THREE ESSAYS ON BEHAVIORAL FINANCE

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Ye Xian
December 2023

Three Essays on Behavioral Finance

In this disseration, I study several topics in the field of behavioral finance. Chapter 1 empirically explores the divergent sentiment shift of partisan investors after shift in political power, by using the 2020 presidential election as an event study. With data from social media platform StockTwits, the empirical findings indicate that after the election, Republican (Democrat) investors become more pessimistic (optimistic) toward future stock returns. These partisan divergent belief shifts are more pronounced when the election results become more solidified and are more concentrated in major stock indices (SPY, QQQ, DIA) and stocks with higher market beta. Additionally, consistent with the theoretical framework of Kruger (2020), further analysis indicates that during the post-election period, partisan disagreement is associated with increased stock liquidity and intraday volatility. These results indicate that partisan investors are willing to trade against those with opposite beliefs on the financial market during the post-election period.

In Chapter 2, I empirically examine the impact of local Covid spread on the net flows of locally headquartered mutual funds. The empirical findings indicate that state-level Covid spread reduces the net flows of locally headquartered mutual funds, which are more pronounced in the Covid crash period. Further analysis indicates that the reducing effect of Covid on fund net flows is more pronounced for retail fund share class, suggesting a heterogeneous response to the pandemic across investor types. Additional analysis shows that Covid-induced fund outflows are more pronounced for funds associated with higher levels of risk, implying that heightened risk aversion during the pandemic is a major driving factor behind the baseline results. Controlling local economic conditions does not significantly alter the main findings, indicating that visceral response is a more plausible explanation than economic shock. Alternative measures using state-level Google search volumes corroborate the main findings.

In Chapter 3, I empirically examine the impact of social media sentiment and attention on IPO pricing. By using social media sentiment from Stocktwits as a proxy for retail valuation, I empirically examine the theoretical predictions in prior studies (Ljungqvist et al. (2006), Cornelli, Goldreich, and Ljungqvist (2006), and Derrien (2005)) that over-optimism of sentiment investors leads to initial overpricing of IPO followed by long-term reversals. Using posts on Stocktwits during the pre-IPO period, I construct investor attention and sentiment measurements. The empirical results are generally consistent with the theoretical predictions that retail investor over-optimism leads to higher IPO first-day price run-up and worse long-term performance. Additionally, using machine learning techniques to classify untagged posts, I find similar results where sentiment measures are constructed with classified untagged posts. Results with sentiment measures constructed by classified untagged posts imply that more optimistic sentiment leads to a higher turnover rate shortly after IPO, implying that informed investors are selling overpriced IPO shares to sentiment retail investors.

BIOGRAPHICAL SKETCH

Ye Xian was born and grew up in Tianjin, China. He spent 18 years in Tianjin, before going to the United States for undergraduate study. He graduated from Brandeis University in 2012 with bachelor's degree in economics and mathematics, with honors. In August 2017, he began his doctoral study at Cornell University, in the field of Applied Economics and Management, with focus on financial economics, particularly behaviroal finance. After graduation, he will join Acadian Asset Management as a quantitative research analyst in Boston, MA.

To my parents

Acknowledgements

I have witnessed extraordinary growth during my six years at Cornell, not only from an intellectual and academic perspective but also from a personal perspective. I would like to reflect on the people who supported me during my Ph.D. study at Cornell.

First, I am deeply indebted to my thesis advisors, Will Lin Cong, Byoung-Hyoun Hwang, Yongmiao Hong, and Shanjun Li, for their support, guidance, and kindness during my Ph.D. studies. My advisors not only inspired me to become a better researcher in the field of economics but also taught me valuable lessons in life. During the course of my Ph.D. studies, I have benefited from their comprehensive knowledge in various fields of economics, as well as their professional and emotional support. Without their help, it would be impossible for me to finish this dissertation.

Additionally, I would like to thank David Ng, Calum Turvey, Arnab Basu, Jawad Addoum, Justin Murfin, Scott Yonker, and Gideon Saar for their guidance and suggestions during my studies at Cornell. I also would like to express my gratitude to my friends at Cornell Hui Zhou, Siguang Li, Yimeng Tang, Haokun Sun, Fan Zhou, Jintao Gu, and many others for their help and company along the way. I am also grateful for the financial support from Dyson School. I would like to thank Penny Sanders for her professional support during the job market process. Last but not least, I would like to express my gratitude to my parents, Guoming Xian and Hong Ding, for their selfless love, support, and encouragement during the course of my Ph.D. study. I would not have made it this far without their support and encouragement.

Contents

List of Tables

List of Figures

1 Partisan Disagreement and Stock Liquidity: Evidence from Social Media Activities around the 2020 Presidential Election

1.1 Introduction

The United States is currently experiencing an unprecedented level of political polarization, which has led to significant divisions between the two major political parties and their respective ideologies. These divisions are now deeply entrenched in the opposing political camps, with each side increasingly resistant to compromise or find common ground. During the COVID-19 pandemic, partisan affiliation became a critical predictor of attitude toward the risk of the virus, as well as policies regarding masking, social distancing, and vaccines (Allcott et al. (2020), Milosh et al. (2020), Barrios and Hochberg (2021)). In the wake of this political and ideological polarization, it has become increasingly important to explore how partisanship shapes individuals' economic beliefs and decisions.

Partisan differences influence beliefs on a wide range of issues, including trade policy, immigration, economic policy, gun control, and racial issues (Milner and Judkins (2004), Gaines et al. (2007), Bartels (2018), Burton et al. (2019)). Over the past thirty years, divergence in beliefs shaped by partisan affiliation has increased dramatically in the United States (Bishop (2008), Abramowitz and Saunders (2008), Kaplan, Spenkuch, and Sullivan (2019)). Recent studies have shown that partisan politics have influenced economic expectations, as agents with different partisan affiliations exhibit divergent economic expectations. Survey-based studies show that households become more optimistic about future economic conditions when the party they are more closely affiliated with control the White House (Bartels (2002), Evans and Anderson (2006), Coibion et al. (2020)). However, whether such divergence in belief translates into economic actions remains uncertain. Several studies have suggested that partisan divergence in economic perceptions does indeed lead to divergent economic behavior and activity (Gerber and Huber (2009, 2010), Gillitzer and Prasad (2018), Benhabib and Spiegel (2019)). However, this notion is challenged by some recent studies (McGrath (2017), Mian et al. (2017)). As such, the question of whether partisan disagreement over economic expectations translates into economic action remains unresolved.

Additionally, though partisan affiliation may influence economic expectations on a wide range of issues, such as trade or tax policy, as suggested by Cookson, Engelberg, and Mullins (2020), it is surprising to see that investors are forming stock expectations based on their partisan identity. There exists a strong incentive for investors to form correct beliefs regarding the future performance of firms, such as cash flow, when making financial decisions, regardless of their partisan affiliation. Even though investors may express divergent sentiments toward the financial market based on their partisan identity, it is even more surprising to see that such disagreement has an impact on investment decisions. In this scenario, it is evident that under the current political division in the United States, partisan identity imposes behavioral bias in the expectation formation process and subsequent investment decisions.

This study is motivated by the desire to empirically determine whether the shift in the political landscape induces divergent beliefs among partisan investors regarding the financial market and whether these beliefs have an impact on investment decisions. Social media platforms provide a useful setting for this study as they provide a wealth of user-generated content that can be analyzed to identify users' partisan affiliations and gauge their beliefs regarding the financial market. Additionally, data from social media posts can be used to construct belief measurements, which in turn can be used to determine whether partisan beliefs translate into investment decisions. Specifically, by combining social media data and stock-level data daily, I can investigate whether partisan beliefs have an effect on investment decisions by examining the association between divergent belief shifts and aggregate trading behavior at the daily stock level.

The social media data I use in this study comes from the social media platform Stocktwits.com, where users can tag their posts as "Bullish" or "Bearish" to explicitly express their sentiment toward given securities. More specifically, I collect posts that mention S&P500

stocks and major stock indices (SPY, QQQ, DIA) in the months around the 2020 presidential election (09/2020–04/2021). Using the sample of posts collected during the months around the 2020 presidential election, I attempt to identify the partisan affiliation of each user by employing a list of political keywords with strong partisan inclination. Users who made posts that contained Republican (Democrat) keywords are identified as Republican (Democrat) investors.

After identifying the partisan identity of users, I examine the shift in sentiment toward stock market performance at a stock-user-date level. Specifically, I examine the heterogenous belief shift in Republican and Democrat investors over the timelines of the 2020 election. After the election outcome was called on 11/06/2020, Republican presidential candidate President Donald Trump refused to admit the loss and made several efforts to overturn the election outcome. Therefore, during the period between 11/07/2020 and the final congressional certification of election results on 01/06/2021, the future state of a Democrat presidency remained uncertain, at least in the minds of many partisan investors. As such, I employ three events during the period to conduct the event study: the election outcome's call on 11/07/2020, the vote of the electoral college on 12/14/2020, and the eventual certification of the election result on 01/06/2021. The empirical results indicate that in the post-election period, Republican investors became more pessimistic toward the stock market, while Democrat investors did the opposite. The divergence in partisan sentiment then peaked after the eventual certification of the election result on 01/06/2021, suggesting that efforts by Republicans to challenge the election outcome introduced uncertainty about the future political climate, which in turn influenced partisan beliefs related to the stock market.

Next, I conduct further analysis of the baseline results. Specifically, I explore whether the peak of partisan divergence after the Jan.6 event was driven by partisan sentiment toward big tech stocks. After the Capitol riot on Jan.6, big tech companies prohibited or limited services related to President Trump and his political campaign. These actions may have resulted in a partisan sentiment shift toward big tech firms, which make up a significant portion of posts in the sample. Although the results indicate that the partisan sentiment divergence was significantly more pronounced for big tech firms after Jan.6, the results from the subsample excluding big tech firms remain statistically and economically significant. Therefore, partisan sentiment shifts toward big tech firms only partially drive the baseline results.

The shift in political climate may result in drastic changes in economic policies, which may be favored by investors more closely affiliated with the incoming presidency and opposed by those who are not. Therefore, I conduct heterogeneity tests to examine whether the divergent partisan sentiment shift comes from heterogenous expectations toward future overall economic conditions. Firstly, I show that the divergent partisan belief shifts are more concentrated in the subsample of major stock indices, which implies that the partisan belief shifts come more from expectations of the overall stock market performance. Furthermore, I categorize stocks into three quantiles based on their exposures to systematic risk, which are measured by the market beta in the months before the sample period. The results indicate that the divergent partisan sentiment shifts are significantly more pronounced in the quantiles with a higher market beta, which implies that the sentiment shift comes from the expectation of overall market condition. Overall, these results indicate that the partisan sentiment shift induced by changes in political climate can be explained by changes in investors' expectations of aggregate market conditions, which are likely to be affected by changes in significant economic policies.

After studying the influence of political climate shifts on partisan investors' beliefs, I further explore whether such partisan disagreement leads to investment action. Specifically, I empirically assess the association between partisan disagreement and stock liquidity across the cross-section of stocks. The results indicate that in the post-election period, higher daily stock-level partisan disagreement was associated with increased liquidity, which is measured as a higher abnormal turnover rate and lower effective percentage spread. Additionally, the empirical results indicate that during the post-election period, stock-level partisan disagreement was associated with higher intraday volatility, implying that partisan investors' divergent opinions were priced into stock prices.

The findings of this study have important implications. Firstly, the study contributes to the emerging studies on the impact of political polarization on the financial market (Dagostino et al. (2022), Meeuwis et al. (2022), Cassidy and Vorsatz (2021), Cookson, Engelberg, and Mullins (2020)). Overall, the empirical findings in this study indicate that the announcement of the 2020 presidential election outcome induced significant divergence in partisan investors' sentiments toward future financial market performance. Further empirical analyses imply that the divergence in partisan beliefs primarily stemmed from expectations of overall economic conditions, as the belief shifts were more pronounced for stock indices and stocks with higher exposure to market risk. Additionally, further empirical results indicate that the presidential election induced divergent partisan belief shifts and that such partisan disagreement leads to investment action. In the post-election period, higher partisan disagreement was associated with increased liquidity and volatility, which implies that partisan investors were trading against each other based on their divergent beliefs regarding asset prices. In the post-election period, higher partisan disagreement was associated with increased liquidity and volatility, suggesting that partisan investors were trading against each other based on their divergent beliefs about asset prices.

Another important implication of this study is that its findings connect the literature on politics and finance with the literature on disagreement and financial markets (Eyster et al. (2019), Kyle, Obizhaeva, and Wang (2018), Kruger (2020), Cookson and Niessner (2020)). Firstly, the empirical results indicate that a shift in the political climate was a key source of heterogeneous beliefs among investors, as the presidential election outcome induced divergent beliefs among investors with different political identities. Additionally, the results indicate that under the current political polarization, the private interpretation of public information leads to increased liquidity in the financial market, as partisan investors tend to overweight their signal in expectation formation.

Lastly, the empirical results and methodologies have important implications for future studies. Firstly, as the policies proposed by the Republican and Democratic Parties continue to become more polarized, the current political division in the United States seems unlikely to wane in the foreseeable future. Whether such political division translates into divergent economic beliefs and actions remains an interesting question to be answered in the future. Future studies can employ the method adopted in this study to explore the impact of future policy changes or political events on the beliefs or investment actions of partisan investors. Moreover, future research can determine the impact of policy changes or political events in relation to partisan issues such as climate change, trade policy, immigration and racial issues, and gun control on partisan investors' belief shifts and potential investment decisions. Additionally, this study indicates that messages on social media platforms can be used as a crucial source of data for identifying partisan investors and their sentiment shifts. Thus, future studies can employ social media data to measure the belief changes of partisan investors and to determine whether such belief shifts affect investment action.

The rest of this paper is organized as follows. In section 2, I discuss the existing literature related to this study. In section 3, I discuss the background of the empirical setting in the 2020 presidential election and the development of the hypothesis to be empirically examined. In section 4, I discuss the data source and the construction of the variables. In section 5, I discuss how the presidential election outcome induced divergent sentiment shifts among partisan investors. This is then used as a means by which to discuss the empirical specifications and baseline empirical results and to further explain the rationale behind the baseline empirical findings. Section 6 discusses the impact of partisan disagreement on the cross-section of stock liquidity and intraday volatility, including the empirical specification and results. In section 7, I summarize the paper and offer a conclusion.

