechanism is supported across various proxies for investor attention, including media attention (Dong & Ni, 2014), social media attention (Ding et al., 2020), and internet search attention (Duz Tan, 2022). These findings suggest that shifts to stock-specific discussions can act as a force that reduces return comovement as investors process and trade on firm-specific signals rather than reacting to broad market trends. In addition, while shifts to stock-level discussions reflect a thematic orientation toward firm-specific content, the volume of stock mentions can also serve as a more direct indicator of how intensely investors are engaging with individual firms. A higher

number of ticker mentions suggests deeper attention to firm-level signals, which may further reinforce the reduction in return comovement.

However, the mechanisms linking attention and return comovement extend beyond individual attention allocation. Investors do not evaluate stocks in isolation but often focus on groups of stocks that share thematic or informational linkages (Barberis & Shleifer, 2003; Peng & Xiong, 2006). Coattention occurs when investors focus on multiple stocks within the same sector or narrative, leading to synchronized trading and stronger return comovement. Studies show that stocks frequently co-mentioned in online financial discussions, clustered in search behaviors, and connected in media coverage exhibit stronger return linkages beyond their fundamental connections (Agarwal et al., 2017; Chen et al., 2021; Jiang et al., 2019; Leung et al., 2017). Thus, we expect a negative association between stock-level attention and comovement, but a positive association between coattention and comovement.

Hypothesis 9a: Shifts to stock-level discussions are associated with lower return comovement.

Hypothesis 9b: Mentions of specific stocks in online discussions are associated with lower return comovement.

Hypothesis 9c: The co-mentioning of multiple stocks in online discussions is associated with higher return comovement.

While stock-specific attention can reduce return comovement, the context in which this attention occurs may influence its effectiveness. Likely, not all stock mentions reflect the same level of informational engagement, and how these mentions emerge affects the market impact. For instance, when investors engage with firm-specific news in the context of pre-allocated stock-level discussions, their

attention is more likely to be focused and informed by relevant information. In contrast, stock mentions that appear in broader market-wide or off-topic conversations may reflect more casual or sentiment-driven attention that lacks depth. Thus, even when two discussions mention the same stock, the differences in the thematic context of discussions may affect attention's impact on return comovement.

The thematic context of ticker mentions introduces a novel dimension to the literature on investor attention and return comovement. While prior studies document that stock-specific mentions reduce synchronicity, they largely treat attention as uniform across mentions. We contribute to the literature by considering the thematic framing of those mentions. Attention embedded in stock-level discussions likely represents more processing of idiosyncratic information. We, therefore, expect that stock mentions within stock-level discussions are more effective at reducing return comovement. This leads to our next hypothesis:

Hypothesis 10a: Mentions of specific stocks embedded in stock-level discussions are more effective at reducing return comovement than mentions within market-wide or inattentive discussions.

In addition, not all shifts toward stock-level discussions may be equally informative. A shift into stock-level themes from a thematically relevant discussion may indicate a deepening of focus on firm-specific content. In contrast, shifts from market-wide or inattentive discussions may reflect a low information processing capability. Such shifts may be triggered by news events but lack the context or continuity of stock-specific discourse. For instance, a user might casually reference a consumer stock to a discussion related to rising inflation without considering its actual earnings or exposure to cost pressures. In contrast, someone already discussing sector dynamics before narrowing the focus to a specific firm is more

likely to process firm-specific signals in depth. Similarly, stock mentions arising from off-topic conversations, such as jokes and memes, may reflect incidental attention rather than informed analysis. As a result, attention shifts toward stock-level discussions may be more effective at reducing return comovement when they build on prior stock-related focus rather than when they originate from broader or off-topic discussions.

Hypothesis 10b: Shifts to stock-level discussions have a weaker effect on reducing return comovement when they originate from market-wide or inattentive discussions.

4.3 Data and methodology

This study examines thematic attention shifts in online investor discussions and their relationship with stock return comovement. To analyze how attention moves between stock-level, market-wide, and inattentive discussions, we use comments posted on the r/stocks subreddit. Our dataset includes all comments posted between January 1, 2019, and June 30, 2022. We obtain stock-related data from Datastream.

Our methodology consists of four main components. First, we apply Latent Dirichlet Allocation (LDA) to classify Reddit comments into six topics, which are then grouped into broader attention categories. Second, we develop attention shift measures by analyzing comment-to-comment interactions and how discussions transition between different themes over time. Third, we extract stock ticker mentions from comment text to measure stock-level attention intensity and coattention. Fourth, we evaluate the market impact of attention using two return comovement metrics: R^2 values from market model regressions and return correlations, both computed over various time windows. As a robustness check, we

replicate our main analyses using alternative attention classifications based on a Random Forest classifier and a keyword-based approach.

The following subsections provide detailed information on topic modeling, comovement measures, and regression models.

4.3.1 Attention classification with latent dirichlet allocation

We use Latent Dirichlet Allocation (LDA) to identify primary themes in Reddit comments. LDA provides a probabilistic framework to uncover latent topics based on word distributions. Since online discussions are unstructured and involve various market-related themes, our study benefits from this aspect of LDA. To apply the model, we first clean our comment data by removing URLs, emojis, mentions of usernames and subreddits, numbers, and special characters. Then, we tokenize the comments into individual words, remove stop words, and lemmatize tokens to their root forms. We filter out comments with fewer than three meaningful words to have a plausible classification.

LDA requires a predefined number of topics. To make a justified selection, we evaluate models with 3 to 15 topics on a random 10% sample of comments. Based on the coherence scores of these preliminary models and the topics' interpretability, we determine that six topics are the most appropriate for our sample. The six-topic model achieves a coherence score of 0.69, which reflects a strong thematic alignment.

Table 21 presents the six topics and the most commonly used representative keywords identified by LDA. We see that the first topic captures discussions surrounding the fundamentals of technology and growth-oriented stocks. We frequently see mentions of “growth,” “Tesla,” “tech,” “earning,” “business,” “pe,”

“Apple,” “Google,” “Amazon,” and “Microsoft.” The first topic seems primarily related to stock-level information about a specific industry. The second topic is defined by words like “buy,” “time,” “sell,” “long,” 'invest,' “hold,” “lot,” and “trade.” The keywords reflect a focus on investment strategies, portfolio management, and trading ideas. The third topic includes keywords such as “oil,” “energy,” “green,” “power,” “war,” “world,” “Ukraine,” “global,” “Biden,” and “Elon.” The third topic is related to global and geopolitical issues and reflects an allocation to developments affecting the entire market.

Table 21. Topics Identified by Latent Dirichlet Allocation (LDA)

Topic Number	Topic Name	Keywords
1	Technology and Growth Stocks	company, growth, tesla, tech, earning, business, grow, pe, apple, car, new, revenue, vr, google, product, big, sale, msft, cheap, amazon, compare, microsoft, industry, build, china, amd, billion, service
2	Investment Strategies	stock, year, buy, go, think, time, good, money, sell, look, long, invest, right, like, hold, day, lot, start, well, month, term, way, lose, big, great, probably, trade, crash
3	Global and Geopolitical Issues	oil, country, energy, green, power, medium, crypto, war, qqq, world, social, gas, ukraine, street, american, global, baba, biden, wall, scale, elon, america, tqqq, lesson, garbage, political, ban, direction
4	Market Fundamentals and Valuations	market, price, high, share, mean, value, low, short, investor, cost, yes, reason, change, increase, expect, base, gt, future, close, valuation, number, cap, likely, current, dollar, issue, matter, large
5	Economic and Monetary Policies	rate, pay, cash, fund, inflation, interest, account, fed, dividend, bank, debt, margin, free, bond, economy, house, spy, tax, raise, loan, asset, home, income, government, hedge, capital, yield, flow
6	Casual Talk	not, people, like, know, thing, say, need, want, work, make, lol, try, use, actually, sure, see, tell, post, yeah, will, think, talk, real, find, understand, agree, stop, read, wrong

This table presents the six topics identified using Latent Dirichlet Allocation (LDA), each labeled based on representative keywords and contextual interpretation. For subsequent analysis, we group these topics into three broader attention categories: Topics 1 and 2 fall under stock-level attention, Topics 3, 4, and 5 reflect market-wide attention, and Topic 6 represents noise/inattention.

The fourth topic, with keywords like “market,” “price,” “value,” “investor,” “future,” and “valuation,” aligns with market fundamentals and valuation discussions. The keywords such as “rate,” “inflation,” “interest,” “account,” “fed,” “debt,” and “economy” in the fifth topic indicate that it is concerned with economic and monetary policies. Like the third topic, the fifth one also relates to information affecting the entire market. Finally, the sixth topic, encompassing keywords like “like,” “thing,” “people,” “know,” and “lol,” captures casual and low-information discussions that are more social in nature. We use common keywords and contextual interpretation and label these topics as follows: “Technology and Growth Stocks,” “Investment Strategies,” “Global and Geopolitical Issues,” “Market Fundamentals and Valuations,” “Economic and Monetary Policies,“ and “Casual Talk.”

Figure 8 displays the 7-day moving averages of each topic’s share in daily comments to demonstrate how investor attention fluctuates over time. We see that the proportion of comments allocated to each topic fluctuates significantly. The “Investment Strategies” topic dominates discussions, accounting for approximately 60% of all comments on average. This share reflects retail investors’ focus on trading ideas and is consistent with the view that investing is a social activity (Shiller et al., 1984). We also see that the Casual Talk topic ranked second, with an average share of around 29%. The Market Fundamentals and Valuations topic represents 8% of the total comments, whereas the Technology and Growth Stocks topic averaged 4%. Broader themes, such as Economic and Monetary Policies and Global and Geopolitical Issues, were less prominent, accounting for 1% and 0.1% of the discussions.