1.2 Related Literature

This study makes valuable contributions to the literature on disagreement and its impact on the financial market. The primary role of disagreement in finance literature is to explain trading volume and turnover rate in a way that a standard rational expectation model with homogenous beliefs cannot achieve. Standard asset pricing models cannot explain the high turnover rate in the financial market. In the model of Grossman (1976), when the market perfectly aggregates the private information of traders, and there is no disagreement of asset valuation among traders, there is no incentive for trade. As such, disagreement provides an appealing rationale for trading and high turnover, which the standard asset pricing model cannot explain.

Hong and Stein (2007) summarize the theoretical and empirical findings of disagreement in the financial market. Disagreement models challenge the traditional homogenous beliefs of asset pricing models. One of the critical features of disagreement models is the proposition that differences in private information, or divergent private interpretations of public information, cause investors to disagree on the values of asset prices. There are several possible explanations for this heterogenous private information. The first possible explanation, introduced by Hong and Stein (1999), involves the gradual flow of information. In this scenario, certain value-relevant information about the stock reaches some investors before it reaches others, leading to disagreement among them. The second possible explanation is that due to limited attention, investors can only process a subset of the available information at any given time, which further leads to heterogenous beliefs among investors. In the model proposed by Hirshleifer and Teoh (2003), we observe this disagreement between attentive and inattentive investors, with inattentive investors failing to update their beliefs to match the prevailing market price.

The third possible explanation for disagreement is that investors employ different economic models and interpret public information differently due to heterogeneous priors. Harris and Raviv's (2015) model states that trading in a speculative market is based on different opinions formed by heterogeneous interpretations. Kandel and Pearson's (1995) model emphasizes the differential interpretation of public information to account for the increased abnormal trading volume observed around earnings announcements. Banerjee and Kremer's (2010) dynamic model highlights that investors disagree on interpreting public information, which helps explain the empirical evidence of the positive correlation between volatility and trading volume. Andrei, Carlin, and Hasler's (2019) model of incomplete information suggests that agents disagree about the business cycle length. Such disagreement generates return volatility, particularly during recessions. Finally, Boot, Gopalan, and Thakor's (2008) model posits that the different priors of managers and investors lead to disagreements regarding the undertaking of projects, which may cause governance challenges and force public companies to go private.

Another strand of literature focuses on the role of behavioral bias in the formation of disagreement and how such disagreement influences trading behavior. Eyster, Rabin, and Vayanos's (2019) model proposes that disagreement should be seen as the result of investors neglecting the informational content of prices about other investors' private information. Several other recent studies focus on the role of overconfidence in disagreement trading. Overconfidence makes investors overestimate the precision of their private information, which in turn makes them more willing to trade against those with different private information. For instance, Kelley and Tetlock's (2013) model states that disagreement generated by the overconfidence of uninformed traders plays a crucial role in explaining stock trading volume. Kyle, Obizhaeva, and Wang's (2018) theoretical framework proposes that overconfidence leads to more disagreement among traders. Kruger's (2020) model suggests that private information increases liquidity and trading in the presence of overconfidence while reducing liquidity and trading in the absence of overconfidence. The findings of this study indicate that partisan investors may overestimate their partisan belief in the financial market performance, ultimately leading to disagreement regarding stock returns following the presidential election outcome.

Another series of studies employs disagreement to explain trading volume and liquidity. The market microstructure literature has traditionally viewed private information as a source of illiquidity. In fear of adverse selection, traders may worry that they are being exploited by counterparties with different information and will therefore demand different prices. Glosten and Milgrom's (1985) model claims that private information leads to a higher effective percentage spread. However, some studies model disagreement on asset value as motivation for trade and, therefore, as a source of liquidity, particularly in the presence of overconfidence. Kyle, Obizhaeva, and Wang's (2018) model views disagreement as the critical motivation for trade for overconfident traders. Lastly, in the framework developed by Kruger (2020), overconfident traders are more willing to trade against those with different private information; therefore, disagreement enhances liquidity in the presence of overconfidence. The empirical findings of this study indicate that partisan disagreement tended to enhance liquidity in the stock market following the election outcome.

Existing studies have also empirically examined the impact of disagreement on asset prices, trading behavior, and liquidity. Ajinkya et al. (1991) find that trading volume is positively related to the dispersion in analyst forecasts. Diether et al. (2002) argue that stocks with higher dispersion in analyst forecasts earn lower returns than similar stocks. Verardo (2009) observes that momentum return is far more significant for the portfolios characterized by higher analyst forecast dispersion. Carlin, Longstaff, and Matoba (2014) find that disagreement is associated with higher expected returns, higher return volatility, and higher trading volume for mortgage-backed security. Another challenge is to identify the source of disagreement empirically. A group of studies attempt to empirically identify the source that may generate investor disagreement on asset returns and examine the impact of such disagreement on the financial market. Many of these studies focus on heterogeneous private information or heterogenous priors that generate differential interpretations of public information, such as experience, background, or investment philosophy. Bailey et al. (2016), for instance, find that differential exposure to friends' experience with housing prices can lead to disagreement regarding real estate investment expectations, which in turn influences one's decision to become a renter or house owner. Chang et al. (2019) reveal that investors living in linguistically diverse areas express more diverse opinions on stock message boards and trade more frequently, indicating that linguistic diversity is a source of differential interpretation of public information. Using social media data, Cookson and Niessner (2020) find that disagreement generated by differential investment philosophy (e.g., technical, fundamental) leads to higher abnormal stock turnover. Cookson, Engelberg, and Mullins (2020) observe that heterogenous opinions expressed by partisan investors on social media platforms led to abnormal turnovers after the onset of the COVID-19 pandemic. This study builds on the work of these studies by demonstrating that partisan differential interpretations of the presidential election outcomes in 2020 served as a source of disagreement, ultimately leading to enhanced liquidity following the election outcome.

This study also contributes to the literature on the relationship between politics and finance. One strand of literature considers the impact of political events and activities on the financial market. In the time series, Santa-Clara and Valkanov (2003) and Blinder and Watson (2016) find that stock performance is significantly better during Democratic presidencies than Republican presidencies, a phenomenon known as the presidential puzzle. Pastor and Veronesi (2020) attempt to explain the presidential puzzle from a time-varying risk aversion perspective. Belo, Gala, and Li (2013) find that firms with high government spending exposure have higher cash flow and stock return during Democratic presidencies, while the opposite is the case during Republican presidencies. Addoum et al. (2019) note that political sensitivity explains a significant portion of momentum profit, particularly around the time of presidential elections. Addoum and Kumar (2016) reveal that political sentiment caused by shifts in political climate generates significant demand-based predictability of stock return. Cooper et al. (2010) find that firm-level contribution to political campaigns is positive and significantly correlated with the cross-section of stock return. Kim et al. (2012) observe that firms more geographically aligned to the presidential parties earn higher raw and risk-adjusted returns. Brown and Huang (2020) establish that corporate executive meetings with White House officials are associated with higher abnormal returns for firms. Chen et al. (2023) find that firms with low exposure to the presidential economic approval rating index earn higher returns than those with high exposure. Hassan et al. (2019) observe that firm-level exposure to political risk is associated with reduced hiring and investment and increased political donation. Using a machine learning-based textual measurement of firm-level political risk, Gorbatikov et al. (2023) reveal that firms with higher political risk exposure earn higher returns. Liu, Shu, and Wei (2017) note that politically sensitive Chinese firms experienced a significant stock price fall after the Bo Xilai scandal in 2012. The findings of this study add to this strand of literature by demonstrating that partisan disagreement drives liquidity and asset prices around the time of presidential elections.

Another study in this area (Ke, 2019) focuses on how partisanship shapes investor belief and financial decisions, revealing that Democrats are less likely than Republicans to participate in the stock market, particularly during Democratic presidencies. Many other studies focus on the influence of partisan belief on investor behavior and asset prices. Dagostino, Gao, and Ma (2022), for example, find that partisan bankers whose party is not affiliated with the US president tend to charge higher loan spreads and that partisan bankers also tend to lead syndicates with co-partisan bankers. Kempf and Tsoutsoura (2021) observe that partisan investors whose party is not affiliated with the US President tend to adjust corporate bond rating downward more frequently, resulting in a significant divergence in rating action between Democrat and Republican analysts after the 2016 election. Engelberg et al. (2022) find that partisan entrepreneurs increase their entrepreneurship when their party is affiliated with the US president but do the opposite when their party is not affiliated with the US president. Meeuwis et al. (2022) observe that after the 2016 election, Republican households increased the equity share and market beta of their portfolios, while Democrat households did the opposite. Cassidy and Vorsatz (2021) claim that Republican mutual fund teams purchased more equity from high-market beta industries around the 2016 election, while Democrat mutual fund teams did the same around the 2020 election. Cookson, Engelberg, and Mullins (2020) find that Republican investors became more optimistic than non-Republican investors after the onset of the COVID-19 pandemic. This study adds to this strand of literature by showing that the 2020 election outcome generated sharp partisan disagreement regarding the future performance of the stock market, particularly for stocks with high market beta.

Lastly, this study adds to the literature that connects online platforms and the financial market. One string of studies examines how firms employ social media platforms to interact with investors. Blankespoor, Miller, and White (2014) show that firms use Twitter to disclose key information among market participants more widely. Jung, Naughton, Tahoun, and Wang (2018) argue that firms strategically disseminate information on Twitter in order to shape their informational environment. Lee, Hutton, and Shu (2015) find that firms employ corporate social media to communicate with investors in an effort to attenuate the adverse price impact of product recall announcements. Finally, Chen et al. (2023) observe that corporate executives' adoption of social media can either improve or harm stock prices depending on the frequency of social media usage.

Another branch of literature uses messages from online platforms to measure investor opinions and then assesses their impact on the financial market. For instance, by conducting textual analyses of SeekingAlpha articles and comments, Chen et al. (2014) find that investor opinions significantly predict future stock return and earnings surprises. Chen and Hwang (2022) find that investors tend to share online articles for impression-management purposes despite the shared articles' quality, which leads to the overpricing of stock prices. Cookson, Niessner, and Schiller (2023) show that social media sentiment negatively predicts the probability of corporate merger withdrawals. Cookson, Engelberg, and Mullins (2022) produce evidence of selective exposure to confirmatory information on social media, which leads to greater trading volume and lower ex-post returns. Cookson et al. (2023) find that, following the run on Silicon Valley Bank, banks with more social media exposure tend to experience more stock value loss and greater outflow of uninsured deposits. This study contributes further to the second strand of studies by showing that partisan investor disagreement on social media resulted in increased liquidity and volatility after the 2020 presidential election outcome.

1.3 Background and Hypothesis Development

1.3.1 Background: 2020 Election and Politicization of the Stock Market

During his term as president, President Trump strengthened the connection between politics and stock market performance. He often cited the rise of the stock market during his presidential term as one of his most significant accomplishments. According to Trump's Twitter archive, from the launch of his presidency in January 2017 until the election on November 3rd, 2020, President Trump tweeted about the stock market 166 times. During the 2020 campaign season, he consistently highlighted the stock market's performance and warned about the potential of a stock market crash if his opponent, Joe Biden, were to win the election. Moreover, during his campaign rally and presidential debates, Trump frequently pointed to America's 401(K) accounts and investment portfolios and continued to emphasize that the stock market would crash if Joe Biden took the presidency. Examples of his stock market-related tweets can be found in Table 1.A.1 of the Appendix.

It has been shown that the politicization of the stock market can influence partisan beliefs regarding stock market performance and, in turn, impact the investment decisions of partisan investors. For instance, a study by Meeuwis et al. (2019) revealed that after the 2016 election, investors from the pro-Republican zip code tended to invest more aggressively, while investors from the pro-Democrat zip code tended to do the opposite. Compared to the 2016 election, the politicization of the stock market may have been even more aggravated in the 2020 election thanks to President Trump making the issue of stock market performance central to his reelection bid. For example, reflecting Trump's repeated warnings on Twitter about a potential stock market crash if Joe Biden were to be elected, Figure 1.1 shows that the Google search index during the 2020 election witnessed a surge in searching volume for 'Stock Market Crash' after the congressional certification of Biden's win on 01/06/2021. This implies that Trump's warning might have triggered concerns among partisan investors regarding the possibility of a stock market crash.

Another factor to consider is that, during the 2020 election, the US economy was hit by the COVID-19 pandemic, which sparked opposing views and policies from Republicans and Democrats. As COVID-related policies were central to the subsequent economic recovery, the shift in political power may have changed partisan investors' expectations of overall economic conditions and financial market performance. On this point, a survey-based study by Coibion et al. (2020) shows that before the 2020 election, Democrats and Republicans firmly believed that their respective presidential candidates would win the election. Additionally, the survey indicates that partisan households predicted improved economic conditions if their preferred candidate were to win the election and negative economic conditions if their opponent were to win. Given these biases, the 2020 election had the potential to induce significant disagreement among partisan investors regarding future stock returns.

Another feature of the 2020 presidential election was the fact that, due to the influence of the COVID-19 pandemic, a significant portion of votes were cast using mail-in ballots. President Trump seized upon this as an opportunity to claim that there could be widespread voter fraud through mail-in ballots. As such, even after the election results were called, Trump and his campaign refused to admit the loss and further claimed that the election results in major swing states were invalid due to massive voter fraud. The Trump campaign then attempted to file several lawsuits to overturn the election results, all of which proved unsuccessful. These efforts to overturn the election outcome eventually led to the Capitol riot on 01/06/2021, after which the election outcome of Biden's win was eventually certified by Congress, and the presidency was transferred peacefully.

During the period between the call and congressional certification of the election result, the Trump campaign's effort to overturn the election might have stirred up a feeling of uncertainty regarding the future political landscape. This uncertainty likely influenced the beliefs of some partisan investors, particularly as the prospect of a Democratic presidency became more concrete. Hence, one interesting aspect to be examined is whether partisan investors' beliefs regarding future stock returns changed as the future state of the Democrat presidency became more solidified. To address this question, I employ three events during the post-election period as an event study: the call of the election outcome on 11/06/2020, the vote cast by the electoral college on 12/14/2020, and the congressional certification of the election outcome on 01/06/2021.

1.3.2 Stocktwits.com

Stocktwits.com is an investment-oriented social media platform that facilitates up-to-date idea sharing between investors. The website introduced a unique feature, a hashtag combined with a stock ticker symbol, to enable users to hold focused discussions regarding specific companies. As shown in Figure 1.A.1, one can use this feature to search for and view all recent posts related to Alphabet Inc. by typing in its ticker symbol "GOOG" into the search bar.

Another critical feature of Stocktwits.com is that it enables users to tag their posts as "Bullish" or "Beariish", or, alternatively, to use no tag at all. In Figure 1.A.1, we can see a post with a "Bullish" tag, "Bearish" tag, and no tag. As shown in Figure 1.A.2, users can set these tags by clicking the buttons located in the bottom-right of the screen. Given that information on social media platforms contains a vast amount of noise, this feature greatly enhances the platform's usability by enabling researchers to identify bullish or bearish sentiments more easily.

Launched in 2008, this website has become one of the most popular and fast-growing online platforms for investors. Remarkably, the active user base has grown from around 230,000 in 2012 to around 3 million in 2020. In this study, I assume that opinions shared on Stocktwits.com are representative of overall investor opinions.

1.4 Hypothesis Development

The 2020 presidential election was characterized by intense partisan polarization, particularly in regard to the election outcome and subsequent economic conditions. As such, I hypothesize that the outcome of the 2020 presidential election induced significant divergent sentiment shifts among Republican and Democrat investors. Republican (Democrat) investors became more pessimistic (optimistic) regarding future economic conditions in the event of a Democratic presidency and therefore held a more bearish (bullish) view on stock market returns. I also hypothesize that after the outcome of the 2020 election, there was an increase (decrease) in the fraction of posts with "Bearish" ("Bullish") tags for Republican investors and vice versa for Democrat investors.

Additionally, the Republican effort to overturn the election outcome might have introduced a sense of uncertainty regarding the future state of the Democrat presidency. However, as the lawsuits filed by the Trump campaign ultimately failed and the election procedures further certified Biden's win through the electoral college vote and final congressional certification of the election outcome, the future state of the Democrat presidency became more solidified. As such, I hypothesize that the divergence of partisan sentiment shift became incrementally more pronounced from the call of the election outcome on 11/06/2020 to the congressional certification on 01/06/2021.

Moreover, according to the theoretical framework of Kruger (2020), in the presence of overconfidence, investors with private information are more willing to trade against those with different private information. Therefore, disagreement leads to increased trading volume and liquidity. In the current climate of intense political polarization, it is likely that partisan investors overestimate their private interpretation of public signals and ignore signals from the equity prices and private information of those with different opinions. As such, overconfident private interpretation of public signals (election outcome) is likely shared among partisan investors. Thus, using Kruger's (2020) theoretical framework, I hypothesize that in the post-election period, stock-level partisan disagreement was associated with increased liquidity, as measured by a higher turnover rate and lower effective percentage spread. Moreover, as divergent private information is incorporated into the trade of partisan investors, I further hypothesize that stock-level partisan disagreement was associated with increased intraday volatility in the post-election period.

1.5 Data and Summary Statistics

1.5.1 Data and Variable Construction

The sample for this study consists of stock in the S&P500 index composite listed from 08/01/2020–03/31/2021. The daily stock-level data for individual stocks is collected from CRSP. I collect data on the daily closing price, trading volume, and number of outstanding shares for each stock. I further construct daily stock return and turnover rate. Additionally, from the CRSP beta suite, I collect data on stock market beta and abnormal returns based on a rolling regression of the Fama-French three-factor model.

The data on social media posts are collected from Stocktwits.com. For each stock in the composite of S&P500, I collect posts on Stocktwits.com that contain its ticker symbol from 08/01/2020 to 03/31/2021. For each post in the sample, I collect information on the post's content, date of posting, user ID, and sentiment. For each stock, user, and date combination, the bullish and bearish sentiment is measured as the percentage of posts tagged as "Bullish" and "Bearish," respectively. Moreover, the overall sentiment is measured as the difference between the ratio of bullish and bearish posts in the total number of posts.

1.6 Identifying Partisan Investors on Stocktwits

In addition to the self-classified sentiment toward stock investment, users on the StockTwits platform often discuss topics other than investment, such as politics (Cookson et al. (2020) and climate change (Santi (2023)). The discussion of these topics can be extracted from the content of posts on the platform and can be used to identify the partisan inclinations of users. Following Cookson et al. (2020) and Gentzkow and Shapiro (2010), I identify a list of politic-related words that can help identify the potential partisanship of the user.

This list of words, as presented in Table 1.1, builds upon the word list used in Cookson et al. (2020) by adding additional politic-related partisan words around the time of the 2020 election. The list contains distinctively partisan Republican or Democratic language. Panel (a) reports the word list for partisan Republicans, many of which express support for President Trump's reelection in 2020, such as #MAGA or #Trump 2024, or express resentment toward Democrats, such as #Sleepy Joe or #Leftist. Panel (b), meanwhile, reports the word list for partisan Democrats, which contains language that expresses support for Democrats in the 2020 presidential election or demonstrates resentment toward Trump or Republicans. Lastly, Panel (c) reports examples of politic-related posts that contain partisan language. To identify partisan investors, a user who made a post that contains distinctively Republican (Democratic) language is identified as Republican or Democrat. To remove confounding users, users who make posts that contain partisan language toward both parties are not identified as partisan investors. Moreover, the sample of securities in this study only contains S&P 500 stocks and major indexes. Therefore, the sample of users may not be as comprehensive as the sample employed by Cookson, Engelberg, and Mullins (2020). Therefore, I adopt a less conservative method rather than identifying user partisanship in the pre-event period, as in Cookson, Engelberg, and Mullins (2020). By assuming that the partisan inclinations of users did not significantly change before and after the election outcome, I classify users into Democrat or Republican based on their posts in the whole sample.

Table 1.2 reports a summary of the raw data collected from StockTwits posts, along with the classification outcomes for the sampled posts on Stocktwits.com. There are 4,841,244 posts in the sample, among which 1,579,343 are tagged as "Bullish" and 464,467 are tagged as "Bearish." Consistent with the findings in the previous study (Cookson and Nessner (2020)), investors appeared to have a behavioral bias toward broadcasting positive rather than negative information on social platforms. In the partisan identification process, 13,403 posts are identified as containing Republican keywords, while 2,195 are identified as containing Democrat keywords. Among the 135,404 distinctive users identified in the sample, 4,453 are identified as Republican users, 1,002 are identified as Democrat users, and 134,872 are not identified as partisan users. Among the total number of 4,841,244 posts in the sample, 964,935 are identified as messages from Republican users, and 240,563 are identified as messages from Democrat users. The summary statistics are consistent with the findings of Cookson, Engelberg, and Mullins (2020) that there are many more Republican users than Democrat users on Stocktwits.com. Additionally, the summary statistics indicate that partisan investors are more active than non-partisan investors. While Republican users comprise 3.28% of total distinctive users in the sample, they are responsible for 19% of the total messages in the sample. Democrat users, meanwhile, account for 0.74% of the total distinctive users and are responsible for 4.9% of total messages.

1.6.1 Construction of Sentiment Measures

To measure the sentiment at the user-stock-date level, I construct a measurement of bullish, bearish, and aggregate sentiment for user *i* toward security *s* at date *t*, based on the tagged posts in the sample. The bullish (bearish) sentiment is represented by the ratio of posts tagged as "Bullish" ("Bearish") in the total number of posts made by user *i* regarding security *s* at date *t*. The aggregate sentiment is measured as bullish sentiment minus bearish sentiment for each user security-date combination. The construction of sentiment measures can be illustrated as

$$Bullish_{i,s,t} = \frac{Number\_of\_Bullish\_Posts_{i,s,t}}{Number\_of\_Posts_{i,s,t}} \tag{1}$$

$$Bearish_{i,s,t} = \frac{Number\_of\_Bearish\_Posts_{i,s,t}}{Number\_of\_Posts_{i,s,t}} \tag{2}$$

$$Sentiment_{i,s,t} = Bullish_{i,s,t} - Bearish_{i,s,t} \tag{3}$$

Table 1.4 presents the summary statistics of sentiment measures. In the full sample, Republican investors tended to be more optimistic toward the stock market than Democrat investors but less optimistic than non-partisan investors. The same pattern held during the subsample of the pre-event period before 11/07/2020, where Republican investors had an average sentiment of 0.1818, Democrat investors had an average sentiment of 0.1083, and non-partisan investors had an average sentiment of 0.2765. In the two post-event periods, 11/07/2020–12/14/2020 and 12/15/2020–01/05/2021, Democrat investors, on average, tended to be more optimistic about the financial market, while Republican and non-partisan investors did not significantly change their opinions. However, after the Jan.6 event, the statistics show that Republican investors, on average, experienced a significant shift toward more pessimistic sentiment toward the financial market, while there was no significant shift in the sentiment of Democrat and non-partisan investors. Overall, then, the summary statistics indicate a significant shift in the sentiment of partisan investors toward the stock market during the 2020 election.

1.7 Evidence of Shift in Beliefs of Partisan Investors

1.7.1 Empirical Specification

In this section, I conduct a regression analysis in order to investigate the shift in beliefs among partisan investors in regard to the stock market during the 2020 presidential election. I specifically focus on the months before and after the 2020 presidential election in November. Using the user-security date sentiment measurement as the dependent variable, I estimate the following regression model in order to track the changes in partisan investor sentiment in the post-election period:

$$\begin{aligned}
Belief_{i,s,t} &= \beta_0 + \beta_1 \times GOP_i + \beta_2 \times Dem_i + \\
\sum_{n=0}^{2} \theta_n \times After_{n,t} &+ \sum_{n=0}^{2} \gamma_n \times After_{n,t} \times GOP_i \\
&+ \sum_{n=0}^{2} \phi_n \times After_{n,t} \times Dem_i + \alpha_s \\
&+ GOP_i \times \alpha_s + Dem_i \times \alpha_s + \epsilon_{i,s,t}
\end{aligned} \tag{4}$$

The dependent variable $Belief_{i,s,t}$ measures the belief of user i toward security s at date t. The dependent variable can be the $Bullish_{i,s,t}$, $Bearish_{i,s,t}$ and $Sentiment_{i,s,t}$, which respectively measure the bullish, bearish, and aggregate belief of the user i toward security s at date t. Additionally, GOP_i is a dummy variable that indicates whether the user is classified as a Republican investor, while Dem_i is a dummy variable that indicates whether the user is classified as a Democrat investor based on the classification rule mentioned previously. The time dummy variables $After_{0,t}$, $After_{1,t}$, and $After_{2,t}$ indicate the three post-election periods, 11/07/2020–12/14/2020, 12/15/2020–01/05/2021, and post 01/06/2021, respectively. The key independent variables are the interaction terms between the three time dummy variables and the partisanship indicators GOP_i and Dem_i. The coefficients on the interaction terms measure the shift in partisan investors' beliefs regarding the return of securities in the post-election period.

Additionally, to control for security-invariant unobservables that may affect investor sentiment toward a given security, I include α_s as the security dummy for security fixed effects. To further absorb the average sentiment of partisan investors toward certain securities, I include interaction terms between partisanship indicators and security dummies for partisanship security fixed effects. $\epsilon_{i,s,t}$refers to the error term. I double-cluster the standard error by security and date to control for serially correlated error terms within the security-date level. This specification is estimated using sample data from 08/01/2020 to 03/31/2021.

Based on my hypothesis, in the post-election period, Republican (Democrat) investors tended to be more pessimistic (optimistic) toward the stock market. Therefore, the coefficients γ_n (ϕ_n) are expected to be negative (positive) when the dependent variable is $Bullish_{i,s,t}$ or $Sentiment_{i,s,t}$ and positive (negative) when the dependent variable is $Bearish_{i,s,t}$.

1.7.2 Empirical Results

In this section, I present the empirical results estimated using the model specified in Equation (4). Table 1.5 presents the baseline empirical results. Columns (1)–(3) present the empirical results for the dependent variables bullish, bearish, and aggregate sentiment, respectively. The coefficients on the partisanship indicators indicate that before the election outcome, Democrat investors tended to be more pessimistic than non-Democrat investors, while there was no significant difference in the aggregate sentiment of Republican and non-Republican investors despite their lower bearish sentiment levels. Consistent with my hypothesis, coefficients on the interaction term between post-election time dummies and Republican (Democrat) indicator are negative (positive) and statistically significant when the dependent variable is bullish or aggregate sentiment, and positive (negative) and statistically significant when the dependent variable is bearish sentiment. Overall, the empirical results indicate the emergence of partisan disagreement regarding the future returns of the financial market, with Republican investors experiencing an increase (decrease) in bullish (bearish) sentiment and Democrat investors experiencing the opposite.

The empirical results also indicate a time-varying response in investors' belief shifts during the 2020 election. Although the interaction terms between partisan indicators and post-election time indicators all have the expected sign of coefficients, the magnitude of the coefficients differ across the post-election time indicators. Specifically, for both Republican and Democrat partisan indicators, the magnitude of the coefficients on the interaction term is most pronounced for the post-Jan.6 event time dummy, relatively less pronounced for the post 11/06/2021 time dummy, and least pronounced for the post 12/15/2020 time dummy.

These results indicate that investors reacted differently to the events during the 2020 election. Notably, although the calling of the election outcome on 11/07/2020 induced a significant shift in the beliefs of partisan investors regarding the stock market, the Jan.6 event—after which the election result was certified by Congress and all the Republican efforts to overturn the election had clearly failed—induced the most pronounced impact on partisan investors' beliefs.

This differential reaction to various events during the course of the election may reflect investors' belief in the possibility of the election being overturned. After the election result was called on 11/06, the Republican candidate President Donald Trump and his associates spread the claim that widespread voter fraud had occurred in key battleground states and attempted to overturn the election result by filing several lawsuits to dispute the election result in these states. However, since all these lawsuits failed, the Republicans then attempted to dispute and invalidate the election outcome in the congressional certification process on 01/06/2021, which eventually resulted in the Jan.6 Capitol riot.

After the election results were called on 11/06, Republican investors may have only partially incorporated the election result into their beliefs regarding the future political landscape and financial market performance. Consequently, until the final certification of the presidential election results on 01/06/2021, the shift in sentiment among Republican investors towards pessimism was tempered by the lingering belief that the election outcome could be overturned. However, once the election results were officially certified by Congress after the events of January 6th, Republican investors fully incorporated the prospect of a Democrat president into their outlook on future financial market performance. Interestingly, the empirical results indicate that Democrat investors may have also incorporated the possibility of the election results being overturned into their belief formation. The magnitude of coefficients on the interaction term with the Democrat indicator also varies across the three post-election time dummies and is most pronounced for the post Jan.6 time dummy.

This change in belief regarding the overturning of the election result may explain the differential magnitude of sentiment shift throughout the election outcome, especially the most pronounced sentiment shift after the Jan.6 event. Overall, the empirical results indicate that although the calling of the election outcome on 11/06 induced a partial shift of partisan investors' beliefs regarding future financial market returns, the possibility of overturning the election attenuated this shift until the eventual certification after the Jan.6 event.