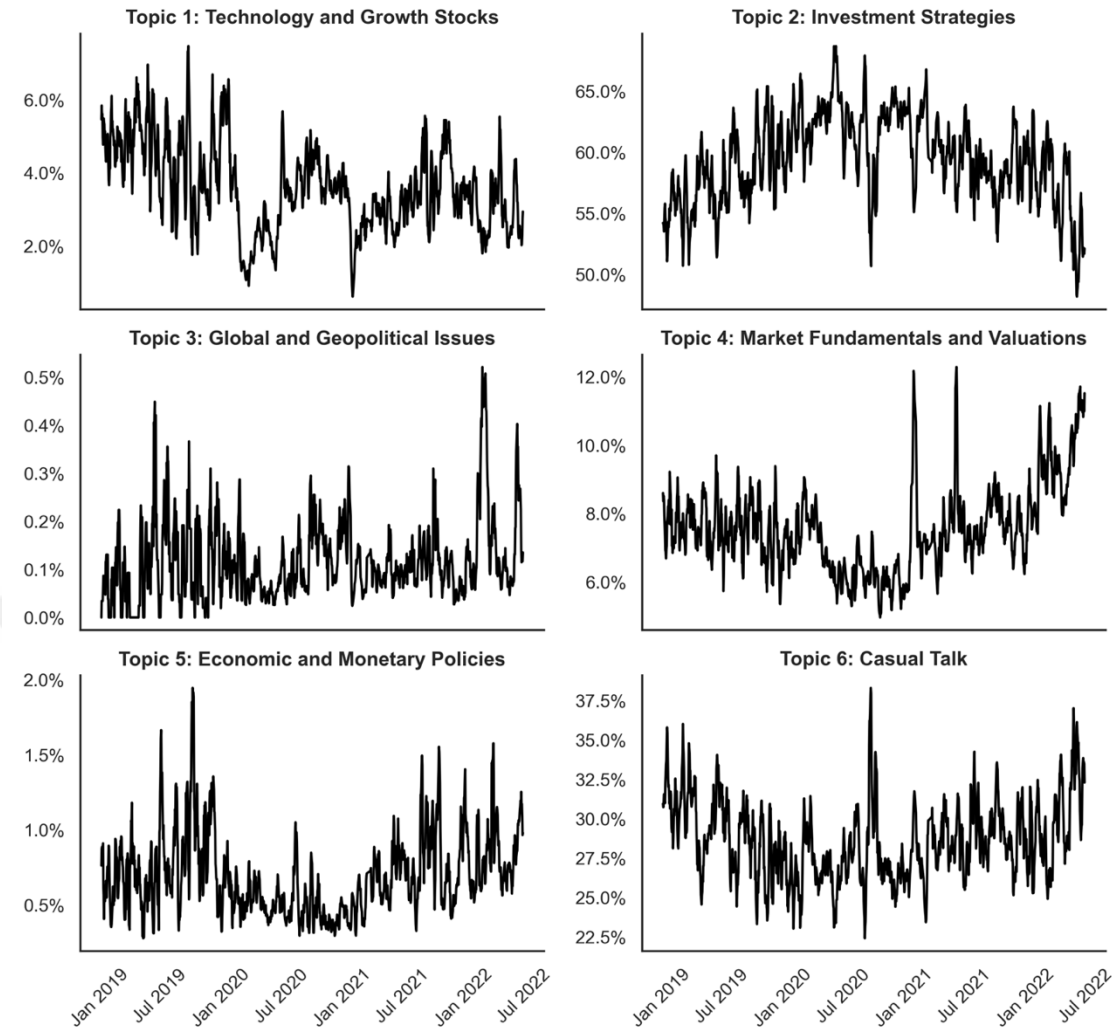

Figure 8. Distribution of topics over time
This figure shows the daily distribution of comment topics over time, based on dominant topic assignments from Latent Dirichlet Allocation (LDA). Each line represents the share of comments assigned to each topic per day

Nevertheless, these less prominent themes exhibit significant spikes during critical events and reflect investors' dynamic and responsive allocation of attention. As shown in Figure 9, the share of the Economic and Monetary Policies topic increases notably following FED rate changes, such as the rate cuts in early 2020 and subsequent hikes in 2022. Similarly, Figure 10 highlights spikes in the Global and Geopolitical Issues topic during significant events like the Gulf of Oman tanker attacks (June 2019), COVID-19 lockdowns (March 2020), and the war in Ukraine

(February 2022). These dynamic variations in attention allocation highlight the importance of examining how attention shifts influence investor behavior and market dynamics.

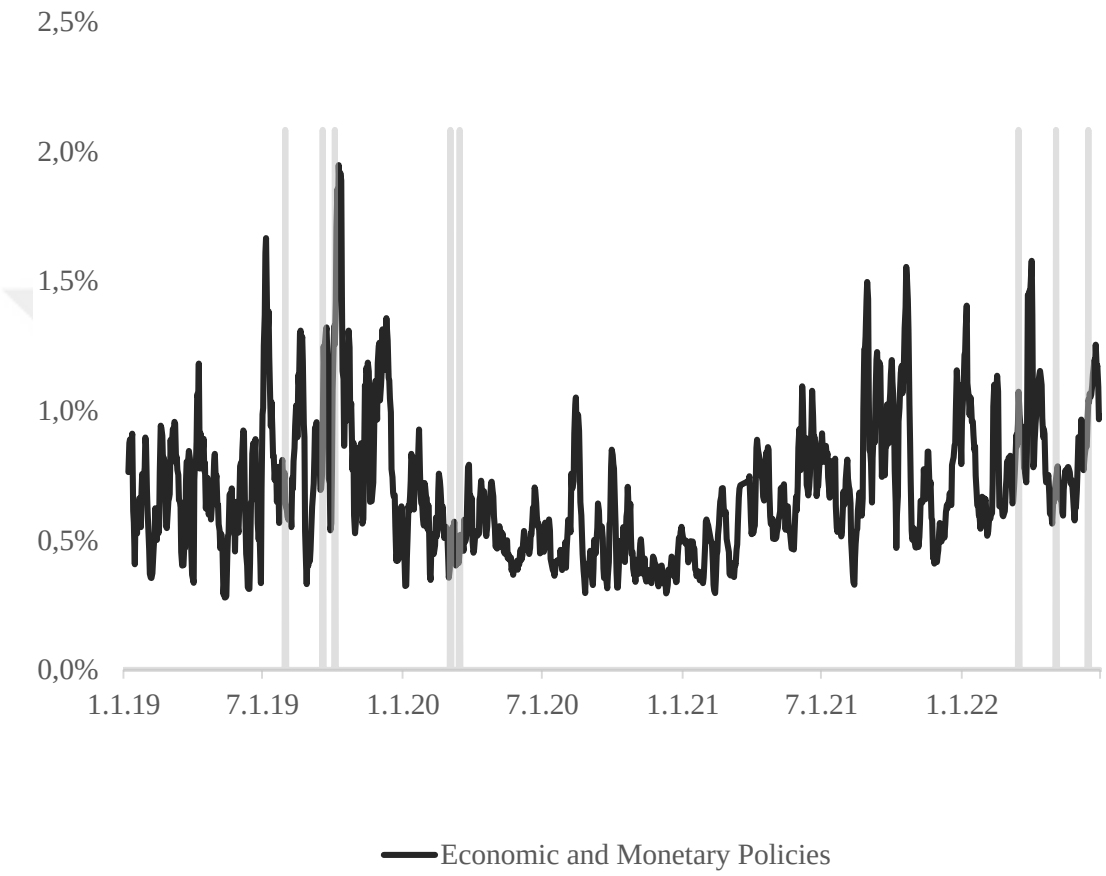

Figure 9. Attention to economic issues and the FED policy rate changes
Vertical lines indicate FED rate change announcements.

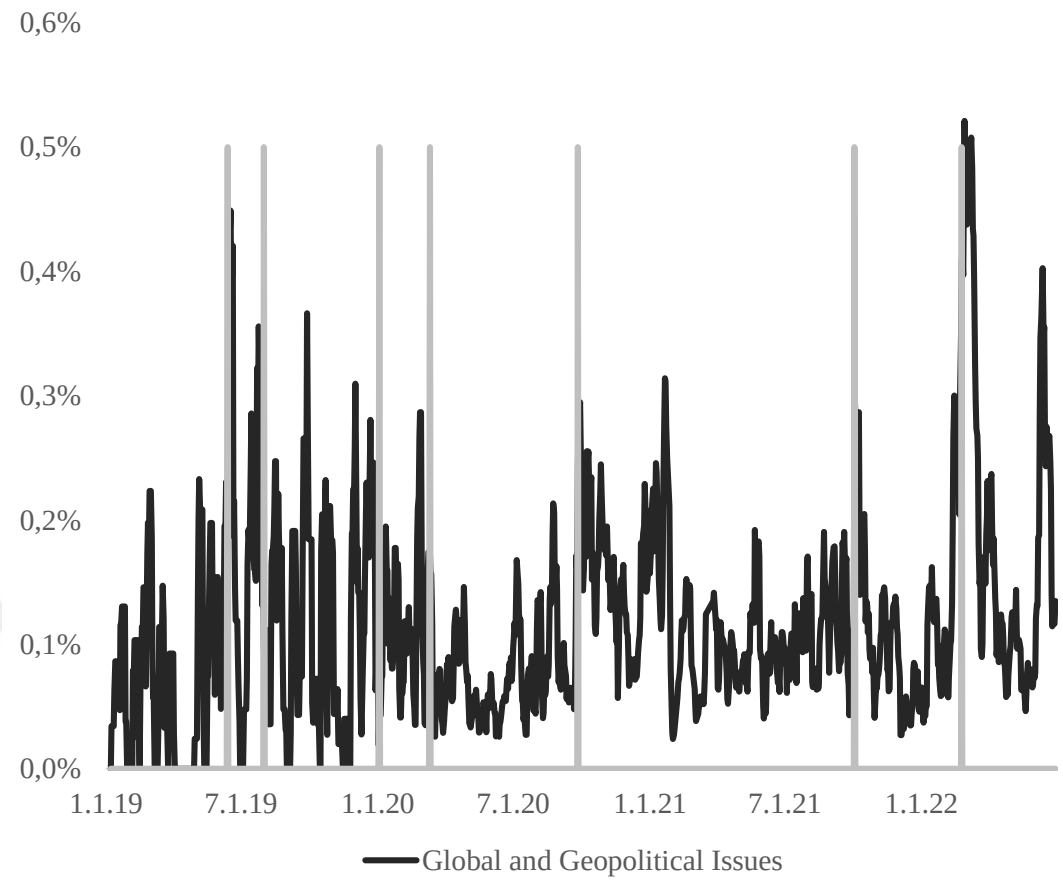

Figure 10. Attention to global issues and significant geopolitical events
Significant geopolitical events are shown in vertical lines. We identify Gulf of Oman tanker attacks, U.S.–China trade war escalation, the assassination of Qasem Soleimani, the COVID-19 Pandemic, the second Nagorno-Karabakh War, the OPEC meeting about an increase in oil output, and the start of the Russia-Ukraine war as significant geopolitical events that caused a peak in comments on global and geopolitical issues.