1.7.3 Alternative Explanation: Partisan Sentiment toward Big Tech Firms

An alternative explanation for the pronounced magnitude of the sentiment shift among partisan investors after the Jan.6 event could be that there was partisan favoritism toward certain firms, such as big tech firms. In this section, I disentangle this alternative explanation and lend more support to the notion that the realization of the future political landscape was the key driver of the partisanship sentiment shift. After the Jan.6 event, several major technology firms took actions to sever their ties with President Donald Trump and his associated businesses or political campaigns, largely due to concerns about his potential involvement in inciting the Capitol riot. For example, Twitter permanently suspended Trump's personal account and banned him from any further use of the platform. Moreover, Amazon removed the pro-Trump social media platform Parler from its cloud hosting services. The list of tech firms that suspended Trump-related operations in the sample, and their discontinuation of Trump-related business in the wake of the Capitol riot, can be found in Table 1.A.2.

An alternative explanation for the pronounced belief shift after the Jan.6 event is that partisan investors changed their attitude toward big tech firms after the Trump ban on social media and related services. Republican investors may have expressed resentment towards the big tech firms and more bearish sentiment on StockTwits, while Democrat investors may have done the opposite. Table 1.6 presents the empirical results of Equation (4) regarding the subsample of big tech firms listed in Table 1.A.2. As indicated by the coefficients on partisanship indicators, before the election outcome, Democrat investors tended to be more optimistic. In contrast, Republican investors tended to be more pessimistic toward big tech stocks, which contradicts the findings from the full sample regression. One possible explanation for this is that during the pre-election period, these big tech firms tended to support Democrat presidential candidate Joe Biden, potentially eliciting favor (resentment) from Democrat (Republican) investors. For example, Twitter suspended a *New York Post* report regarding a scandal involving Joe Biden's son.

Although the interaction term for the time dummy of 12/15/2020–01/05/2021 generally has no significant coefficients, the coefficients on the interaction terms have the same sign as the results from the full sample regression. However, the magnitude of these coefficients indicates that the sentiment shift among partisan investors after the Jan.6 event was much more pronounced for these big tech firms than for the other firms across the whole sample. For Republican (Democrat) investors, the magnitude of the decrease (increase) in aggregate sentiment toward big tech firms was -0.08355 (0.1270) versus 0.05394 (0.06040) for the whole sample. Big tech firms were popular among retail investors on StockTwits, constituting 14.26% of the sample (341,669 in 2,396,089). Meanwhile, in the subsample of big tech firms, whether the more pronounced sentiment shift after Jan.6 was caused by partisan investors' favoritism or resentment induced by the Trump ban remains uncertain. Due to the significant weight of posts related to big tech firms across the whole sample, it is essential to determine whether sentiment shifts toward big tech firms are the sole driver of the results in the whole sample regression.

I run the regression of Equation (4) in a subsample that excludes big tech firms, the results for which are shown in Table 1.7. Overall, the results do not differ significantly from those for the whole sample regression. For the interaction term between the partisanship indicator and the post-Jan.6 time dummy, the coefficients are similar to those in the whole sample regression in magnitude and statistical significance, albeit with slightly smaller values. Overall, the results indicate that the full sample regression results presented in Table 1. 5 are not driven by the potential changes in partisan favoritism toward big tech firms after the Trump ban, as demonstrated by the fact that the coefficients are not all that different in the subsample without big tech firms. These results support the hypothesis that the pronounced partisan belief shift after the Jan.6 event was driven by the full realization of the future political landscape after the congressional certification of the presidential election outcome.

1.8 Further Analysis: Systematic Risk Exposure

The existing literature shows that investors update their beliefs about economic conditions differently according to their partisan inclinations. To provide further evidence corroborating the hypothesis that heterogenous updates in the beliefs of partisan investors are driven by expectations regarding future economic conditions, I empirically examine how heterogenous partisan sentiment shifts change according to exposure to market risk. The rationale behind this line of inquiry is that if investors heterogeneously update their beliefs about the future performance of financial assets based on changes in economic conditions induced by shifts in the political landscape, then such heterogeneous belief updates will be more pronounced for securities with higher exposure to changes in overall economic conditions, i.e., systematic risk. As such, I hypothesize that after the outcome of the 2020 election, Republican investors tended to be more pessimistic about securities with higher exposure to systematic risk, while Democrat investors tended to feel the opposite. This further analysis is conducted from two perspectives. First, I empirically explore whether the heterogenous belief updates of partisan investors are more pronounced for major stock indices (SPY, QQQ, DIA). This is because, compared to individual stocks, these indices represent a greater portion of the overall performance of the financial market and are therefore more susceptible to systematic risk. Second, I investigate whether these heterogenous belief updates are more pronounced for individual securities with elevated exposure to market risk. This is accomplished by running the regression of Equation (4) in subsamples categorized by historical market beta.

Table 1.8 reports the empirical results for the subsample of posts containing major stock indices (SPY, QQQ, and DIA). For aggregate sentiment in the subsample of major stock indices, the coefficients on the interaction term between time dummies and partisan indicators are -0.05230, -0.06889, and -0.8531 for Republican investors, compared to -0.03616, -0.02760, and -0.05394 for the full sample regression. For Democrat investors, the coefficients are 0.09398, 0.05494, and 0.08284 for the subsample of indices and 0.04324, 0.02484, and 0.06040 for the full sample regression. Overall, these empirical results indicate a larger magnitude of coefficients on the interaction term between time dummy and partisan indicators than for the full sample regression. This finding aligns with my hypothesis that the expectation of future economic conditions played a significant role in driving the belief updates among partisan investors after the election outcome.

To show that these outcomes are not solely driven by indices and big techs, I also run the regression in the subsample without indices or big tech firms. The results for this are shown in Table 1.9, from which it can be seen that there are no significant changes in the coefficients.

To provide further evidence for the hypothesis, I run the regression of Equation (4) in subsamples of individual stock categorized by exposure to market risk. For each stock in the sample, I run a monthly regression of the Fama-French three-factor model in the 36 months before 08/2020 to measure its market beta. I then further divide the full sample of stocks into three quantiles based on their estimated market betas and run the subsample regression of Equation (4) based on these three quantiles.

Tables 1.10–1.12 report the regression results for the subsample of stocks categorized into the first, second, and third quantile of market beta, where the first quantile represents stocks with the lowest market beta, and the third quantile covers stocks with the highest market beta. Overall, the results indicate that the heterogenous belief update is more pronounced for stocks with higher market risk exposure because the coefficients on the interaction term become more pronounced from the first to the third quantile. In the subsample of the first market beta quantile, the coefficients for the interaction term between time dummies and the Republican indicator are -0.02670, -0.02248 for the post 11/06/2020 and post 01/06/2021 time dummy, and statistically insignificant for the time dummy for 12/15/2020–01/05/2021. Meanwhile, the coefficients for the Democrat indicator are 0.01683, 0.01927, and 0.01467. In the subsample of the second market beta quantile, the coefficients are -0.03978, -0.04393, and -0.06713 for Republicans and 0.06530, 0.02638, and 0.05836 for Democrats. In the third market beta quantile subsample, the coefficients are -0.02762, -0.02968, and -0.05733 for Republicans and 0.02801, 0.02187, and 0.06628 for Democrats. The coefficients are more pronounced in the second and third market beta quantiles in terms of magnitude and statistical significance than in the first market beta quantile.

This increasing magnitude of coefficients across market beta quantiles indicates that the heterogenous belief updates among partisan investors were more concentrated in stocks with higher market risk exposure. The empirical results imply that exposure to systematic risk was a crucial factor driving the heterogenous partisan belief updates about the financial market. The sentiment shift among partisan investors after the election can be attributed to their heterogeneous private model of processing public information about the future political landscape, which in turn resulted in a heterogenous expectation of future economic conditions and, therefore, financial market performance. As such, major stock indices and stocks more exposed to systematic risk experienced a more pronounced shift in the sentiment of partisan investors.

1.9 Partisan Disagreement and Stock Liquidity

As shown in the previous section, the post-election period witnessed the emergence of heterogeneous belief updates among partisan investors. Republican investors tended to become more pessimistic toward the financial market, while Democrat investors tended to display the opposite sentiment. This heterogeneous belief update led to disagreement regarding the financial market, with investors forming heterogeneous expectations about future stock returns. Such partisan disagreement may have affected market liquidity, particularly when partisan investors engaged in trading activities that reflected their divergent beliefs about future stock performance. In light of this, in this section, I explore how partisan disagreement after the 2020 election led to changes in the liquidity of individual stocks. Specifically, I empirically examine the impact of partisan disagreement on stock-level liquidity measures, including abnormal turnover rate and effective percentage spread.

1.9.1 Variable Construction and Summary Statistics

To assess the impact of partisan disagreement on daily stock liquidity, I first construct a measurement of daily stock-level partisan disagreement. Given the relatively limited representation of Democrat investors in the dataset compared to Republican investors, measuring disagreement between Republican and Democrat investors at the daily stock level will result in a significant amount of missing data. Therefore, in this analysis, I focus on the disagreement between Republican investors and non-Republican investors as a means by which to assess the impact of partisan disagreement on daily stock-level liquidity.

Following Cookson, Engelberg, and Mullins (2020), the daily stock-level partisan disagreement is measured using the following equation:

$$Partisan\_Disagreement_{s,t} = |Republican\_Sentiment_{s,t} - Non-Republican\_Sentiment_{s,t}| \tag{5}$$

Specifically, for stock s at date t, partisan disagreement is calculated as the absolute value of the difference between the average value of the aggregate sentiment of Republican investors and non-Republican investors; $Republican_Sentiment_{s,t}$is measured as the mean of the aggregate sentiment of Republican investors on stock s at date t; and $Non-Republican\_Sentiment_{s,t}$ is defined analogously for Non-Republican investors.

1.9.2 Cross Section of Stock Liquidity

To analyze the cross-section of daily stock-level liquidity, I concentrate on a daily panel of stock-level data from 08/01/2020 to 03/31/2021. Using data on daily stock-level information from CRSP and TAQ, I construct daily measurements of liquidity and other control variables. In particular, I explore the impact of partisan disagreement before and after the release of the election outcome. To do this, I empirically estimate the impact of daily stock-level partisan disagreement on liquidity by using the following empirical specification:

$$\begin{aligned} \text{Liquidity}_{s,t} &= \beta_0 + \beta_1 \times \text{Partisan\_Disagreement}_{s,t} + \sum_{n=0}^{2} \theta_n \times \text{After}_{n,t} \\ &\sum_{n=0}^{2} \gamma_n \times \text{After}_{n,t} \times \text{Partisan\_Disagreement}_{s,t} + X_{s,t} + \alpha_s + \theta_t + \epsilon_{s,t} \end{aligned} \tag{6}$$

The dependent variable, $Liquidity_{s,t}$, can represent either the abnormal turnover rate or the effective percentage spread for stock s at time t. The turnover rate is calculated by dividing the total number of traded shares by the total number of outstanding shares. The abnormal turnover rate for stock s at date t is calculated as the difference between the turnover rate at time t and the average turnover rate over the past 20 days. The effective percentage spread of stock s at date t is measured as the average of the effective percentage spread, which is calculated using the method introduced by Holden and Jacobsen (2014). The trade is then classified as buyer- or seller-initiated by using the method proposed by Lee and Ready (1991).

The key independent variables are $Partisan\_Disagreement_{s,t}$ and its interaction with the time dummies indicating the post-election period. $X_{s,t}$ refers to a list of control variables, which include idiosyncratic volatility from t-5 to t-1, cumulative abnormal return from t-5 to t-1, cumulative abnormal return from t-30 to t-5, and overall disagreement at time t. Overall disagreement is calculated as the standard deviation of the aggregate sentiment of users toward stock s at date t. Detailed descriptions of variable construction can be found in Table 1.A.3. Meanwhile, Table 1.13 presents the summary statistics of the variables used in Equation (6) as input for analysis. Additionally, I include stock fixed effect α_s and time fixed effect θ_t to control for any stock-level or time-level invariant unobservable. $epsilon_{s,t}$ refers to the error term. To control for serial or within-day correlation of the error term, I cluster the standard error by stock and date.

The independent variable of primary interest is partisan disagreement and its interaction with multiple time dummies indicating the post-election period. I expect partisan disagreement to increase stock liquidity, especially during the post-election period. Increased liquidity is associated with a higher turnover rate and lower trading costs, as measured by the effective percentage spread. Therefore, when the dependent variable is the abnormal turnover rate, I expect the coefficients on $Partisan\_Disagreement_{s,t}$ and its interaction term with time dummies to be positive. When the dependent variable is the effective percentage spread, I expect the coefficients to be negative.

Additionally, I run the regression on quantiles categorized by the market beta of stocks in order to determine the interactive impact of systematic risk exposure and partisan disagreement on stock liquidity. I expect the effect of disagreement on stock liquidity to be more pronounced for stocks with higher exposure to systematic risk.

1.9.3 Empirical Results for the Cross Section of Stock Liquidity

Tables 1.14 and 1.15 present the empirical results of the regression from Equation (6), with abnormal turnover rate and effective percentage spread serving as the dependent variables, respectively. Column 1 presents the results for the full sample regression, while Columns 2–4 present the results for the subsample of the first, second, and third quantiles based on stock market beta.

Table 1.14 presents the results with the abnormal turnover rate as the dependent variable. In contrast to the hypothetical expectation, the coefficient on the independent variable $Partisan\_Disagreement$ is negative and statistically significant in the full sample regression and all subsamples divided by market beta. The coefficients on the interaction term between $Partisan\_Disagreement$ and the time dummies of post-election periods are generally positive and statistically significant except for the subsample of the lowest market beta quantile. Notably, the coefficients' magnitude and statistical significance generally increase from the first to the third market beta quantile. Overall, these results indicate that partisan disagreement led to an increased turnover rate after the announcement of the election outcome. This increase was significant both after the calling of the election result on 11/06/2020 and after the Jan.6 event, but it was more pronounced for the former. Contrary to the results with sentiment as the dependent variable, these results indicate that partisan disagreement may have induced more turnover between investors with heterogenous partisan beliefs immediately after the election outcome than the Jan.6 event. These results indicate that although the disparity in partisan sentiment became most pronounced after the Jan.6 event, the trading behavior induced by such a disparity was more pronounced after the election outcome called on 11/06/2020. Thus, investors may have incorporated the state of the future political landscape of the Democrat White House into their trading action, primarily immediately after the call of the election outcome.

Additionally, I find that the magnitude and statistical significance of the coefficients generally increase from the first to the third market beta quantiles. Overall, the coefficients on the interaction term are statistically insignificant for the subsample of the first beta quantile but become significant for the subsample of the second and third beta quantiles. The empirical results indicate that the abnormal turnover rate induced by partisan disagreement was more concentrated in stocks with higher exposure to market risk. This implies that after the announcement of the election outcome, partisan investors may have traded against those with different beliefs on future economic conditions, which then resulted in a more pronounced increase in the abnormal turnover rate in stocks with higher exposure to systematic risk.

Table 1.15 presents the results with the effective percentage spread serving as the dependent variable. The coefficients on $Partisan\_Disagreement$ and its interaction terms with time dummies are generally negative and statistically significant. The empirical results indicate that partisan disagreement increased stock-level liquidity by reducing the effective percentage spread, and this effect became more pronounced in the post-election periods. The magnitude and significance of coefficients are mixed for the interaction term with time dummies across quantiles divided by market beta. In the first and third market beta quantiles, the magnitude of coefficients is most pronounced for the interaction term with the post-Jan.6 time dummy, while for the second quantile, the coefficient is positive and statistically insignificant.

Tables 1.A.4 and 1.A.5 present the results with dollar-weighted percentage spread and share-weighted percentage spread serving as the dependent variables. Unlike the results with an equal-weighted average, the coefficients on $Partisan\_Disagreement$ and its interaction term with time dummies are generally statistically insignificant. The results imply that the decrease in effective percentage spread induced by partisan disagreement is mostly concentrated in small-size trades in terms of share and dollar value since the coefficients are statistically significant when the spread is equally weighted across trades while becoming insignificant when the spread is weighted by the number of shares or dollar value. This further implies that the sentiment-driven decrease in the effective spread is primarily concentrated in trades initiated by small-size trades submitted by retail investors. Thus, partisan disagreement on social media platforms may have largely reflected the beliefs of retail investors, whose investment decisions and trading behavior were more likely to be sentiment-driven.

Overall, the empirical results indicate that partisan disagreement tended to increase stock-level liquidity by increasing abnormal turnover rate and decreasing liquidity, particularly in the post-election period. For abnormal turnover rate, the increasing effect was more pronounced in the period immediately after the call of the election result on 11/06/2020 and more concentrated among stocks with higher exposure to market risk. This indicates that partisan disagreements were mostly incorporated into the trading actions of investors immediately after the call of the election results, particularly for stocks with higher exposure to systematic risk. Meanwhile, for effective percentage spread, the magnitude and significance of coefficients are relatively mixed.

1.9.4 Intraday Volatility

Another issue that needs to be addressed is the potential impact of partisan disagreement on the intraday volatility of stocks after the election outcome. Increased partisan disagreement may have led to more dispersion of belief priced into the stock price, which would have further increased the volatility of the stock price. To test this hypothesis, I run the regression of Equation (4) with abnormal intraday volatility serving as the dependent variable. Intraday volatility of stock s at date t is calculated as the standard deviation of the quoted price for stock s within date t. Abnormal intraday volatility of stock s at date t is calculated as the difference between the intraday volatility of stock at date t and the average intraday volatility from $t-30$ to $t-1$. Under the hypothesis of partisan disagreement increasing intraday volatility, I expect the coefficients on the $Partisan\_Disagreement$ and its interaction term with post-election time dummies to be positive and statistically significant.

Table 1.16 presents the empirical results from the regression with abnormal intraday volatility serving as the dependent variable. Contrary to expectation, the coefficients on the variable $Partisan\_Disagreement$ are negative and statistically significant. The coefficients on the interaction term between $Partisan\_Disagreement$ and post-election period time dummies are positive and statistically significant for some of the time dummies in the full sample regression, which implies that after the election outcome, partisan disagreement was associated with increased intraday quote-based volatility. Moreover, for the subsamples divided by market beta quantile, the coefficients are most pronounced for the second market beta quantile, attenuated for the first quantile, and muted for the third quantile. Overall, these results lend some support to the hypothesis that partisan disagreement increased stock-level intraday volatility after the announcement of the election outcome. However, the impact of market risk exposure on the volatility increased by partisan disagreement remains uncertain, as I can detect no consistent pattern across market beta quantiles.

1.9.5 Discussion

In this subsection, I discuss the heterogenous effect of partisan disagreement on various liquidity measures and intraday volatility. Specifically, when we consider the results using the abnormal turnover rate or abnormal intraday volatility as the dependent variable, partisan disagreement appears to have a dampening effect in the pre-election period but an amplifying effect in the post-election period. However, when assessing the results with the effective percentage spread serving as the dependent variable, partisan disagreement has a dampening effect in both the pre-and post-election periods. These results indicate that partisan disagreement reduces the abnormal turnover rate and intraday volatility while decreasing the effective percentage spread in the pre-election period. This partially contradicts the notion that partisan disagreement improves liquidity by motivating partisan investors to trade against each other and that partisan disagreement increases volatility by pricing more heterogenous beliefs into asset prices, particularly during the pre-election period.

Here I offer an explanation for these seemingly contradictory results. According to the theoretical framework of Kruger (2020), heterogeneous belief increases liquidity only in the presence of overconfidence, where investors with private information overestimate their own signals and are willing to trade against those with different private information. In the pre-election period, the election result remained uncertain, meaning that no concrete public information was available for the overconfident partisan investors to interpret with their own private model. In such a scenario, partisan investors, despite holding different views on the impact of potential election outcomes on stock returns, may have withheld from trading in the pre-election period, waiting for the release of public information to be processed with their private models. This may explain why partisan disagreement is associated with the lower abnormal turnover rate in the pre-election period. Additionally, as partisan investors withheld from trading in the pre-election period, a lower amount of their heterogenous private information was priced into stock prices, which may explain the negative association between partisan disagreement and intraday volatility.

As for the negative association between partisan disagreement and effective percentage spread in the pre-election period, this can also be explained by the lack of a definitive public signal. During the pre-election period, as partisan investors withheld from trading due to a lack of public signal, the market's fraction of traders with private information decreased. As such, there likely existed a higher fraction of liquidity traders in the market, which reduced the effective percentage spread. In such a scenario, higher partisan disagreement in the pre-election period led to lower market participation among partisan investors with heterogenous private information due to a lack of clear public signals. In turn, the fraction of liquidity traders increased, resulting in a lower market effective percentage spread.

1.10 Conclusion

Political polarization is a major ongoing issue in contemporary American politics. Since the election of President Trump in 2016, political polarization has divided public opinion on a wide range of topics, including climate change, trade policy, tax policy, immigration, racial issues, and gun control. Recent literature focuses on the impact of political division on belief divergence in regard to economic issues and whether such partisan disagreement has an impact on economic action.

This study is one of the few, though, that explores the connection between political polarization and belief divergence regarding the financial market. Using data from social media posts from one of the most popular investment platforms in the United States, I empirically examine the impact of political climate shifts on partisan disagreement about future stock returns, using the 2020 presidential election as an event study. The empirical results indicate that the election outcome, which resulted in a Democrat presidency, led to a substantial increase in partisan disagreement between Republican and Democrat investors in regard to future stock returns. Moreover, as the election outcome of the Democrat presidency became more solidified in the post-election period, the partisan disagreement on future stock returns intensified. Further analysis indicates that the partisan disagreement induced by the election outcome was more concentrated in major stock indices (SPY, QQQ, DIA) and stocks with higher market beta. These results suggest that partisan disagreement about stock returns primarily stemmed from heterogenous expectations of future market conditions. In essence, partisan disagreement regarding the financial market was largely determined by how Republican and Democrat investors interpreted the potential policy changes accompanying shifts in political power, and how these changes might have impacted the overall economic landscape as perceived by partisan investors.

Furthermore, this study also connects the literature on partisan division and the strand of research on disagreement and trading behavior. Specifically, by using the 2020 election as an event study, I explore the impact of partisan investor disagreement on stock liquidity and intraday volatility. Overall, the results indicate that during the post-election period, partisan disagreement was associated with increased liquidity, as measured by a higher abnormal turnover rate and lower effective percentage spread. These results are consistent with the theoretical framework of Kruger (2020), in which investors with heterogenous private information are willing to trade against each other in the presence of overconfidence, resulting in increased liquidity. In the post-election period, partisan investors may have overestimated their own interpretation of the election outcome and actively sought to trade against those with opposing views, leading to increased stock market liquidity. Additionally, further results indicate that in the post-election period, partisan disagreement was associated with higher intraday volatility, implying that heterogenous beliefs were priced into stock prices.

The findings of this study have important implications for future research. Firstly, the results indicate that major political events may trigger remarkable disagreement among partisan investors toward stock returns in the current context of political division. This partisan disagreement may also influence aggregate trading behavior, further shaping stock liquidity and volatility. As partisan politics continue to prevail in the United States, political polarization will persist and lead to further belief division in the country. Whether such belief division influences the formation of investor expectations and subsequent investment action is a topic worthy of exploration.

Additionally, future studies could use political events other than the presidential election as a lens through which to explore potential partisan disagreement. For example, major policy changes regarding partisan issues such as climate change, trade policy, and tax policy may trigger significant partisan disagreement on the economic impact as partisan agents may hold remarkably different opinions on the pros and cons of these issues. Whether these policy changes induce divergence in expectations of stock returns is worth further exploration. For example, drawing on the fact that Republicans and Democrats hold different views on climate issues, future studies could empirically examine whether climate-related policy shifts influence partisan investor belief in regard to ESG-related investments.

References

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth?. *The Journal of Politics*, 70(2), 542-555.

Addoum, J. M., & Kumar, A. (2016). Political sentiment and predictable returns. *The Review of Financial Studies*, 29(12), 3471-3518.

Addoum, J. M., Delikouras, S., Ke, D., & Kumar, A. (2019). Underreaction to political information and price momentum. *Financial Management*, 48(3), 773-804.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, 191, 104254.

Andrei, D., Carlin, B., & Hasler, M. (2019). Asset pricing with disagreement and uncertainty about the length of business cycles. *Management Science*, 65(6), 2900-2923.

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

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.

Banerjee, S., & Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance*, 65(4), 1269-1302.

Barrios, J. M., & Hochberg, Y. (2020). Risk perception through the lens of politics in the time of the COVID-19 pandemic (No. w27008). National Bureau of Economic Research.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24, 117-150.

Bartels, L. M. (2018). Partisanship in the Trump era. *The Journal of Politics*, 80(4), 1483-1494.

Belo, F., Gala, V. D., & Li, J. (2013). Government spending, political cycles, and the cross section of stock returns. *Journal of Financial Economics*, 107(2), 305-324.

Benhabib, J., & Spiegel, M. M. (2019). Sentiments and economic activity: Evidence from US states. *The Economic Journal*, 129(618), 715-733.

Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *The Accounting Review*, 89(1), 79-112.

Blinder, A. S., & Watson, M. W. (2016). Presidents and the US economy: An econometric exploration. *American Economic Review*, 106(4), 1015-1045.

Bishop, B. (2008). *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart.* New York, Houghton Mifflin.

Brown, J. R., & Huang, J. (2020). All the president's friends: Political access and firm value. *Journal of Financial Economics*, 138(2), 415-431.

Burton, A. L., Logan, M. W., Pickett, J. T., Cullen, F. T., Jonson, C. L., & Burton Jr, V. S. (2021). Gun owners and gun control: shared status, divergent opinions. *Sociological inquiry*, 91(2), 347-366.

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, 114(2), 226-238.

Cassidy, W., & Vorsatz, B. (2021). Partisanship and portfolio choice: Evidence from mutual funds. *Working Paper.*

Chang, Y. C., Hong, H. G., Tiedens, L., Wang, N., & Zhao, B. (2015). Does diversity lead to diverse opinions? Evidence from languages and stock markets. *Working Paper*, (168), 13-16.

Chen, Z., Da, Z., Huang, D., & Wang, L. (2023). Presidential economic approval rating and the cross-section of stock returns. *Journal of Financial Economics*, 147(1), 106-131.

Chen, H., De, P., Hu, Y., & 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.

Chen, H., & Hwang, B. H. (2022). Listening in on investors' thoughts and conversations. *Journal of Financial Economics*, 145(2), 426-444.

Chen, H., Hwang, B. H., & Peng, Z. (2023). Revisiting the Cross-Section of Expected Stock Returns: Evidence from a Textual Analysis of Buy Recommendations. *Working Paper.*

Coibion, O., Gorodnichenko, Y., & Weber, M. (2020). Political polarization and expected economic outcomes (No. w28044). National Bureau of Economic Research.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2020). Does partisanship shape investor beliefs? Evidence from the COVID-19 pandemic. *The Review of Asset Pricing Studies*, 10(4), 863-893.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2023). Echo chambers. *The Review of Financial Studies*, 36(2), 450-500.

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., & Schiller, C. (2023). Social media as a bank run catalyst. *Working Paper.*

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.

Cookson, J. A., Niessner, M., & Schiller, C. (2023). Can social media inform corporate decisions? Evidence from merger withdrawals. Evidence from Merger Withdrawals. *Working Paper.*

Cooper, M. J., Gulen, H., & Ovtchinnikov, A. V. (2010). Corporate political contributions and stock returns. *The Journal of Finance*, 65(2), 687-724.

Dagostino, R., Gao, J., & Ma, P. (2020). Partisanship in loan pricing. *Working Paper.*

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5), 2113-2141.

Engelberg, J., Guzman, J., Lu, R., & Mullins, W. (2022). Partisan entrepreneurship (No. w30249). *Working Paper.*

Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., & Verkuilen, J. (2007). Same facts, different interpretations: Partisan motivation and opinion on Iraq. *The Journal of Politics*, 69(4), 957-974.

Gerber, A. S., & Huber, G. A. (2009). Partisanship and economic behavior: Do partisan differences in economic forecasts predict real economic behavior?. *American Political Science Review*, 103(3), 407-426.

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2016). Why people vote: Estimating the social returns to voting. *British Journal of Political Science*, 46(2), 241-264.

Gillitzer, C., & Prasad, N. (2018). The effect of consumer sentiment on consumption: Cross-sectional evidence from elections. *American Economic Journal: Macroeconomics*, 10(4), 234-269.

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

Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance*, 31(2), 573-585.

Gorbatikov, E., van Lent, L., Naik, N. Y., Sharma, V., & Tahoun, A. (2019). Is firm-level political exposure priced?. *Working Paper*

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *The Review of Financial Studies*, 6(3), 473-506.

Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4), 2135-2202.

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

Hong, H., & Stein, J. C. (1999). Differences of opinion, rational arbitrage and market crashes. National Bureau of Economic Research

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, 21(2), 109-128.

Jung, M. J., Naughton, J. P., Tahoun, A., & Wang, C. (2018). Do firms strategically disseminate? Evidence from corporate use of social media. *The Accounting Review*, 93(4), 225-252.