We further group six LDA topics into broader attention categories based on their primary focus and consistent with the literature: stock-level attention, market-wide attention, and noise/inattention. Stock-level attention includes discussions that focus on individual companies, their fundamentals, and investment strategies related to specific stocks. We categorize the topics of Technology and Growth Stocks and Investment Strategies under this group, as these discussions often reflect how retail investors evaluate and select specific stocks. Market-wide attention represents investor discussions that focus on broader economic, geopolitical, and monetary

policy developments that affect the stock market as a whole. This category includes the topics of Global and Geopolitical Issues, Market Fundamentals and Valuations, and Economic and Monetary Policies. References to macroeconomic indicators such as inflation, interest rates, central bank policy, and global market events characterize these discussions. Unlike stock-level attention, which emphasizes firm-specific signals, market-wide attention reflects how investors process systemic risks and macroeconomic conditions when making investment decisions. The final category is noise/inattention, which captures non-financial or low-information content that is common in online communities. This category includes casual conversations, memes, social discussions, and unrelated topics that do not contribute directly to investor decision-making. Grouping LDA topics into these three broad categories allows us to examine thematic attention shifts at both a general and detailed level.

4.3.2 Attention classification with ChatGPT assistance

In addition to the LDA-based topic classification, we develop two independent ChatGPT-assisted classification methods as alternative attention classification frameworks. The alternative frameworks serve as robustness checks by constructing attention categories using different techniques. The first approach uses a supervised classification model using Random Forests, and the second one follows a keyword-based classification framework.

In the first approach, we randomly sample 1% of comments and ask ChatGPT to classify them into one of the three broad categories based on the context of the comment and relevant keywords. We also ask it to give a confidence level between 0 and 1 for its classification. We drop comments with confidence levels smaller than 0.5 and use ChatGPT-labeled comments to train a random forest

classifier. This approach allows us to replicate attention category assignments using supervised machine learning with human-like categorization based on semantic meaning, context, and topic coherence**.** To prepare the data for classification, we transform the comment text into numerical features using Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF transformation captures the importance of words within each comment while reducing the impact of commonly used terms. We extract the 5,000 most relevant features and use them as input variables for the model. The Random Forest classifier is trained on 90% of the labeled dataset, while the remaining 10% is set aside for testing. The model achieves an accuracy score of 93% on the test set, indicating that the ChatGPT-provided labels are learnable from textual features. Using the trained model, we predict a classification for each comment in the dataset and compare them with LDA-based categories. The results show that the Random Forest model classifies comments into the same category as LDA in 52% of cases.

In the second approach, we employ a keyword-based classification framework. Again, we use ChatGPT to generate an extensive list of keywords for stock-level attention, market-wide attention, and noise/inattention. We then scan the dataset and classify each comment based on keyword matches. We assign a comment to the most frequently matched topic if it contains multiple topic-related keywords. This method successfully classifies 62% of comments; the others do not match predefined keywords. Among the classified comments, 57% of the results overlap with the LDA-based classification.

The alternative classification approaches indicate that while LDA classifications are not identical to ChatGPT-based methods, they align well at the broader category level. The fact that LDA matches keyword-based classifications at

a higher rate than the supervised model suggests that word distributions strongly influence topic assignments. Although LDA operates on a probabilistic framework and lacks contextual awareness, its classifications align meaningfully with alternative classification techniques. We re-estimate our main models using these ChatGPT-based attention indices to evaluate whether our findings are robust to alternative attention classifications.

4.3.3 Thematic attention shifts

We investigate comment-to-comment interactions to examine how investors shift their attention between different themes. For each interaction between comments, we define a "parent category" as the dominant attention category assigned by LDA to the initial comment and a "child category" as the dominant attention category of the reply. We analyze how frequently discussions transition from one category to another and quantify the likelihood that an investor conversation shifts between specific categories. The probability of shifting from category k to category j on a given day d is calculated as:

$$S_{k,j,d} = \frac{N_{k\to j,d}}{\sum_{j=1}^{K} N_{k\to j,d}} \tag{4.1}$$

where $N_{k\to j,d}$ represents the number of observed transitions from category k to category j on day d, and the denominator represents the total number of thematic transitions from category k on that day. Our normalization ensures that transition probabilities are interpreted as conditional probabilities. In other words, transition probabilities show the likelihood of shifting to another category, given that the discussion starts as category k. For example, we are interested in the probability of

stock-level attention shifting to market-wide attention, noise/inattention, or remaining in the same category.

Since discussions often contain elements of multiple themes, we conduct a robustness check using probability-weighted transitions at the attention category level. Rather than assigning each comment to a single dominant category, we aggregate its full topic distribution into category-level probabilities. For instance, a comment might be classified by LDA as 40% stock-level, 10% market-wide, and 50% noise/inattention. Instead of assigning this comment to just noise/inattention category, we allow it to contribute proportionally to all three based on these probabilities. This approach captures the possibility that secondary themes influence the conversation flow. For each comment interaction, we compute the transition probability from one category to another by weighting the influence of both the parent and child comments according to their category-level probabilities. Given a parent comment with probabilities $P_{i,k}$ across categories and a child comment with probabilities $C_{i,j}$, the probability-weighted transition score from category k to category j is computed as:

$$S_{k,j,d} = \frac{\sum_{i \in P} P_{i,k} \times C_{i,j}}{\sum_{i \in P} P_{i,k}} \tag{4.2}$$

where P is the set of all parent comments on day d. This approach ensures that comments with mixed thematic content contribute proportionally to multiple category shifts. For example, if a parent comment has a 60% probability of being stock-level and a child comment has a 70% probability of being market-wide, this pair contributes 0.42 to the transition probability from stock-level to market-wide.

4.3.4 Stock ticker mentions

In addition to thematic classifications, we analyze the presence and intensity of stock-level attention by extracting stock ticker mentions from comment text. We

begin by compiling a list of all stocks traded on major U.S. exchanges during the sample period and filter out tickers that overlap with common English words (e.g., "IT," "ON") to avoid potential ambiguities. To quantify stock-level attention using ticker mentions, we use three complementary measures. First, we compute the number of distinct stock tickers mentioned in a given day to capture the breadth of attention. Second, we calculate the total number of comments mentioning at least one stock ticker to get the volume of attention. Lastly, we count the number of distinct stock pairs co-mentioned in the same comment to measure coattention intensity.

Table 22 presents descriptive statistics for these stock-level attention measures. On average, 148 unique tickers are mentioned daily, with 517 comments mentioning at least one ticker. The number of co-mentioned stock pairs averages 248 per day, suggesting that investors frequently reference multiple stocks within the same discussion. The wide range and variance in these values reflect the dynamic nature of stock-specific attention, possibly driven by market events or company announcements.

To understand the attention themes of these mentions, we classify them according to the attention category of the comment. We find that 78% of all ticker mentions occur within the stock-level attention category. We see that retail investors most often mention individual stocks when exchanging trading ideas and discussing portfolio decisions. Only 5% of mentions appear within the market-wide attention category, suggesting that macro discourses typically occur without specific firm references. The remaining 17% occur in the noise/inattention category, where stock mentions are likely to be incidental rather than analytically driven.

Table 22. Descriptive Statistics of Stock Ticker Mentions

	Mean	St. Dev.	Q1	Median	Q3	Min.	Max.
$unique$	148	78	66	166	205	5	507
$count$	517	439	134	500	726	5	5049
$pairs$	387	389	99	321	551	1	4034

This table reports descriptive statistics for three stock-level attention measures based on ticker mentions in Reddit discussions. $unique$ refers to the number of unique stock tickers mentioned in a day. $count$ is the total number of comments that mention at least one ticker. $pairs$ represents the number of stock pairs co-mentioned within the same comment. Q1 and Q3 refer to the first and third quartiles, respectively.

4.3.5 Return comovement

Stock return comovement captures the degree to which individual stocks move together, and serves as a key indicator of whether investors rely on broad market signals or firm-specific information. High return comovement suggests that investors rely more on market-wide signals rather than firm-specific information, while lower comovement indicates greater incorporation of idiosyncratic factors into stock prices. We construct two measures to examine how thematic attention shifts influence stock return comovement: R^2 values from rolling market model regressions and rolling return correlations. Both R^2 values and correlation measures are heavily used in studies examining the relationship between attention and return comovement (Ding et al., 2020; Drake et al., 2017; Hu et al., 2021; Huang et al., 2019).

First, we estimate rolling R^2 values from market model regressions to capture the proportion of stock return variation explained by market movements. We construct a value-weighted index of all stocks mentioned in Reddit discussions and regress each individual stock's return on this index within a rolling window:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t} \quad (4.3)$$

where $r_{i,t}$ is the return of stock i, $r_{m,t}$ is the market return, and R^2 is obtained from the regression. We compute the rolling R^2 using a 20-day rolling window and then average them across all stocks as the primary measure of comovement on a given day. Higher R^2 values indicate that a larger share of stock return variation is explained by market-wide movements, which typically occurs when investors pay less attention to firm-specific fundamentals. As an alternative comovement measure, we compute rolling return correlations, which measure the average correlation between stocks' daily returns for all stock pairs within a 20-day rolling window.

We repeat these calculations using 10-day, 40-day, and 60-day rolling windows as robustness checks. Shorter rolling windows capture more reactive short-term movements, while longer windows assess whether trends in comovement persist over time. Alternative rolling windows ensure that our results are not driven by a specific window length.

4.3.6 Regression models

To test Hypotheses 8a through 8c, which examine how market-wide conditions and firm-level information affect the likelihood of shifting into different attention categories, we estimate a set of regressions in which the dependent variable is the daily probability of shifting into a specific attention category. For example, one regression focuses on the likelihood that discussions shift into stock-level attention, while others examine shifts into market-wide or noise/inattention categories. We regress these attention shift probabilities on a set of explanatory variables that capture both market conditions and information flow.