Kandel, E., & Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4), 831-872.

Kaplan, E., Spenkuch, J., & Sullivan, R. (2019, January). Measuring geographic polarization: Theory and long-run evidence. In American Political Science Association conference, Boston, August.

Ke, D. (2022). Left behind: Partisan identity and wealth inequality. *Working Paper*.

Kelley, E. K., & Tetlock, P. C. (2013). Why do investors trade?. *Working Paper*.

Kempf, E., & Tsoutsoura, M. (2021). Partisan professionals: Evidence from credit rating analysts. *The Journal of Finance*, 76(6), 2805-2856.

Kim, C. F., Pantzalis, C., & Park, J. C. (2012). Political geography and stock returns: The value and risk implications of proximity to political power. *Journal of Financial Economics*, 106(1), 196-228.

Kruger, S. (2020). Disagreement and liquidity. *Working Paper*.

Kyle, A. S., Obizhaeva, A. A., & Wang, Y. (2018). Smooth trading with overconfidence and market power. *The Review of Economic Studies*, 85(1), 611-662.

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, 53(2), 367-404.

Liu, L. X., Shu, H., & Wei, K. J. (2017). The impacts of political uncertainty on asset prices: Evidence from the Bo scandal in China. *Journal of Financial Economics*, 125(2), 286-310.

Marston, C., Renedo, A., & Miles, S. (2020). Community participation is crucial in a pandemic. *The Lancet*, 395(10238), 1676-1678.

Meeuwis, M., Parker, J. A., Schoar, A., & Simester, D. (2022). Belief disagreement and portfolio choice. *The Journal of Finance*, 77(6), 3191-3247.

Mian, A., Sufi, A., & Khoshkhou, N. (2021). Partisan bias, economic expectations, and household spending. *Review of Economics and Statistics*, 1-46.

McGrath, M. C. (2017). Economic behavior and the partisan perceptual screen. *Quarterly Journal of Political Science*, 11(4), 363-383.

Milner, H. V., & Judkins, B. (2004). Partisanship, trade policy, and globalization: Is there a left–right divide on trade policy?. *International Studies Quarterly*, 48(1), 95-119.

Pástor, Ľ., & Veronesi, P. (2020). Political cycles and stock returns. *Journal of Political Economy*, 128(11), 4011-4045.

Santa-Clara, P., & Valkanov, R. (2003). The presidential puzzle: Political cycles and the stock market. *The Journal of Finance*, 58(5), 1841-1872.

Verardo, M. (2009). Heterogeneous beliefs and momentum profits. *Journal of Financial and Quantitative Analysis*, 44(4), 795-822.

Tables and Figures

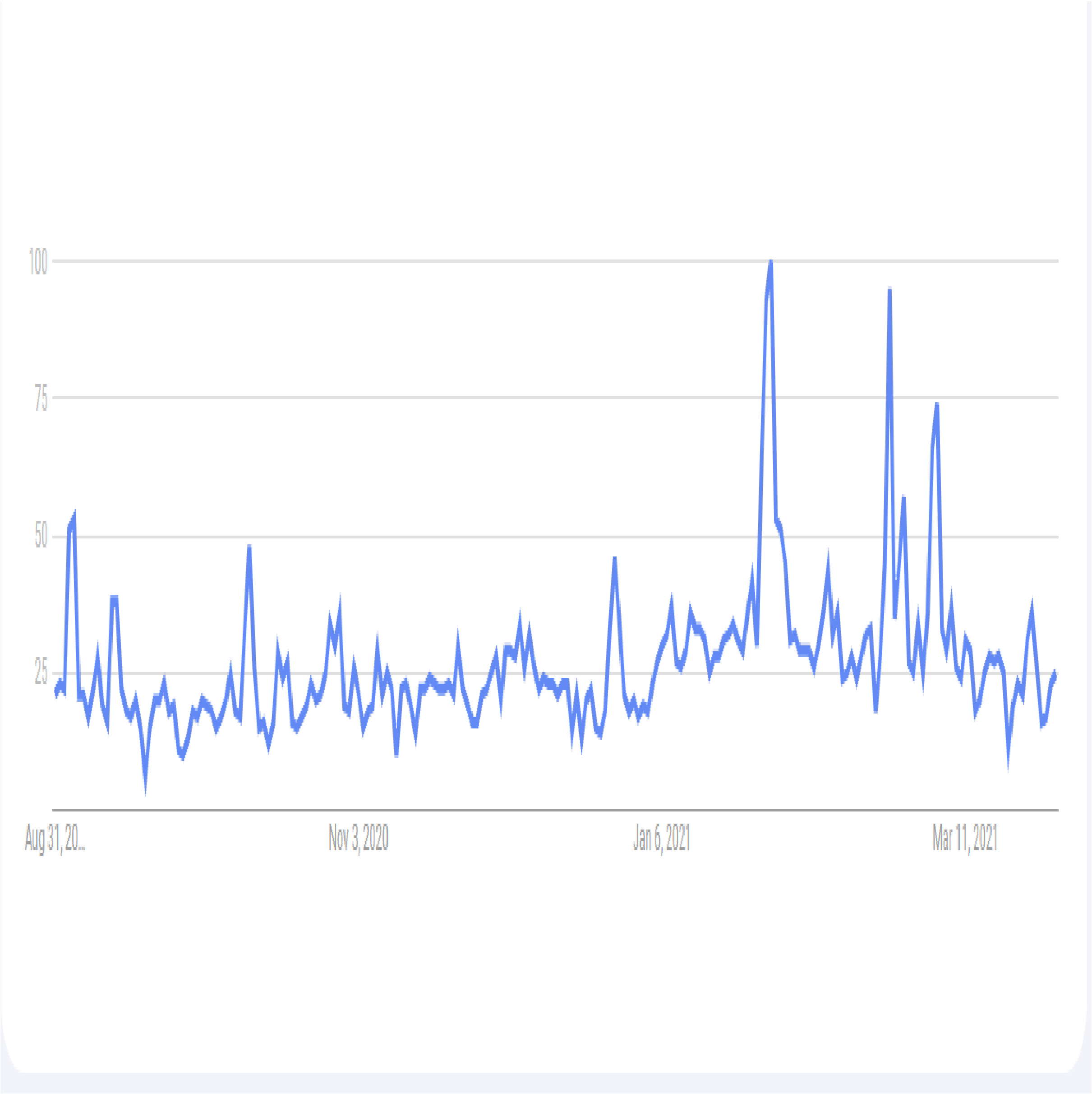


Figure 1.1: Google Search Index for “Stock Market Crash” around the 2020 Election

File outline Imported from Another project/Table2 Tryout.tex, at 10:46 am Today

Table 1.1: Classification of Partisan Posts

This table reports the classification of partisan posts. Panel (a) reports partisan keywords to identify the partisan inclination for a given post. Panel (b) reports examples of posts with Republican or Democrat inclinations.

(a) Partisan Keywords	
Republican Keywords	’stop the steal’,’election fraud’,’fake election’, ’false election’, ’sleepy joe’, ’hussein obama’, ’obummer’, ’fake news media’, ’crooked hillary’, ’snowflake’, ’liberal Media’, ’libs’, ’trump hater’, ’typical liberal’, ’liberal agenda’, ’your liberal’, ’russian collusion’, ’stupid dems’, ’trump derangement’, ’maga 2020’, ’maga2020’, ’maga 2024’, ’maga 2024’,’stupid democrats’, ’liberal democrat’, ’you liberal’, ’maga’, ’russian hoax’, ’the liberals’, ’leftist’, ’socialist’,’commie’, ’libtard’, ’trump won’, ’libtard’, ’america first’,’trump2020’, ’trump2024’, ’trump 2020’, ’trump 2024’, ’trump win’
Democrat Keywords	’drumpf’, ’trump nationalism’, ’trumptard’, ’magatard’, ’trump is a liar’, ’big lie’, ’biglie’, ’idiot trump’,’faux news’, ’clown child’, ’stupid trump’, ’pig clown’, ’liar in chief’, ’liar trump’, ’f***trump’, ’orange colored’, ’orange man’, ’orangeman’, ’idiot in chief’, ’criminal potus’, ’trump is an idiot’, ’clown trump’, ’imbecile trump’, ’trump is an imbecile’, ’orange scum’, ’scumping clown’, ’lyingtrump’, ’lying trump’, ’impeachtrump’, ’impeach trump’, ’#impeachtrump’, ’biden2020’, ’biden 2020’
(b) Examples of Partisan Tweets	
Republican Posts	
12/11/20	$ORCL are all of the liberals going to vote the same crazy polices that caused them to leave ca or start voting smart for lower taxes and personal freedom?!?!?!?!
12/4/20	$CCL $RCL ok bulls, explain to me how you think the biden admin will allow ships to sail in spring (or even summer) when hes demanding 100 day mandatory mask wearing and the vax wont even be out in large numbers until fall 2021? all those people visiting foreign countries and bring the virus back in? mind u, i think covid is no worse than a bad flu season (numbers back it up) but libtards voted for this clown. dont see rainbows and cupcakes anytime soon, just dark days and lockdowns.
12/21/20	$SPY america first right?!
Democrat Posts	
12/18/20	$BORR $OXY $UAL $AAL $save final popular vote count. biden- 81,283,485 drumpf- 74,223,744 electoral college biden- 306 drumpf 232 biden wins in a landside, with no proven fraud. every court case has been thrown out for lack of evidence. who else is excited for drumpf to be sent to federal prison next month?!
12/19/20	$QQQ clear bearish engulfing candle on the daily... then we go over to $dxy and see an inside bar on the daily as well! this comes as stimulus negotiations are being held up by bernie, the orange man, and other pro stimulus check representatives... congress was just barely able to avoid a government shutdown which would’ve cost thousands of more jobs on top of the 885k unemployment claims, 10% above the 800k forecasted... the virus is still raging on and cdc is predicting almost 400k deaths by january 9th! glta, clear your mind this weekend! spend time with family and enjoy the holiday season! $SPY $DIA
12/10/20	$SPY $QQQ crazy trump will punish tech on his way out. watch out. crazy clown trump

Table 1.2: Summary Statistics on StockTwits Data

This table reports the summary statistics for data on posts from StockTwits. Panel (a) reports the summary statistics for sample posts from StockTwits. Panel (b) reports the summary statistics for the partisan classification results for the sample posts, based on the keyword list method introduced in Table 1.

(a) Summary Statistics on Posts on Stocktwits.com			
Total Number of Posts	4841244		
Number of Bullish Posts	1,579,343	Number of Posts Contain Republican Keywords	13,403
Number of Bearish Posts	464,467	Number of Posts Contain Democrat Keywords	2,195
Number of Untagged Posts	2,797,434		
(b) Summary Statistics for the Partisan Classification of Users and Posts			
Total Number of Users	135,404		
Repunlican Users	4,453	Number of Posts Made by Republican Users	964,935
Democrat Users	1,002	Number of Posts Made by Democrat Users	240,563
Other Users	134,872	Number of Posts Made by Other Users	3,806,364

Table 1.3: Summary Statistics on User-Security-Date Data

This table reports the summary statistics for data at the user-security-date level..

	Sample	Before 11/07/2020	11/07/2020 - 12/14/2020	12/15/2020 - 01/05/2021	After 01/06/2021
Number of User-Security-Date Observations	2,396,089	987,807	339,198	185,255	883,829
Number of Ticker	503	501	503	503	503
Number of Days	243	98	38	23	84
Number of Users	135,404	72,055	42,559	30,400	84,802

Table 1.4: Summary Statistics on Sentiment by Partisanship

This table reports the summary statistics for the bullish, bearish, and aggregate sentiment of StockTwits posts posted by Republican, Democrat, and non-partisan investors. The summary statistics for subsamples based on the event period are also reported.

Whole Sample				
		Num.of.Obs	Mean	SD
	Bullish	2,396,089	0.3260	0.4542
	Bearish	2,396,089	0.0722	0.2485
	Sentiment	2,396,089	0.2538	0.5594
Republican				
	Bullish	632491	0.2399	0.4089
	Bearish	632491	0.0739	0.2473
	Sentiment	632491	0.1660	0.5106
Democrat				
	Bullish	195940	0.1931	0.3781
	Bearish	195940	0.0660	0.2333
	Sentiment	195940	0.1271	0.4691
Other				
	Bullish	1725477	0.3559	0.4654
	Bearish	1725477	0.0708	0.2476
	Sentiment	1725477	0.2852	0.5714
Before 11/07/2020				
Republican				
	Bullish	269893	0.2511	0.4138
	Bearish	269893	0.0693	0.2391
	Sentiment	269893	0.1818	0.5102
Democrat				
	Bullish	82324	0.1824	0.3687
	Bearish	82324	0.0741	0.2461
	Sentiment	82324	0.1083	0.4696
Other				
	Bullish	701531	0.3505	0.4628
	Bearish	701531	0.0739	0.2522
	Sentiment	701531	0.2765	0.5725
11/07/2020 - 12/14/2020				
Republican				
	Bullish	94023	0.2429	0.4110
	Bearish	94023	0.0718	0.2443
	Sentiment	94023	0.1711	0.5104
Democrat				
	Bullish	29568	0.1984	0.3822
	Bearish	29568	0.0603	0.2232
	Sentiment	29568	0.1381	0.4654
Other				
	Bullish	239823	0.3629	0.4680
	Bearish	239823	0.0669	0.2420
	Sentiment	239823	0.2960	0.5698

Table 1.4: Continued

12/15/2020 - 01/05/2021				
Republican				
	Bullish	49089	0.2566	0.4188
	Bearish	49089	0.0669	0.2350
	Sentiment	49089	0.1896	0.5110
Democrat				
	Bullish	15759	0.2061	0.3876
	Bearish	15759	0.0602	0.2214
	Sentiment	15759	0.1459	0.4700
Other				
	Bullish	133092	0.3674	0.4695
	Bearish	133092	0.0587	0.2267
	Sentiment	133092	0.3087	0.5599
After 01/06/2021				
Republican				
	Bullish	219486	0.2210	0.3988
	Bearish	219486	0.0820	0.2607
	Sentiment	219486	0.1390	0.5100
Democrat				
	Bullish	68289	0.2008	0.3849
	Bearish	68289	0.0601	0.2240
	Sentiment	68289	0.1407	0.4690
Other				
	Bullish	651031	0.3570	0.4664
	Bearish	651031	0.0713	0.2485
	Sentiment	651031	0.2857	0.5730

Table 1.5: Baseline Results

This table reports the baseline empirical results with OLS estimation of Equation (4). The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.05689*	0.01614***	-0.07303**
	(-1.75)	(9.73)	(-2.24)
GOP	-0.03167	-0.02726***	-0.004408
	(-1.27)	(-3.76)	(-0.18)
After0	0.01464***	-0.005361***	0.02000***
	(5.53)	(-3.12)	(4.85)
After1	0.01662***	-0.01423***	0.03085***
	(5.70)	(-8.25)	(7.32)
After2	0.01522***	-0.005814***	0.02103***
	(6.18)	(-3.57)	(5.42)
After0 × GOP	-0.02063***	0.01553***	-0.03616***
	(-8.79)	(12.08)	(-11.76)
After1 × GOP	-0.01212***	0.01547***	-0.02760***
	(-4.14)	(9.78)	(-7.42)
After2 × GOP	-0.03038***	0.02356***	-0.05394***
	(-16.02)	(19.87)	(-20.21)
After0 × Dem	0.02201***	-0.02123***	0.04324***
	(7.73)	(-11.30)	(11.08)
After1 × Dem	0.01040***	-0.01444***	0.02484***
	(2.70)	(-6.25)	(5.00)
After2 × Dem	0.03426***	-0.02614***	0.06040***
	(15.39)	(-17.22)	(19.59)
N	2396089	2396089	2396089
adj. R^2	0.075	0.039	0.055

Table 1.6: Results with Big Tech Firms

This table reports the empirical results with OLS estimation of Equation (4) on the subsample of big tech firms involved in the "Trump Ban" after the Capitol riot on 01/06/2021. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	0.05172***	-0.1170***	0.1687***
	(4.51)	(-8.37)	(7.15)
GOP	-0.3049***	0.2572***	-0.5621***
	(-33.82)	(14.00)	(-22.32)
After0	0.004878	-0.006286	0.01116
	(0.65)	(-1.35)	(0.98)
After1	0.02950***	-0.02050***	0.04999***
	(4.25)	(-5.26)	(5.17)
After2	-0.0006518	0.01901***	-0.01966*
	(-0.10)	(3.52)	(-1.78)
After0 × GOP	-0.02339***	0.01889***	-0.04228***
	(-3.13)	(4.89)	(-4.42)
After1 × GOP	-0.008569	0.001913	-0.01048
	(-1.17)	(0.48)	(-1.17)
After2 × GOP	-0.03749***	0.04605***	-0.08355***
	(-7.49)	(11.66)	(-11.31)
After0 × Dem	0.02101**	-0.03329***	0.05430***
	(2.13)	(-4.84)	(3.85)
After1 × Dem	0.004477	-0.002035	0.006512
	(0.41)	(-0.32)	(0.49)
After2 × Dem	0.07620***	-0.05083***	0.1270***
	(10.54)	(-8.56)	(11.30)
N	341669	341669	341669
adj. R^2	0.031	0.067	0.061

Table 1.7: Results without Big Tech Firms

This table reports the empirical results with OLS estimation of Equation (4) on the subsample excluding big tech firms involved in the "Trump Ban" after the Capitol riot on 01/06/2021. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.05398*	0.01344***	-0.06742**
	(-1.66)	(9.30)	(-2.07)
GOP	0.06892***	0.02160	0.05391*
	(3.13)	(1.31)	(1.94)
After0	0.01620***	-0.005799***	0.02200***
	(5.74)	(-3.12)	(4.97)
After1	0.01467***	-0.01344***	0.02811***
	(4.69)	(-7.10)	(6.18)
After2	0.01800***	-0.01025***	0.02825***
	(6.75)	(-6.09)	(6.82)
After0 × GOP	-0.02028***	0.01488***	-0.03516***
	(-8.20)	(10.83)	(-10.79)
After1 × GOP	-0.01235***	0.01717***	-0.02952***
	(-3.89)	(10.31)	(-7.37)
After2 × GOP	-0.02916***	0.02015***	-0.04931***
	(-14.19)	(16.32)	(-17.13)
After0 × Dem	0.02108***	-0.01945***	0.04052***
	(7.10)	(-10.07)	(10.03)
After1 × Dem	0.01133***	-0.01631***	0.02764***
	(2.80)	(-6.91)	(5.40)
After2 × Dem	0.02760***	-0.02103***	0.04863***
	(11.87)	(-14.59)	(15.81)
N	2054420	2054420	2054420
adj. R^2	0.080	0.032	0.054

Table 1.8: Results with Major Stock Indices

This table reports the baseline empirical results with OLS estimation of Equation (4) on the subsample major stock indices (SPY, QQQ, and DIA). The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.03288***	0.01192***	-0.04480***
	(-11.87)	(4.58)	(-10.35)
GOP	0.04462***	0.005856	0.03876***
	(8.07)	(1.42)	(4.79)
After0	0.01835***	0.002911	0.01544
	(3.05)	(0.54)	(1.45)
After1	0.04209***	-0.02945***	0.07154***
	(4.84)	(-5.32)	(5.60)
After2	0.04528***	-0.01840***	0.06369***
	(9.96)	(-5.29)	(8.71)
After0 × GOP	-0.03295***	0.01935***	-0.05230***
	(-7.36)	(5.95)	(-8.34)
After1 × GOP	-0.03920***	0.02969***	-0.06889***
	(-6.42)	(7.64)	(-8.36)
After2 × GOP	-0.04743***	0.03788***	-0.08531***
	(-12.91)	(14.70)	(-16.16)
After0 × Dem	0.05429***	-0.03970***	0.09398***
	(10.57)	(-8.39)	(11.50)
After1 × Dem	0.02743***	-0.02716***	0.05459***
	(3.65)	(-4.62)	(5.04)
After2 × Dem	0.04934***	-0.03350***	0.08284***
	(12.07)	(-9.08)	(13.08)
N	385610	385610	385610
adj. R^2	0.002	0.003	0.003

Table 1.9: Results without Indices or Big Tech Firms

This table reports the baseline empirical results with OLS estimation of Equation (4) on the subsample excluding major stock indices (SPY, QQQ, and DIA) and big tech firms involved in the 'Trump Ban' after the Capitol riot on 01/06/2021. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.04734	0.009466***	-0.05681*
	(-1.46)	(7.81)	(-1.74)
GOP	-0.03465	-0.02256***	-0.01209
	(-1.39)	(-3.15)	(-0.49)
After0	0.01534***	-0.007151***	0.02249***
	(4.92)	(-3.75)	(4.72)
After1	0.009357***	-0.01042***	0.01978***
	(2.77)	(-5.19)	(4.04)
After2	0.01225***	-0.008512***	0.02076***
	(4.02)	(-4.51)	(4.40)
After0 × GOP	-0.01548***	0.01168***	-0.02717***
	(-5.69)	(8.44)	(-7.80)
After1 × GOP	-0.006403*	0.01507***	-0.02147***
	(-1.74)	(8.20)	(-4.68)
After2 × GOP	-0.02574***	0.01477***	-0.04050***
	(-11.24)	(11.83)	(-13.14)
After0 × Dem	0.008190**	-0.01227***	0.02046***
	(2.47)	(-6.59)	(4.91)
After1 × Dem	0.002948	-0.01074***	0.01369**
	(0.63)	(-4.84)	(2.46)
After2 × Dem	0.01767***	-0.01559***	0.03326***
	(6.61)	(-11.18)	(10.05)
N	1668810	1668810	1668810
adj. R^2	0.089	0.023	0.051

Table 1.10: Results with $Beta\_Q = 1$

This table reports the baseline empirical results with OLS estimation of Equation (4) on the subsample of stocks in the first market beta quantile. Sample stocks are divided into three quantiles based on their market beta calculated by a 36-months regression of the Fama-French three-factor model prior to 09/2020. The first market beta quantile contains the stocks with the lowest market betas, while the third quantile contains the stocks with the highest market betas. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.04370	0.005406***	-0.04911
	(-1.35)	(4.04)	(-1.51)
GOP	-0.2227***	-0.02331***	-0.1994***
	(-9.55)	(-8.88)	(-8.41)
After0	0.01251***	-0.005743***	0.01826***
	(2.68)	(-3.25)	(3.09)
After1	-0.01298**	0.003237	-0.01621**
	(-2.44)	(1.49)	(-2.34)
After2	0.002408	-0.002209	0.004617
	(0.78)	(-1.30)	(1.08)
After0 × GOP	-0.01634***	0.01036***	-0.02670***
	(-3.27)	(5.04)	(-4.66)
After1 × GOP	0.01654***	0.005498**	0.01104
	(2.60)	(2.01)	(1.50)
After2 × GOP	-0.01510***	0.007372***	-0.02248***
	(-4.01)	(4.28)	(-5.00)
After0 × Dem	0.01196*	-0.004868*	0.01683**
	(1.93)	(-1.67)	(2.30)
After1 × Dem	0.01166	-0.007609**	0.01927**
	(1.56)	(-2.15)	(2.22)
After2 × Dem	0.005189	-0.009486***	0.01467***
	(1.17)	(-4.77)	(2.89)
N	407366	407366	407366
adj. R^2	0.104	0.011	0.073

Table 1.11: Results with $Beta\_Q = 2$

This table reports the baseline empirical results with OLS estimation of Equation (4) on the subsample of stocks in the second market beta quantile. Sample stocks are divided into three quantiles based on their market beta calculated by a 36-months regression of the Fama-French three-factor model prior to 09/2020. The first market beta quantile contains the stocks with the lowest market betas, while the third quantile contains the stocks with the highest market betas. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.02961***	0.01659***	-0.04620***
	(-4.49)	(8.94)	(-6.36)
GOP	-0.1236***	-0.02886***	-0.09476***
	(-8.38)	(-4.06)	(-6.67)
After0	0.01037***	-0.001441	0.01181**
	(3.09)	(-0.57)	(2.19)
After1	0.02297***	-0.02500***	0.04797***
	(5.27)	(-9.83)	(7.88)
After2	0.03180***	-0.01820***	0.05001***
	(11.71)	(-9.60)	(11.90)
After0 × GOP	-0.02323***	0.01655***	-0.03978***
	(-6.95)	(8.30)	(-9.32)
After1 × GOP	-0.02323***	0.02070***	-0.04393***
	(-5.53)	(9.46)	(-8.42)
After2 × GOP	-0.04006***	0.02708***	-0.06713***
	(-15.63)	(16.13)	(-19.02)
After0 × Dem	0.03569***	-0.02961***	0.06530***
	(8.72)	(-10.22)	(11.43)
After1 × Dem	0.01100*	-0.01538***	0.02638***
	(1.89)	(-3.99)	(3.35)
After2 × Dem	0.03625***	-0.02211***	0.05836***
	(11.81)	(-9.88)	(13.59)
N	874630	874630	874630
adj. R^2	0.055	0.033	0.043

Table 1.12: Results with $Beta\_Q = 3$

This table reports the baseline empirical results with OLS estimation of Equation (4) on the subsample of stocks in the third market beta quantile. Sample stocks are divided into three quantiles based on their market beta calculated by a 36-months regression of the Fama-French three-factor model prior to 09/2020. The first market beta quantile contains the stocks with the lowest market betas, while the third quantile contains the stocks with the highest market betas. The dependent variable is the sentiment measure at the user-stock-date level. The key independent variables are the interaction term between post-election time dummies and the partisanship indicator.
t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	(1) bull	(2) bear	(3) sentiment
Dem	-0.03123	0.009811	-0.04104
	(-1.09)	(1.37)	(-1.40)
GOP	-0.3730***	-0.07998***	-0.2930***
	(-17.24)	(-9.97)	(-12.48)
After0	0.01751***	-0.008417***	0.02593***
	(3.84)	(-2.95)	(3.64)
After1	0.02326***	-0.01429***	0.03755***
	(4.98)	(-5.04)	(5.49)
After2	0.01063**	-0.004453	0.01509**
	(2.37)	(-1.59)	(2.16)
After0 × GOP	-0.01432***	0.01330***	-0.02762***
	(-3.84)	(6.79)	(-5.61)
After1 × GOP	-0.01228***	0.01739***	-0.02968***
	(-2.71)	(6.38)	(-5.02)
After2 × GOP	-0.03113***	0.02620***	-0.05733***
	(-10.26)	(13.17)	(-13.16)
After0 × Dem	0.01013**	-0.01788***	0.02801***
	(2.17)	(-5.82)	(4.45)
After1 × Dem	0.007855	-0.01402***	0.02187***
	(1.17)	(-4.09)	(2.66)
After2 × Dem	0.03739***	-0.02889***	0.06628***
	(9.50)	(-12.03)	(12.66)
N	1071396	1071396	1071396
adj. R^2	0.058	0.017	0.033

Table 1.13: Summary Statistics

This tables reports the summary statistics for the variables used to estimate Equation (6). Please refer to Table 1.A.3 for the definition of these variables.