First, we compute the daily value-weighted return of all stocks mentioned in the r/stocks subreddit by ticker. We then use market returns in excess of the risk-free

rate to account for overall market performance. Second, we use abnormal trading volume, defined as the ratio of actual total trading volume to its 20-day moving average, to capture changes in investor activity. Third, we include market volatility, measured by the log difference between value-weighted indices of daily high and low stock prices, to capture overall uncertainty in the financial environment. Fourth, we add a macroeconomic announcement dummy, which equals one on days when major macroeconomic indicators are released. We focus on GDP, inflation, employment, and Federal Reserve announcements for this purpose. Finally, we construct a measure of firm-level information volume, defined as the total number of 8-K, 10-K, and 10-Q filings across all U.S. firms on a day, to capture the aggregate flow of firm-specific information to the market.

Our specification to examine determinants of attention shifts is as follows:

$$Shift_{k,t} = \alpha + \beta_1 mret_t + \beta_2 avol_t + \beta_3 prange_t + \beta_4 anon_econ_t + \beta_5 anon_stock_t + \varepsilon_t \quad (4.4)$$

where $Shift_{k,t}$ is the share of comment transitions on day t that move to category k, $mret$ represents excess market returns, $prange$ is the intraday price range, $avol$ reflects abnormal trading volume, $anon_econ$ is a binary variable that equals 1 on days with economic announcements, and $anon_stock$ measures the volume of stock-specific 8-K, 10-K, and 10-Q filings.

Hypotheses 9a through 9c predict that stock-level attention reduces return comovement, while coattention increases it. To test these hypotheses, we examine whether attention shifts, stock mentions, and coattention are associated with changes in stock return comovement. We estimate regressions in which the dependent variable is return comovement, measured by the rolling R^2 values.

We begin with regressing daily return comovement measures on the attention shifts probabilities. These probabilities show the likelihood that discussions shift into stock-level, market-wide, or inattentive categories. We then turn to stock-level attention metrics derived from ticker mentions. In separate regressions, we include the log number of unique tickers mentioned daily ($lnunique$) and the log number of total ticker mentions ($lncount$) as explanatory variables. To test the influence of coattention, we add the log number of co-mentioned stock pairs ($lnpairs$) and its interaction with $lncount$. This specification allows us to assess whether coattention weakens the market impact of stock-specific discussions.

All regressions include a common set of controls: market return in excess of the risk-free rate ($mret$), abnormal trading volume ($avol$), intraday market volatility ($prange$), a binary macroeconomic announcement indicator ($anon_econ$), and a measure of firm-level information flow ($anon_stock$). Our main specification is:

$$\begin{aligned} Comove_t = \alpha &+ \beta_1 Attention_t + \beta_2 mret_t + \beta_3 avol_t \\ &+ \beta_4 prange_t + \beta_5 anon_econ_t + \beta_6 anon_stock_t \\ &+ \varepsilon_t \end{aligned} \tag{4.5}$$

where $Attention_t$ is either the probability of shifting to a given attention category, volume of unique ticker mentions or total ticker mention volume. In another model, we augment the model with $lnpairs$ and its interaction with $lncount$, as well.

For Hypothesis 10a, we disaggregate the total number of daily ticker mentions ($lncount$) into three variables based on the attention category of the comment: $lncount_stock_level$, $lncount_market_wide$, and $lncount_noise$. This decomposition allows us to assess whether ticker mentions embedded in stock-level discussions are more effective at reducing return comovement than those embedded in market-wide or inattentive discussions.

For Hypothesis 10b, we examine whether the origin of stock-level attention shifts matters. Specifically, we construct a variable measuring the proportion of shifts into stock-level attention that originate from other categories. We then interact this proportion with the overall stock-level shift variable to create $external_shift_interaction$. This variable captures the extent to which the impact of attention shifts toward stock-level discussions depends on whether they emerge from different thematic contexts. The following regression specification summarizes our estimation.

We use these new variables as explanatory variables in Equation 4.5, replacing the main attention variable with either the decomposed mention counts or the interaction variable $external_shift_interaction$.

4.4 Findings

This section presents the study's empirical findings. We begin by examining the factors that influence investors' attention allocation. In particular, we test whether attention shifts into stock-level, market-wide, or noise/inattentive categories are systematically related to market conditions and information flow. Next, we assess the market consequences of attention allocation by analyzing the relationship between attention measures and return comovement. We then explore whether the thematic context of stock mentions and the origin of attention shifts moderate attention allocation's effect on return comovement. Finally, we test the robustness of our findings by using alternative attention classification approaches and different model specifications.

4.4.1 Determinants of attention shifts

Table 23 presents regression results examining the market-related and informational determinants of attention shifts in online investor discussions. The dependent variable in each panel is the daily probability of discussions shifting into one of the three attention categories: stock-level (Panel A), market-wide (Panel B), or noise/inattention (Panel C). The results offer strong support for the notion that investor attention is not randomly allocated but systematically influenced by market conditions and information flows.

Market-related variables reveal interesting patterns. Notably, excess market returns are negatively associated with shifts to stock-level attention and positively associated with shifts to noise/inattention. These attention shifts may reflect a sentiment-driven mechanism: on days when the market performs well, investor mood may become more relaxed and divert discussions toward casual or off-topic content rather than engaging with firm-specific fundamentals. Meanwhile, abnormal trading volume is positively associated with shifts to stock-level attention. One possible explanation is that greater disagreement reflected in the increased trading volume can divert discussions to specific companies. In contrast, abnormal trading volume is associated with declines in both market-wide and inattentive shifts, indicating that active trading environments concentrate investor attention on firm-level developments. Interestingly, market volatility, proxied by intraday price range, is not a significant predictor of shifts to stock-level or market-wide attention. This suggests that volatility, in the absence of specific informational triggers, may not meaningfully affect attention reallocation. However, it is weakly associated with a reduction in shifts to noise/inattention. This suggests that volatile conditions may

modestly suppress casual conversations in favor of more information-sensitive themes.

Furthermore, information flow variables reveal expected results. Specifically, we find that macroeconomic announcements significantly reduce the probability of attention shifts into stock-level discussions while simultaneously increasing the likelihood of shifts into market-wide themes. This pattern confirms that macro events, such as GDP or inflation announcements, divert investor attention toward aggregate conditions. We also observe that the volume of firm-specific disclosures is positively and significantly associated with shifts toward stock-level attention. Conversely, stock-specific disclosure volume is negatively related to shifts into inattentive content but has insignificant effects on shifts to market-wide attention. This result suggests that firm-level information flows decrease inattentiveness and divert attention to stock-level information.

Table 23. The Effect of Market factors on Attention Shifts in Online Interactions

	Panel A: Shifts to Stock-Level Attention	Panel B: Shifts to Market-Wide Attention	Panel C: Shifts to Noise/Inattention
$mret$	-0.3844***	0.0011	0.3867***
	(0.1425)	(0.0627)	(0.1263)
$avol$	0.0366***	-0.0110**	-0.0254**
	(0.0111)	(0.0049)	(0.0099)
$prange$	0.1766	0.0751	-0.2560*
	(0.1738)	(0.0765)	(0.1541)
$anon_econ$	-0.0090**	0.0040**	0.0052
	(0.0042)	(0.0019)	(0.0037)
$anon_stock$	0.0001***	-0.0000	-0.0000***
	(0.0000)	(0.0000)	(0.0000)
Constant	0.5789***	0.0864***	0.3349***
	(0.0059)	(0.0026)	(0.0052)
Observations	682	681	682
R^2	0.0535	0.0160	0.0504

This table presents regression results examining how market-level variables and announcements influence the probability of shifting to different attention categories in online interactions. Panel A shows shifts to stock-level attention, Panel B to market-wide attention, and Panel C to noise/inattention. Independent variables include daily market returns ($mret$), abnormal trading volume ($avol$), and the intraday price range ($prange$). $anon_econ$ is a binary variable that equals 1 on days with economic announcements. $anon_stock$ measures the volume of stock-specific 8-K, 10-K, and 10-Q filings. The dependent variable in each panel is the daily probability of shifting to the corresponding attention category. Standard errors are shown in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Overall, these findings reveal a systematic and intuitive structure to investors' dynamic attention allocation and provide strong support for our first set of hypotheses. Consistent with Hypothesis 8a, attention shifts toward stock-level discussions decline in the presence of major macroeconomic announcements, which instead redirect focus to market-wide themes. However, we do not observe a

statistically significant effect of market volatility on shifts to stock-level discussions. In line with Hypothesis 8b, we find that firm-specific information releases encourage investors to shift attention toward individual companies and significantly increase the probability of stock-level attention shifts. Finally, Hypothesis 8c is supported by the negative association between both abnormal trading volume and firm-specific disclosure volume with inattentive shifts. Active information environments appear to reduce the space for casual or off-topic conversations.

4.4.2 Market implications of attention allocation

We now turn to the implications of investor attention allocation for return comovement. Specifically, we test whether shifts in attention, stock mentions in discussions, and coattention are associated with the degree to which stock returns move together.

Table 24 presents the results of regressions testing the relationship between attention shifts and return comovement. We find that shifts toward stock-level attention are significantly associated with a reduction in return comovement, while shifts toward market-wide attention and noise/inattention significantly increase comovement. These results support Hypothesis 2a, which predicts that stock-level attention facilitates the incorporation of idiosyncratic information and reduces return comovement. In contrast, when discussions concentrate on market-wide issues, investors may rely more heavily on shared signals instead of firm-specific information. Similarly, shifts to noise/inattentiveness can directly affect the incorporation of information by decreasing informational efficiency.

Table 24. Attention Shifts in Online Interactions and Return Comovement

	Dependent Variable: Return Comovement		
	Panel A: Shifts to Stock-Level Attention	Panel B: Shifts to Market-Wide Attention	Panel C: Shifts to Noise/Inattention
$category_shift$	-0.2552***	0.3662***	0.2350***
	(0.0502)	(0.1157)	(0.0570)
$mret$	0.1416	0.2391	0.1488
	(0.1871)	(0.1884)	(0.1886)
$avol$	-0.0972***	-0.1025***	-0.1006***
	(0.0147)	(0.0148)	(0.0147)
$prange$	5.0046***	4.9323***	5.0197***
	(0.2272)	(0.2300)	(0.2290)
$anon_econ$	-0.0047	-0.0039	-0.0037
	(0.0055)	(0.0056)	(0.0056)
$anon_stock$	-0.0001**	-0.0001***	-0.0001**
	(0.0000)	(0.0000)	(0.0000)
Constant	0.2578***	0.0784***	0.0313
	(0.0301)	(0.0126)	(0.0206)
Observations	682	681	682
R-squared	0.4357	0.4227	0.4285

This table reports regression results on the relationship between attention shifts in online investor interactions and stock return comovement. The dependent variable is the average daily R^2 from rolling market model regressions. Each panel examines one attention category as the key independent variable: Panel A focuses on shifts to stock-level attention, Panel B on shifts to market-wide attention, and Panel C on shifts to noise/inattentive discussions. We control for market returns ($mret$), abnormal trading volume ($avol$), intraday price range ($prange$), the presence of economic announcements ($anon_econ$), and the volume of stock-specific 8-K, 10-K, and 10-Q filings ($anon_stock$). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Together, these findings indicate that dynamic attention allocation in online investor discussions has significant implications for how information is processed in financial markets. Stock-level attention encourages idiosyncratic signal processing, whereas market-wide attention fosters the use of common signals. On the other hand,

shifts to noise or inattentive discussions damage the overall quality of information incorporation. The consistent patterns across all three categories emphasize the importance of attention allocation on the information processing and return comovement.

To get a more detailed view, we next examine attention shifts using the six LDA topics to examine whether particular discussion themes have distinct effects on return comovement. Table 25 reports the findings. Within the stock-level attention category, shifts to Technology and Growth Stocks and Investment Strategies significantly reduce return comovement and reinforce the support for Hypothesis 9a. These discussions often involve active consideration of stock selection, valuation, or trading strategies, which likely encourage the processing of firm-specific information.

On the other hand, results for market-wide topics are more mixed. Only shifts to Market Fundamentals and Valuation are associated with a significant increase in return comovement. In contrast, attention shifts to Global and Geopolitical Issues and Economic and Monetary Policies do not show a statistically significant effect. One potential explanation is that these two topics account for a small share of overall attention. With a daily proportion generally below 1%, they may have a limited influence at the aggregate level. Alternatively, their impact may be conditional, emerging only when discussions are tied to market-wide valuation signals.

Table 25. Thematic Attention Shifts in Online Interactions and Return Comovement

	Dependent Variable: Return Comovement					
	Shift to T1: Tech & Growth	Shift to T2: Invest. Strategies	Shift to T3: Global Issues	Shift to 4: Market Fund.	Shift to T5: Econ. Policies	Shift to T6: General Disc.
$topics_shift$	-0.2340*	-0.190***	0.6255	0.3995***	0.5432	0.2350***
	(0.1362)	(0.0470)	(1.7056)	(0.1259)	(0.4757)	(0.0570)
$mret$	0.2536	0.1569	0.0926	0.2371	0.1882	0.1488
	(0.1896)	(0.1885)	(0.2762)	(0.1884)	(0.1928)	(0.1886)
$avol$	-0.104***	-0.102***	-0.127***	-0.103***	-0.113***	-0.101***
	(0.0149)	(0.0147)	(0.0201)	(0.0148)	(0.0159)	(0.0147)
$prange$	4.8112***	5.1169***	5.3811***	4.9190***	4.9843***	5.0197***
	(0.2479)	(0.2319)	(0.3946)	(0.2302)	(0.2397)	(0.2290)
$anon_econ$	-0.0026	-0.0038	-0.0081	-0.0036	-0.0020	-0.0037
	(0.0056)	(0.0056)	(0.0082)	(0.0056)	(0.0059)	(0.0056)
$anon_stock$	-0.0001***	-0.0001**	-0.0000	-0.0001***	-0.0001**	-0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.1229***	0.2093***	0.0926***	0.0791***	0.1025***	0.0313
	(0.0110)	(0.0257)	(0.0136)	(0.0125)	(0.0093)	(0.0206)
Obs.	680	682	379	681	635	682
R^2	0.4174	0.4280	0.3701	0.4227	0.4170	0.4285

This table presents regression results on the relationship between thematic attention shifts in online interactions and stock return comovement. The dependent variable is the average daily R^2 from rolling market model regressions. Each column represents a separate regression in which the key independent variable is the probability of shifting to one of the six topics identified via Latent Dirichlet Allocation (LDA): Topic 1 (Technology and Growth Stocks), Topic 2 (Investment Strategies), Topic 3 (Global and Geopolitical Issues), Topic 4 (Market Fundamentals and Valuations), Topic 5 (Economic and Monetary Policies), and Topic 6 (Casual Talk). We control for market returns ($mret$), abnormal trading volume ($avol$), intraday price range ($prange$), the presence of economic announcements ($anon_econ$), and the volume of stock-specific 8-K, 10-K, and 10-Q filings ($anon_stock$). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Overall, the LDA topic-level analysis suggests that not all forms of market-wide attention contribute equally to return comovement. Discussions around

valuation and pricing fundamentals appear to have more market effects than broader macroeconomic or geopolitical comments.

While topic classifications allow us to examine the thematic context of attention, they may not fully capture the extent to which investors engage with specific firms. To complement our earlier analysis, we turn to direct measures of stock-level attention based on ticker mentions. Table 26 presents the regression results examining the relationship between ticker mentions and return comovement. Panel A shows that the log number of unique tickers mentioned in a day ($lnunique$) is negatively and significantly associated with return comovement. Panel B shows a similarly negative and significant association for the total number of ticker mentions ($lncount$). These findings are consistent with Hypothesis 9b and provide further support for the idea that investor attention allocation to firm-level information enhances the incorporation of idiosyncratic signals into prices. We see that this effect is valid for both the breadth and volume of stock-specific investor attention. In Panel C, we assess the role of coattention by including the number of co-mentioned ticker pairs ($lnpairs$) and its interaction with the total number of ticker mentions. The results reveal that coattention is positively and significantly associated with return comovement. This finding supports Hypothesis 9c and aligns with prior research suggesting that when multiple stocks are discussed together, it encourages category-based processing and leads to synchronized trading behavior. The significant influence of coattention likely reflects underlying commonalities in information demand, affected by the shared market events or sectors. The commonality in such information contexts can lead to more correlated expectations and investment behavior. However, the interaction term between total ticker mentions and

coattention is significantly negative, suggesting that the comovement-enhancing effect of coattention is decreased by a high volume of stock-level attention.

Table 26. Stock Ticker Mentions and Return Comovement

	Dependent Variable: Return Comovement		
	Panel A: Effect of Unique Ticker Count	Panel B: Effect of Total Ticker Count	Panel C: Interaction with Total Pair Count
$lnunique$	-0.0086**		
	(0.0040)		
$lncount$		-0.0094***	0.0091
		(0.0028)	(0.0131)
$lnpairs$			0.0330***
			(0.0109)
$lncount \times lnpairs$			-0.0049**
			(0.0020)
$mret$	0.2306	0.2188	0.2170
	(0.1891)	(0.1882)	(0.1871)
$avol$	-0.1102***	-0.1104***	-0.1070***
	(0.0149)	(0.0148)	(0.0148)
$prange$	5.1320***	5.2313***	5.1823***
	(0.2443)	(0.2429)	(0.2422)
$anon_econ$	-0.0023	-0.0025	-0.0025
	(0.0056)	(0.0056)	(0.0055)
$anon_stock$	-0.0001***	-0.0001**	-0.0001**
	(0.0000)	(0.0000)	(0.0000)
Constant	0.1465***	0.1565***	0.0323
	(0.0188)	(0.0157)	(0.0569)
R^2	0.4180	0.4240	0.4319

This table reports regression results on the relationship between stock-specific attention in online discussions and return comovement. The dependent variable is the average daily R^2 from rolling market model regressions. Panel A examines the effect of the log number of unique stock tickers mentioned ($lnunique$). Panel B focuses on the total number of ticker mention ($lncount$). Panel C includes both $lncount$ and the log of number of co-mentioned ticker pairs ($lnpairs$), along with their interaction term. All regressions control for market returns ($mret$), abnormal trading volume ($avol$), intraday price range ($prange$), the presence of economic announcements ($anon_econ$), and the volume of stock-specific 8-K, 10-K, and 10-Q filings ($anon_stock$). Each model has 682 observations. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Together, the results in Table 26 strengthen the view that stock-specific attention is significantly associated with information processing and comovement dynamics. Attention to individual stocks contributes to the processing of firm-specific information and weakens comovement, while coattention to multiple stocks reflects a shared information demand that reinforces synchronized investor behavior and contributes to return comovement.

4.4.3 Moderating role of thematic context and origins of attention

In this section, we examine whether the thematic context of ticker mentions and the origin of shifts to stock-level attention affect their relationship with return comovement. The goal is to assess whether all forms of stock-specific attention have equal informational value or whether their impact depends on where and how they arise within online discussions.

Panel A of Table 27 presents results for the effect of stock ticker mentions by their attention category on return comovement. We disaggregate the total number of daily ticker mentions into three variables based on whether the comment belongs to the stock-level, market-wide, or noise/inattention category. The results reveal a negative and statistically significant association between ticker mentions embedded in stock-level discussions and return comovement. This finding supports Hypothesis 10a and suggests that the informational effectiveness of ticker mentions is greater when they emerge in stock-level discussions. By contrast, ticker mentions embedded in market-wide discussions are positively and significantly associated with return comovement. This result indicates that referring to individual stocks within market-wide discussions may introduce a common frame for price action, possibly shaped by the market conditions. Moreover, mentions in inattentive discussions do not show a

statistically significant effect. Although specific stocks are mentioned in casual or off-topic contexts, they appear to neither facilitate nor systematically disrupt the processing of stock-level information, which emphasizes the importance of the context of ticker mentions.

Table 27. Attention Categories, Ticker Mentions, and Return Comovement

	Dependent Variable: Return Comovement	
	Panel A: Mentions by Attention Category	Panel B: Shifts to Stock-Level Attention from Other Categories
$lncount_stock_level$	-0.0451**	
	(0.0198)	
$lncount_market_wide$	0.0366**	
	(0.0173)	
$lncount_noise$	-0.0140	
	(0.0212)	
$to_stock_level_from_other$		0.3325**
		(0.1293)
Observations	157	682
R^2	0.4567	0.4198

This table reports regression results on the relationship between return comovement and attention dynamics in online discussions. The dependent variable is the average daily R^2 from rolling market model regressions. Panel A examines the effect of stock ticker mentions by attention category. The independent variables are the log number of ticker mentions from comments classified under stock-level, market-wide, and noise/inattention discussions. Panel B analyzes whether the impact of shifts to stock-level attention depends on its origin. It includes an interaction between the probability of shifting to stock-level attention and the fraction of those shifts originating from other categories. All models include controls for market returns, abnormal trading volume, intraday price range, economic announcements, and stock-related 8-K, 10-K, and 10-Q filings (coefficients not reported). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel B of Table 27 evaluates Hypothesis 10b by testing whether the origin of attention shifts into stock-level discussions influences their impact. We construct a

variable capturing the interaction between the overall share of attention shifting into stock-level discussions and the proportion of those shifts originating from other attention categories. The results reveal a positive and statistically significant association with return comovement. This finding supports Hypothesis 10b and suggests that shifts into stock-level attention are less effective at facilitating the incorporation of firm-specific information when they originate from market-wide or inattentive discussions. In other words, when attention shifts into stock-level discussions from market-wide or unrelated discussions, it may lack firm-specific depth, limiting its ability to reduce return comovement.

Taken together, these results highlight the importance of discussion themes and continuity in shaping the market relevance of attention. Thematic context and continuity both shape whether attention fosters or hinders the incorporation of firm-specific information.

4.4.4 Robustness checks

We conduct a series of robustness checks to ensure the validity of our main findings. These checks examine whether the observed relationships between attention dynamics and return comovement hold under alternative attention classifications and model specifications.

4.4.4.1 Alternative attention classification methods

To test the robustness of our results to different classification approaches, we replicate our primary analyses using two alternative attention measures based on ChatGPT-assisted classifications: a Random Forest classifier trained on labeled comments and a keyword-based categorization using predefined attention-related

terms. Table 28 reports the summary of results. Across both methods, we find consistent evidence that shifts toward stock-level attention are associated with a significant reduction in return comovement and shifts to market-wide attention are associated with a significant increase. While the Random Forest model also reproduces a significant positive effect of inattentive shifts, the keyword-based model yields a positive but statistically insignificant relationship. These findings suggest that the relationship between attention allocation and return comovement is robust across classification strategies, though sensitivity may vary for inattentive attention depending on the method used.

Table 28. Attention Shifts from Alternative Classification Methods

	Dependent Variable: Return Comovement	
	Panel A: Random Forest Model	Panel A: Keyword Model
to_stock_level	-0.2692***	-0.1778***
	(0.0599)	(0.0436)
to_market_wide	0.2688**	0.2210***
	(0.1111)	(0.0459)
to_noise	0.1902***	-0.0410
	(0.0604)	(0.0658)

This table reports robustness checks using alternative classification methods to construct attention shift measures. The dependent variable is the average daily R^2 from rolling market model regressions. The independent variables are the probabilities of shifting to stock-level, market-wide, and noise/inattentive discussions based on Random Forest (RF) and keyword-based category classifications. For brevity, only the shift-to-category coefficients are shown. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

4.4.4.2 Probabilistic attention shifts

As an alternative to the dominant topic assignment in LDA, we construct a probabilistic measure of attention shifts using the full topic distribution of each comment. This approach accounts for the possibility that a single comment may

reflect multiple thematic elements. Table 29 reports the summary of results. We find that shifts to stock-level attention, when measured probabilistically, are associated with a significant decrease in return comovement, while probabilistic shifts to both market-wide and inattentive categories are significantly associated with increased comovement. These results mirror the findings from our main specification and confirm that the relationship between attention allocation and comovement is not driven by the dominant topic labeling.

Table 29. Probabilistic Attention Shifts

	Dependent Variable: Return Comovement
	Probabilistic Attention Shifts
to_stock_level	-1.2823***
	(0.2054)
to_market_wide	1.0912***
	(0.2358)
to_noise	1.0858***
	(0.3317)

This table presents robustness checks using probability-weighted attention shift measures instead of dominant topic assignments. Transition probabilities are calculated based on the full topic distribution of each comment. The dependent variable is the average daily R^2 from rolling market model regressions. The independent variables are the probabilities of shifting to stock-level, market-wide, and noise/inattentive discussions. For brevity, only the shift-to-category coefficients are shown. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

4.4.4.3 Alternative comovement measures

To ensure that our results are not sensitive to how return comovement is measured, we repeat our primary analyses using an alternative metric and for different rolling windows. Panel A of Table 30 presents results based on average R^2 values from rolling market model regressions across rolling windows of 10, 20, 40, and 60 trading days. Panel B uses average pairwise return correlations across the same rolling windows. The results are broadly consistent with our main findings. Shifts to

stock-level attention are significantly associated with lower comovement across shorter rolling windows (10 to 40 days). In contrast, shifts to market-wide and inattentive attention are significantly associated with higher comovement over shorter windows, although these effects taper off over longer horizons.

Table 30. Alternative Return Comovement Measures

	10-Day RW	20-Day RW	40-Day RW	60 Day RW
Panel A: Comovement from average R^2				
to_stock_level	-0.2435***	-0.2552***	-0.1120**	-0.0213
	(0.0577)	(0.0502)	(0.0480)	(0.0480)
to_market_wide	0.4342***	0.3662***	0.2473**	0.0213
	(0.1319)	(0.1157)	(0.1094)	(0.1092)
to_noise	0.1974***	0.2350***	0.0829	0.0218
	(0.0655)	(0.0570)	(0.0543)	(0.0541)
Panel B: Comovement from average pairwise correlations				
to_stock_level	-0.2356***	-0.2299***	-0.0971**	-0.0090
	(0.0544)	(0.0455)	(0.0444)	(0.0440)
to_market_wide	0.3807***	0.2861***	0.1295	-0.0734
	(0.1246)	(0.1048)	(0.1012)	(0.1001)
to_noise	0.2032***	0.2252***	0.0947*	0.0305
	(0.0617)	(0.0515)	(0.0501)	(0.0496)

This table presents robustness checks using alternative measures of stock return comovement. Panel A reports results using the average daily R^2 values from rolling market model regressions, while Panel B uses average rolling return correlations. Each column corresponds to a different rolling window (RW) ranging from 10 to 60 days. The independent variables are the probabilities of shifting to stock-level, market-wide, and noise/inattentive discussions. For brevity, only the shift-to-category coefficients are shown. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Results for alternative rolling windows suggest that the market impact of thematic attention shifts is most pronounced in the short term, where shifts in discussion themes can quickly influence the way investors interpret and respond to

information. The consistency of the results across both comovement measures reinforces the robustness of our conclusions.

4.5 Conclusion

This study investigates how the themes of investor attention in online financial discussions shape stock return comovement. We introduce a novel framework that captures the dynamic reallocation of attention across stock-level, market-wide, and inattentive themes using a large dataset of Reddit interactions. In doing so, we focus more on the evolving structure of attention themes in response to market and information environments rather than the measures of attention intensity.

Our findings reveal a systematic pattern in how attention flows between themes in response to market conditions and information events. Shifts to stock-level attention increase following firm-specific disclosures and trading activity, but decline on days with major macroeconomic announcements. Inattentiveness, on the other hand, diminishes in more information-rich environments. These results suggest that investor attention responds intuitively to the informational environment.

Furthermore, we show that shifts in attention allocation have significant market consequences. We find that shifts to stock-level attention reduce return comovement, while shifts to market-wide and inattentive discussions increase it. These findings are robust across multiple attention classification methods and comovement measures. We further demonstrate that both the volume and breadth of ticker mentions in discussions are negatively associated with comovement, which strengthens the view that stock-level attention facilitates idiosyncratic information processing. In contrast, coattention to multiple stocks within the same discussion increases comovement, which may be reflecting the shared information environment between those stocks.

Moreover, we show that the thematic context and pathway of shifts to stock-specific attention matters. We find that ticker mentions in comments with stock-level attention reduce comovement, while those in market-wide discussions increase it. Finally, shifts into stock-level attention are less effective at reducing comovement when they originate from market-wide or inattentive discussions.

Our study contributes to multiple strands of literature. It extends the investor attention literature by focusing on how attention dynamically moves between themes rather than only measuring intensity. It also contributes to the literature on return comovement by demonstrating that themes of discussions, attention shifts, and their pathway critically influence how information is processed in prices. In conclusion, our study offers new tools to monitor investor attention in real time and suggests that thematic attention allocation provides a pathway through which information flows and affects prices.

CHAPTER 5

CONCLUSION

This dissertation explores how online social interactions shape investor beliefs and influence key outcomes in financial markets. By analyzing extensive discussion data from Reddit, the dissertation examines belief dissemination, disagreement, and attention allocation as key behavioral dynamics in online investor communities. Across three empirical studies, the findings provide a comprehensive view of how sentiment spreads, how disagreements evolve, and how thematic shifts in attention affect return comovement. Together, the studies highlight that online interactions are mechanisms through which financial beliefs are formed, transmitted, and acted upon, with observable consequences for market behavior.

The first study, presented in Chapter 2, investigates the dissemination of sentiment within a stock-related Reddit community. The findings show that highly active users exert significant influence over less active users, particularly when they have low internal disagreement or when market volatility is high. This dynamic became even more pronounced during the COVID-19 pandemic, when increased online engagement and market uncertainty reshaped network structures. Notably, sentiment disseminated through these interactions has short-term predictive power for stock market returns. The study also demonstrates the practical value of network sentiment by presenting a market timing strategy that reduces drawdowns. However, the influence within the network is not unidirectional; while central users lead sentiment, less active users can reinforce sentiment trends. These findings emphasize the complexity and multi-directionality of sentiment flow in social networks.

The second study, in Chapter 3, focuses on social media disagreement and its market effects, with a comparative perspective on traditional stocks and Bitcoin. Using cross-user sentiment dispersion as a measure of disagreement, the study finds that higher disagreement levels are strongly associated with increased trading volume, volatility, and extreme negative returns in the stock market. In contrast, disagreement in Bitcoin discussions has a weaker and more limited effect. The results also indicate that stock-level disagreement has a greater influence than index-level disagreement. This chapter highlights how belief dispersion plays a significant role in financial markets, especially in traditional equities, where investors are more responsive to short-term disagreement signals. It also shows how asset-specific community characteristics, such as ideological focus in Bitcoin discussions, can moderate the behavioral impact of disagreement.

Chapter 4 shifts the focus to investor attention and investigates how thematic shifts in discussion topics affect stock return comovement. Rather than measuring attention intensity, this study introduces a classification-based framework to monitor transitions between stock-level, market-wide, and inattentive themes in online discussions. The results show that investor attention adjusts intuitively to information environments. Specifically, attention shifts to stock-level themes after firm disclosures or unusual trading activity and toward market-wide themes during macroeconomic announcements. These shifts have clear implications for asset pricing. Specifically, stock-level attention reduces return comovement, while market-wide and inattentive themes increase it. The study also finds that the context and pathway of attention matter. Stock ticker mentions reduce comovement only when they appear in stock-specific discussions, and shifts into stock-level attention are less effective when they originate from broader or noisy themes. These findings highlight

the importance of not just what investors focus on, but also how and when attention is redirected within social environments.

Taken together, the three studies in this dissertation offer new insights into the behavioral mechanisms through which online discussions influence financial market outcomes. They show that sentiment, disagreement, and attention are not static or isolated variables, but emerge from interaction patterns, user roles, and thematic structures. Across chapters, the dissertation identifies online social activity as an important source of short-term market signals and emphasizes its relevance for return predictability, trading volume, volatility, and return co-movement. In doing so, the dissertation contributes to the growing body of literature on behavioral finance, investor attention, and digital social interactions in financial contexts.

The findings also have practical implications. From a trading perspective, incorporating network-based sentiment or attention data can improve market timing strategies and downside protection. For policymakers, the ability of a small number of users to influence large communities raises concerns about belief manipulation, herding, and rapid information diffusion in loosely moderated environments. For researchers, the dissertation presents replicable approaches for measuring belief dynamics and thematic attention flows using online textual and interaction data.

Nonetheless, several limitations should be acknowledged. The studies rely on data from Reddit, which, while rich and informative, may not fully represent the broader population of investors. Future research can expand the scope to include other platforms and investigate how belief dynamics differ across demographic, regional, or institutional investor groups. Additionally, user-level attributes such as reputation, identity, or external influence were not explicitly modeled but could play a meaningful role in understanding influence patterns. Lastly, while this dissertation

focuses on short-term market implications, longer-term outcomes of belief-driven trading remain important topics for future research.

In conclusion, this dissertation provides empirical evidence that online social interactions are central to modern investor behavior.

REFERENCES

Agarwal, A., Leung, A. C. M., Konana, P., & Kumar, A. (2017). Cosearch attention and stock return predictability in supply chains. *Information Systems Research*. https://doi.org/10.1287/isre.2016.0656

Ahern, K. R., Duchin, R., & Shumway, T. (2014). Peer effects in risk aversion and trust. *The Review of Financial Studies*, *27*(11), 3213–3240. https://doi.org/10.1093/rfs/hhu042

Ahn, Y., & Kim, D. (2020). Sentiment disagreement and Bitcoin price fluctuations: A psycholinguistic approach. *Applied Economics Letters*, *27*(5), 412–416. https://doi.org/10.1080/13504851.2019.1619013

Ajinkya, B. B., Atiase, R. K., & Gift, M. J. (1991). Volume of trading and the dispersion in financial analysts' earnings forecasts. *The Accounting Review*, *66*(2), 389–401.

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091. https://doi.org/10.1111/1540-6261.00454

Al-Nasseri, A., & Menla Ali, F. (2018). What does investors' online divergence of opinion tell us about stock returns and trading volume? *Journal of Business Research*, *86*, 166–178. https://doi.org/10.1016/j.jbusres.2018.01.006

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, *17*(2), 223–249. https://doi.org/10.1016/0304-405X(86)90065-6

Andrei, D., & Hasler, M. (2015). Investor attention and stock market volatility. *Review of Financial Studies*, *28*(1), 33–72. https://doi.org/10.1093/rfs/hhu059

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modelling*, *35*, 674–681. https://doi.org/10.1016/j.econmod.2013.08.034

Arrondel, L., Calvo-Pardo, H., Giannitsarou, C., & Haliassos, M. (2022). Informative social interactions. *Journal of Economic Behavior & Organization*, *203*, 246–263. https://doi.org/10.1016/j.jebo.2022.09.006

Atiase, R. K., & Bamber, L. S. (1994). Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of Accounting and Economics*, *17*(3), 309–329. https://doi.org/10.1016/0165-4101(94)90031-0

Atmaz, A., & Basak, S. (2018). Belief dispersion in the stock market. *The Journal of Finance*, *73*(3), 1225–1279. https://doi.org/10.1111/jofi.12618

Bailey, M., Cao, R., Kuchler, T., & Stroebel, J. (2018). The economic effects of social networks: Evidence from the Housing market. *Journal of Political Economy*, *126*(6), 2224–2276.

Bailey, M., Dávila, E., Kuchler, T., & Stroebel, J. (2019). House price beliefs and mortgage leverage choice. *The Review of Economic Studies*, *86*(6), 2403–2452. https://doi.org/10.1093/restud/rdy068

Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, *104*(2), 272–287. https://doi.org/10.1016/j.jfineco.2011.11.002

Bamber, L. S., Barron, O. E., & Stober, T. L. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review*, *72*(4), 575–597.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498. https://doi.org/10.1126/science.1236498

Banerjee, S. (2011). Learning from prices and the dispersion in beliefs. *The Review of Financial Studies*, *24*(9), 3025–3068. https://doi.org/10.1093/rfs/hhr050

Banerjee, S., & Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance*, *65*(4), 1269–1302. https://doi.org/10.1111/j.1540-6261.2010.01570.x

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161–199. https://doi.org/10.1016/S0304-405X(03)00064-3

Barron, O. E. (1995). Trading volume and belief revisions that differ among individual analysts. *The Accounting Review*, *70*(4), 581–597.

Bartoszek, G., Ranney, R. M., Curanovic, I., Costello, S. J., & Behar, E. (2022). Intolerance of uncertainty and information-seeking behavior: Experimental manipulation of threat relevance. *Behaviour Research and Therapy*, *154*, 104125. https://doi.org/10.1016/j.brat.2022.104125

Bayer, P., Mangum, K., & Roberts, J. W. (2021). Speculative fever: Investor contagion in the housing bubble. *American Economic Review*, *111*(2), 609–651. https://doi.org/10.1257/aer.20171611

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Blau, B. M., & Whitby, R. J. (2017). Range-based volatility, expected stock returns, and the low volatility anomaly. *PLOS ONE*, *12*(11), e0188517. https://doi.org/10.1371/journal.pone.0188517

Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, *2*(1), 113–120. https://doi.org/10.1080/0022250X.1972.9989806

Brown, J. R., Ivković, Z., Smith, P. A., & Weisbenner, S. (2008). Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance*, *63*(3), 1509–1531. https://doi.org/10.1111/j.1540-6261.2008.01364.x

Burnside, C., Eichenbaum, M., & Rebelo, S. (2016). Understanding booms and busts in housing markets. *Journal of Political Economy*, *124*(4), 1088–1147. https://doi.org/10.1086/686732

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, *82*(4), 1273–1301. https://doi.org/10.3982/ECTA11991

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*(2), 226–238. https://doi.org/10.1016/j.jfineco.2014.06.007

Charness, G., Karni, E., & Levin, D. (2013). Ambiguity attitudes and social interactions: An experimental investigation. *Journal of Risk and Uncertainty*, *46*(1), 1–25. https://doi.org/10.1007/s11166-012-9157-1

Chen, H., De, P., Hu, Y. (Jeffrey), & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

Chen, Z., Guo, L., & Tu, J. (2021). Media connection and return comovement. *Journal of Economic Dynamics and Control*, *130*, 104191. https://doi.org/10.1016/j.jedc.2021.104191

Chen, Z.-H., Wu, W.-L., Li, S.-P., Bao, K., & Koedijk, K. G. (2024). Social media information diffusion and excess stock returns co-movement. *International Review of Financial Analysis*, *91*, 103036. https://doi.org/10.1016/j.irfa.2023.103036

Cohen, L., Frazzini, A., & Malloy, C. (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy*, *116*(5), 951–979. https://doi.org/10.1086/592415

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *The Journal of Finance*, *75*(1), 173–228. https://doi.org/10.1111/jofi.12852

Cookson, J. A., & Niessner, M. (2023). Investor disagreement: Daily measures from social media (SSRN Scholarly Paper 4529594). https://doi.org/10.2139/ssrn.4529594

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dellavigna, S., & Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance*, *64*(2), 709–749. https://doi.org/10.1111/j.1540-6261.2009.01447.x

DeMarzo, P., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, *118*(3), 909–968. https://doi.org/10.1162/00335530360698469

Ding, R., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money*, *37*, 12–26. https://doi.org/10.1016/j.intfin.2015.04.001

Ding, R., Zhou, H., & Li, Y. (2020). Social media, financial reporting opacity, and return comovement: Evidence from Seeking Alpha. *Journal of Financial Markets*, *50*, 100511. https://doi.org/10.1016/j.finmar.2019.100511

Dong, Y., & Ni, C. (2014). Does limited attention constrain investors' acquisition of firm-specific information? *Journal of Business Finance & Accounting*, *41*(9–10), 1361–1392. https://doi.org/10.1111/jbfa.12098

Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The comovement of investor attention. *Management Science*, *63*(9), 2847–2867. https://doi.org/10.1287/mnsc.2016.2477

Duflo, E., & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics*, *85*(1), 121–148. https://doi.org/10.1016/S0047-2727(01)00098-6

Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, *118*(3), 815–842. https://doi.org/10.1162/00335530360698432

Duz Tan, S. (2022). Correlated information consumption and comovement in the stock market. *Borsa Istanbul Review*, *22*(4), 812–827. https://doi.org/10.1016/j.bir.2022.05.001

Ehrmann, M., & Jansen, D.-J. (2022). Stock return comovement when investors are distracted: More, and more homogeneous. *Journal of International Money and Finance*, *129*, 102742. https://doi.org/10.1016/j.jimonfin.2022.102742

Eierle, B., Klamer, S., & Muck, M. (2022). Does it really pay off for investors to consider information from social media? *International Review of Financial Analysis*, *81*, 102074. https://doi.org/10.1016/j.irfa.2022.102074

Ellison, G., & Fudenberg, D. (1995). Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, *110*(1), 93–125. https://doi.org/10.2307/2118512

Foster, F. D., & Viswanathan, S. (1993). Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models. *The Journal of Finance*, *48*(1), 187–211. https://doi.org/10.1111/j.1540-6261.1993.tb04706.x

Gan, B., Alexeev, V., & Yeung, D. (2024). Tweets versus broadsheets: Sentiment impact on stock markets around the world. *Journal of Financial Research*. https://doi.org/10.1111/jfir.12380

Giannini, R., Irvine, P., & Shu, T. (2019). The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. *Journal of Financial Markets*, *42*, 94–120. https://doi.org/10.1016/j.finmar.2018.12.003

Graham, J. R., & Harvey, C. R. (1996). Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. *Journal of Financial Economics*, *42*(3), 397–421. https://doi.org/10.1016/0304-405X(96)00878-1

Guo, F., Lyu, B., Lyu, X., & Zheng, J. (2024). Social media networks and stock price synchronicity: Evidence from a Chinese stock forum. *Abacus*. https://doi.org/10.1111/abac.12341

Han, B., & Hirshleifer, D. A. (2012). Self-enhancing transmission bias and active investing (SSRN Scholarly Paper 2032697). https://doi.org/10.2139/ssrn.2032697

Han, B., Hirshleifer, D., & Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, *57*(1), 390–412. https://doi.org/10.1017/S0022109021000077

Heimer, R. Z. (2014). Friends do let friends buy stocks actively. *Journal of Economic Behavior & Organization*, *107*, 527–540. https://doi.org/10.1016/j.jebo.2014.04.019

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies*, *29*(11), 3177–3209. https://doi.org/10.1093/rfs/hhw063

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. https://doi.org/10.1111/j.1540-6261.2009.01501.x

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock-market participation. *The Journal of Finance*, *59*(1), 137–163. https://doi.org/10.1111/j.1540-6261.2004.00629.x

Hong, H., Kubik, J. D., & Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, *60*(6), 2801–2824. https://doi.org/10.1111/j.1540-6261.2005.00817.x

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, *21*(2), 109–128. https://doi.org/10.1257/jep.21.2.109

Hu, Y., Li, X., Goodell, J. W., & Shen, D. (2021). Investor attention shocks and stock co-movement: Substitution or reinforcement? *International Review of Financial Analysis*, *73*, 101617. https://doi.org/10.1016/j.irfa.2020.101617

Huang, S., Huang, Y., & Lin, T.-C. (2019). Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics*, *132*(2), 369–383. https://doi.org/10.1016/j.jfineco.2018.10.006

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225.

Hvide, H. K., & Östberg, P. (2015). Social interaction at work. *Journal of Financial Economics*, *117*(3), 628–652. https://doi.org/10.1016/j.jfineco.2015.06.004

Ivković, Z., & Weisbenner, S. (2007). Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *The Review of Financial Studies*, *20*(4), 1327–1357. https://doi.org/10.1093/revfin/hhm009

Jain, P. C., & Joh, G.-H. (1988). The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, *23*(3), 269–283. https://doi.org/10.2307/2331067

Jiang, L., Liu, J., & Yang, B. (2019). Communication and comovement: Evidence from online stock forums. *Financial Management*, *48*(3), 805–847. https://doi.org/10.1111/fima.12245

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

Kantorovitch, I., & Heineken, J. (2021). Does dispersed sentiment drive returns, turnover, and volatility for Bitcoin? (SSRN Scholarly Paper 3920987). https://doi.org/10.2139/ssrn.3920987

Kaustia, M., & Knüpfer, S. (2012). Peer performance and stock market entry. *Journal of Financial Economics*, *104*(2), 321–338. https://doi.org/10.1016/j.jfineco.2011.01.010

Kim, D.-Y., & Ahn, Y. (2023). Emotional reactions, sentiment disagreement, and Bitcoin trading. *Asia-Pacific Journal of Business*, *14*(4), 37–48. https://doi.org/10.32599/apjb.14.4.202312.37

Kiymaz, H., & Berument, H. (2003). The day of the week effect on stock market volatility and volume: International evidence. *Review of Financial Economics*, *12*(4), 363–380. https://doi.org/10.1016/S1058-3300(03)00038-7

Knittel, M., Pitts, S., & Wash, R. (2019). "The most trustworthy coin": How ideological tensions drive trust in Bitcoin. *Proc. ACM Hum.-Comput. Interact.*, *3*(CSCW), 36:1-36:23. https://doi.org/10.1145/3359138

Kuchler, T., Li, Y., Peng, L., Stroebel, J., & Zhou, D. (2022). Social proximity to capital: Implications for investors and firms. *The Review of Financial Studies*, *35*(6), 2743–2789. https://doi.org/10.1093/rfs/hhab111

Lee, C., Shin, J., & Hong, A. (2018). Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. *Telematics and Informatics*, *35*(1), 245–254. https://doi.org/10.1016/j.tele.2017.11.005

Leung, A. C. M., Agarwal, A., Konana, P., & Kumar, A. (2017). Network analysis of search dynamics: The case of stock habitats. *Management Science*, *63*(8), 2667–2687. https://doi.org/10.1287/mnsc.2016.2470

Li, D., & Li, G. (2021). Whose disagreement matters? Household belief dispersion and stock trading volume. *Review of Finance*, *25*(6), 1859–1900. https://doi.org/10.1093/rof/rfab005

Li, T., van Dalen, J., & van Rees, P. J. (2018). More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of Information Technology*, *33*(1), 50–69. https://doi.org/10.1057/s41265-016-0034-2

Lin, K.-Y., & Lu, H.-P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, *27*(3), 1152–1161. https://doi.org/10.1016/j.chb.2010.12.009

Loria, S. (2020). Textblob documentation. *Release 0.16.0*. https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf

Maturana, G., & Nickerson, J. (2019). Teachers teaching teachers: The role of workplace peer effects in financial decisions. *The Review of Financial Studies*, *32*(10), 3920–3957. https://doi.org/10.1093/rfs/hhy136

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, *42*(3), 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

Mirc, N., & Parker, A. (2020). If you do not know who knows what: Advice seeking under changing conditions of uncertainty after an acquisition. *Social Networks*, *61*, 53–66. https://doi.org/10.1016/j.socnet.2019.08.006

Morris, M. R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why? A survey study of status message Q&A behavior. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1739–1748. https://doi.org/10.1145/1753326.1753587

Ouimet, P., & Tate, G. (2020). Learning from coworkers: Peer effects on individual investment decisions. *The Journal of Finance*, *75*(1), 133–172. https://doi.org/10.1111/jofi.12830

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, *53*(1), 61–65. https://doi.org/10.1086/296071

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Peng, L., Xiong, W., & Bollerslev, T. (2007). Investor attention and time-varying comovements. *European Financial Management*, *13*(3), 394–422. https://doi.org/10.1111/j.1468-036X.2007.00366.x

Pool, V. K., Stoffman, N., & Yonker, S. E. (2015). The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance*, *70*(6), 2679–2732. https://doi.org/10.1111/jofi.12208

Rantala, V. (2019). How do investment ideas spread through social interaction? Evidence from a Ponzi scheme. *The Journal of Finance*, *74*(5), 2349–2389. https://doi.org/10.1111/jofi.12822

Shalen, C. T. (1993). Volume, volatility, and the dispersion of beliefs. *The Review of Financial Studies*, *6*(2), 405–434. https://doi.org/10.1093/rfs/6.2.405

Shiller, R. J., Fischer, S., & Friedman, B. M. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, *1984*(2), 457–510. https://doi.org/10.2307/2534436

Shiller, R. J., & Pound, J. (1989). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, *12*(1), 47–66. https://doi.org/10.1016/0167-2681(89)90076-0

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking & Finance*, *78*, 130–141. https://doi.org/10.1016/j.jbankfin.2017.02.005

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Su, F., & Wang, X. (2021). Investor co-attention and stock return co-movement: Evidence from China's A-share stock market. *The North American Journal of Economics and Finance*, *58*, 101548. https://doi.org/10.1016/j.najef.2021.101548

Su, M.-H., Suk, J., & Rojas, H. (2022). Social media expression, political extremity, and reduced network interaction: An imagined audience approach. *Social Media + Society*, *8*(1), 20563051211069056. https://doi.org/10.1177/20563051211069056

Tan, S. D., & Tas, O. (2021). Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance*, *22*(2), 221–234. https://doi.org/10.1080/15427560.2020.1772261

Tantaopas, P., Padungsaksawasdi, C., & Treepongkaruna, S. (2016). Attention effect via internet search intensity in Asia-Pacific stock markets. *Pacific-Basin Finance Journal*, *38*, 107–124. https://doi.org/10.1016/j.pacfin.2016.03.008

Vlahavas, G., & Vakali, A. (2024). Dynamics between Bitcoin market trends and social media activity. *FinTech*, *3*(3), Article 3. https://doi.org/10.3390/fintech3030020

Wang, J., Wu, K., Pan, J., & Jiang, Y. (2023). Disagreement, speculation, and the idiosyncratic volatility. *Journal of Empirical Finance*, *72*, 232–250. https://doi.org/10.1016/j.jempfin.2023.03.011

Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal*, *16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041

Yang, S. Y., Mo, S. Y. K., & Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, *15*(10), 1637–1656. https://doi.org/10.1080/14697688.2015.1071078

Zhang, A. C., Fang, J., Jacobsen, B., & Marshall, B. R. (2018). Peer effects, personal characteristics and asset allocation. *Journal of Banking & Finance*, *90*, 76–95. https://doi.org/10.1016/j.jbankfin.2018.03.001

Zhaunerchyk, K., Haghighi, A., & Oliver, B. (2020). Distraction effects on stock return co-movements: Confirmation from the Shenzhen and Shanghai stock markets. *Pacific-Basin Finance Journal*, *61*, 101301. https://doi.org/10.1016/j.pacfin.2020.101301