	Num of Obs	Mean	Std	Min	Max
Abnormal Turnover Rate	46232	-0.0001	0.0104	-0.1925	0.3733
Effective Percent Spread	46232	0.0004	0.0002	0.0000	0.0034
Abnormal Intraday Volatility	46232	0.0000	0.0000	-0.0000	0.0000
Partisan Disagreement	46232	0.1853	0.2221	0.0000	2.0000
Cumulative Abnormal Return (Last 5 Days)	46232	0.0005	0.0151	-0.0763	0.0778
Cumulative Abnormal Return (Last 25 Days)	46232	0.0003	0.0152	-0.0763	0.1238
Idiosyncratic Volatility (Last 5 Days)	46232	0.0008	0.0006	0.0000	0.0135
Overall Disagreement	46232	0.3161	0.2412	0.0000	1.4142

Table 1.14: Results with Abnormal Turnover Rate

This table reports the OLS estimation of Equation (6) with abnormal turnover rate as the dependent variable. The key independent variables are the interaction terms between post-election time dummies and Partisan Disagreement. Column (1) reports the results for the full sample. Columns (2)–(4) report the results for the subsamples of the market beta quantiles. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	Abnormal Turnover Rate			
	Full Sample	Beta_Q = 1	Beta_Q = 2	Beta_Q = 3
After0	0.003463***	0.002242***	0.001431	0.006119***
	(3.39)	(3.22)	(0.78)	(2.78)
After1	0.005644***	0.006104**	0.001351	0.008938***
	(3.97)	(2.21)	(0.76)	(3.38)
After2	0.002351***	0.002394***	0.0008361	0.003309*
	(2.60)	(3.61)	(0.47)	(1.80)
Partisan Disagreement	-0.005105***	-0.002945***	-0.004802***	-0.006738***
	(-12.70)	(-6.63)	(-8.45)	(-7.75)
After0 × Partisan Disagreement	0.003927***	0.0007425	0.002932***	0.006850***
	(5.56)	(1.18)	(3.67)	(4.08)
After1 × Partisan Disagreement	0.0007376	0.00009897	0.00007767	0.001830
	(1.06)	(0.13)	(0.10)	(1.09)
After2 × Partisan Disagreement	0.001696***	0.0003445	0.002151***	0.002344**
	(3.82)	(0.91)	(3.68)	(2.22)
Overall Disagreement	0.005940***	0.004155***	0.006149***	0.006820***
	(17.36)	(12.41)	(12.28)	(8.70)
N	46232	15067	15735	15262
adj. R^2	0.187	0.149	0.341	0.159

Table 1.15: Results with Effective Percent Spread-Equally Weighted

This table reports the OLS estimation of Equation (6) with equal-weighted effective percentage spread as the dependent variable. The key independent variables are the interaction terms between post-election time dummies and Partisan Disagreement. Column (1) reports the results for the full sample. Columns (2)–(4) report the results for the subsamples of the market beta quantiles.
t statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

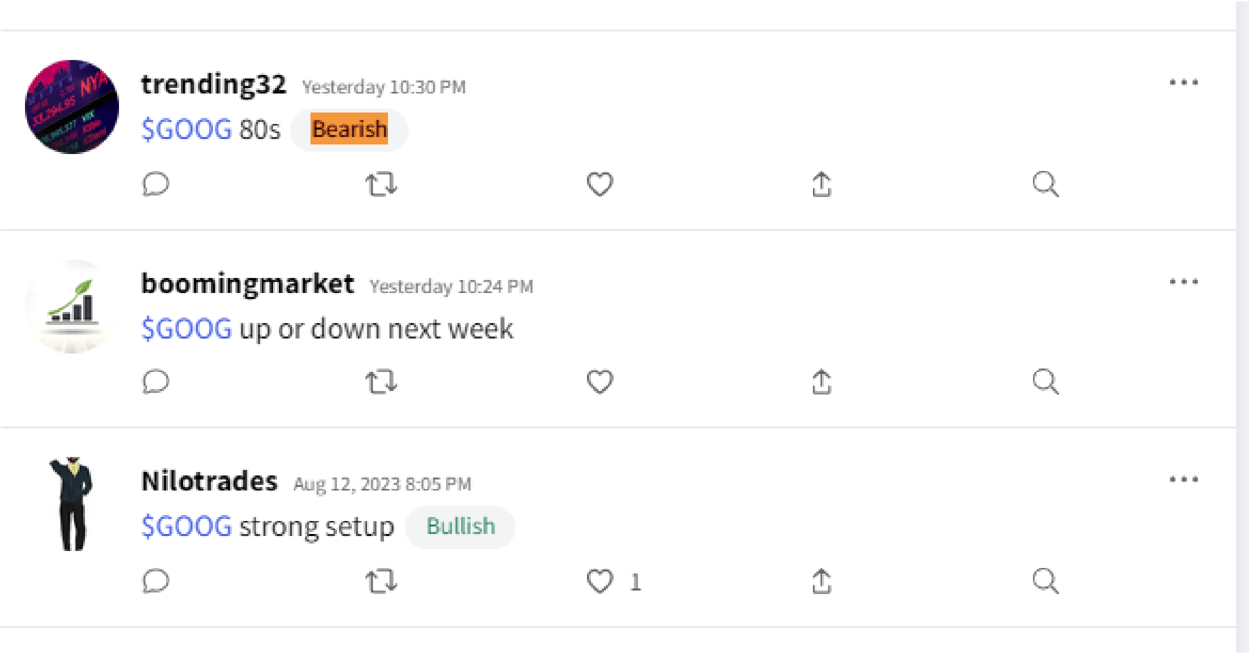
	Effective Percent Spread			
	Full Sample	Beta_Q = 1	Beta_Q = 2	Beta_Q = 3
After0	0.003799***	0.002166*	0.005648***	0.003730**
	(4.78)	(1.87)	(4.12)	(2.52)
After1	0.01058***	0.007955***	0.01153***	0.01216***
	(12.77)	(6.00)	(7.77)	(8.26)
After2	0.0006352	0.000002472	0.0004260	0.001406
	(0.78)	(0.00)	(0.35)	(0.80)
Partisan Disagreement	-0.001025***	-0.0002294	-0.002124***	-0.0003647
	(-2.66)	(-0.39)	(-3.74)	(-0.47)
After0 × Partisan Disagreement	-0.002522***	-0.002568***	-0.002664***	-0.002405**
	(-4.73)	(-2.94)	(-2.99)	(-2.42)
After1 × Partisan Disagreement	-0.002119***	-0.001034	-0.001713*	-0.003427***
	(-3.39)	(-1.11)	(-1.65)	(-2.75)
After2 × Partisan Disagreement	-0.002133***	-0.003066***	0.001000	-0.004309***
	(-4.32)	(-4.34)	(1.31)	(-4.31)
Overall Disagreement	0.002815***	0.002015***	0.003483***	0.002660***
	(8.02)	(4.11)	(6.40)	(3.61)
N	46232	15067	15735	15262
adj. R^2	0.822	0.827	0.825	0.798

Table 1.16: Results with Quote-Based Abnormal Intraday Volatility

This table reports the OLS estimation of Equation (6) with abnormal intraday volatility as the dependent variable. The key independent variables are the interaction terms between post-election time dummies and Partisan Disagreement. Column (1) reports the results for the full sample. Columns (2)–(4) report the results for the subsamples of the market beta quantiles. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	Abnormal Intraday Volaility			
	Full Sample	Beta_Q = 1	Beta_Q = 2	Beta_Q = 3
After0	-0.02607***	-0.01827***	-0.02372**	-0.03563***
	(-5.87)	(-4.38)	(-2.27)	(-4.90)
After1	0.009564*	0.004902	0.01619	0.009555*
	(1.96)	(0.54)	(1.52)	(1.75)
After2	-0.03254***	-0.01881***	-0.03332***	-0.04640*
	(-3.63)	(-3.59)	(-5.86)	(-1.79)
Partisan Disagreement	-0.03804***	-0.02568***	-0.04483***	-0.03978***
	(-9.32)	(-5.92)	(-6.01)	(-5.22)
After0 × Partisan Disagreement	0.003853	-0.004627	0.003664	0.01668
	(0.30)	(-0.52)	(0.12)	(1.24)
After1 × Partisan Disagreement	0.03191***	0.03367***	0.04078***	0.01943
	(4.70)	(4.07)	(3.12)	(1.54)
After2 × Partisan Disagreement	0.02598***	0.002643	0.05319***	0.02079*
	(4.56)	(0.43)	(5.23)	(1.78)
Overall Disagreement	0.03040***	0.02747***	0.03600***	0.02592***
	(7.24)	(5.39)	(4.91)	(2.96)
N	46232	15067	15735	15262
adj. R^2	0.057	0.113	0.036	0.081

Appendix



trending32 Yesterday 10:30 PM
$GOOG 80s Bearish

boomingmarket Yesterday 10:24 PM
$GOOG up or down next week

Nilotrades Aug 12, 2023 8:05 PM
$GOOG strong setup Bullish

Figure 1.A.1: Examples of Stocktwits Posts with 'GOOG'

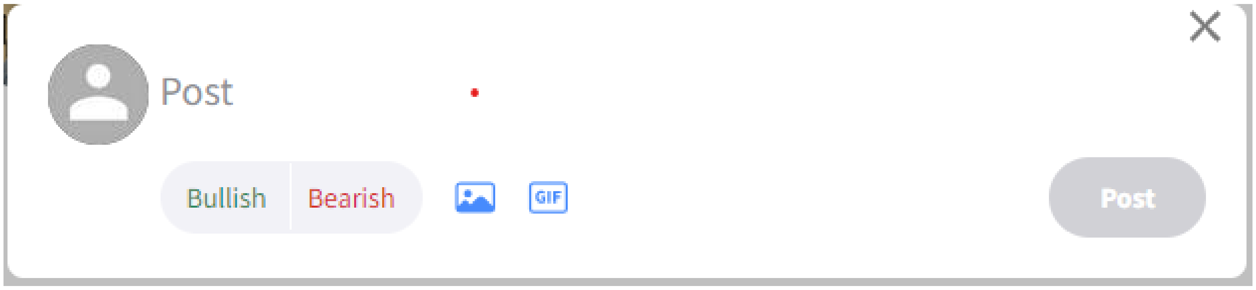


Figure 1.A.2: Stocktwits Post Option

Table 1.A.1: Examples of President Trump's Tweets about the Stock Market

This table reports examples of President Trump's tweets about the stock market during the 2020 presidential election.

Tweet Time	Message
Oct 20th 2020 - 10:02:31 AM EST	Stock Markets will hit new highs if President Trump wins. Tremendous growth like never before. If Biden wins, it's strangulation. Not good." @SteveKalayjian @Varneyco
Oct 12th 2020 - 12:54:15 PM EST	STOCK MARKET UP ANOTHER 300 POINTS - GREATEST LEADING INDICATOR OF THEM ALL!!! DON'T RUIN IT WITH SLEEPY JOE!!!
Oct 12th 2020 - 12:47:04 PM EST	The Economy is about ready to go through the roof. Stock Market ready to break ALL-TIME RECORD. 401k's incredible. New Jobs Record. Remember all of this when you VOTE. Sleepy Joe wants to quadruple your Taxes. Depression!!! Don't let it happen! #MAGA
Oct 12th 2020 - 10:16:29 AM EST	Stock Market Up Big. Do I get no credit for this? Never even mentioned by the Fake News. A New Record for Stocks and Jobs Growth. Remember, "it's the Economy Stupid". VOTE!!!
Aug 26th 2020 - 11:05:08 AM EST	Just In: Chinese State Media and Leaders of CHINA want Biden to win "the U.S. Election". If this happened (which it won't), China would own our Country, and our Record Setting Stock Markets would literally CRASH!
Aug 24th 2020 - 6:49:50 PM EST	Joe Biden has said he would lock down the Country again. That's crazy! We're having record job growth and a booming stock market, but Joe would end it all and close it all down. Ridiculous!

Table 1.A.2: List of Companies in the Sample that Issued ’Trump Ban’

This table reports the list of big tech firms involved in the ”Trump Ban” after the Capitol riot on 01/06/2021, as well as the specific actions that each firm took as part of the “Trump Ban".

Company Name	Ticker Symbole	Actions in the Trump Ban
Alphabet Inc.	GOOG	Google banned the pro-Trump social media platform, Parler, from Google Paly Store.
Amazon.com, Inc	AMZN	Amazon prohibited Parler, a pro-Trump social media platform, from accessing Amazon Web Services. Without Amazon Web Services, Parler was left without the infrastructure needed to continue operating.
Apple Inc.	AAPL	Apple banned the pro-Trump social media platform, Parler, from App Store.
Facebook, Inc	FB	Facebook suspended services from Trump for at least two years.
Twitter, Inc.	TWTR	Twitter issued a permanent ban on Trump. Trump’s Twitter account was wiped out from the platform.

Table 1.A.3: Description of Variable Construction

This table describes the construction of the variables used to estimate Equation (6).

Variable Name	Variable Construction Description
Abnormal Turnover Rate	Turnover rate minus average turnover rate in the past 20 days. Turnover rate is measured as the ratio of number of traded shares and number of shares outstanding.
Effective Percent Spread	Equal average of the effective percent spread, which is calculated by using the method introduced by Holden and Jacobsen (2014). The trade is classified as buyer or seller initiated by using the method in Lee and Ready (1991).
Abnormal Intraday Volatility	Intraday volatility minus average intraday volatility in the past 20 days. Intraday volatility is measured as the standard deviation of quote price at daily stock level.
Partisan Disagreement	The absolute value of difference between average aggregate sentiment of Republican and Democrat investors, at daily stock level.
Cumulative Abnormal Return Last 5 Days	Cumulative sum of abnormal return in the past 5 days. Abnormal return is calculated as return minus expected return. Expected return is calculated with Fama-French 3-factor model, where the factor loading is estimated in a 30-days rolling regression.
Cumulative Abnormal Return Previous 25 Days	Cumulative sum of abnormal return in the past 25 days. Abnormal return is calculated as return minus expected return. Expected return is calculated with Fama-French 3-factor model, where the factor loading is estimated in a 30-days rolling regression.
Recent Idiosyncratic Volatility Last 5 Days	Standard deviation of abnormal return in the past 5 days. Abnormal return is calculated as return minus expected return. Expected return is calculated with Fama-French 3-factor model, where the factor loading is estimated in a 30-days rolling regression.
Overall Disagreement	Standard deviation of Stocktwits users’ aggregate sentiment at daily stock level.

Table 1.A.4: Results with Effective Percent Spread-Dollar Weighted

This table reports the OLS estimation of Equation (6) with dollar-weighted effective percentage spread as the dependent variable. The key independent variables are the interaction terms between post-election time dummies and Partisan Disagreement. Column (1) reports the results for the full sample. Columns (2)–(4) report the results for the subsamples of the market beta quantiles. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	Effective Spread			
	Full Sample	Beta_Q = 1	Beta_Q = 2	Beta_Q = 3
After0	0.0001932***	0.0001641***	0.0002132***	0.0002040**
	(4.43)	(3.29)	(2.82)	(2.21)
After1	0.0003300***	0.0003325***	0.0003882***	0.0002861***
	(7.37)	(4.22)	(4.58)	(4.03)
After2	0.00001416	0.000006098	0.00005480	-0.00001633
	(0.59)	(0.28)	(1.02)	(-0.39)
Partisan Disagreement	6.670e-07	-0.00002320	0.00004431	-0.00001876
	(0.03)	(-0.76)	(1.29)	(-0.65)
After0 × Partisan Disagreement	-0.00002296	-0.00001038	-0.00004213	-0.00001014
	(-0.68)	(-0.25)	(-0.62)	(-0.20)
After1 × Partisan Disagreement	-0.00006702*	-0.00008274	-0.00002283	-0.0001078**
	(-1.93)	(-1.21)	(-0.44)	(-2.13)
After2 × Partisan Disagreement	-0.00004108*	-0.00003183	-0.00006373	-0.00002018
	(-1.68)	(-0.82)	(-1.61)	(-0.55)
Overall Disagreement	0.00002802*	0.00004473	0.000001172	0.00003137
	(1.83)	(1.47)	(0.05)	(1.36)
N	46232	15067	15735	15262
adj. R^2	0.217	0.183	0.219	0.228

Table 1.A.5: Results with Effective Percent Spread-Share Weighted

This table reports the OLS estimation of Equation (6) with share weighted effective percentage spread as the dependent variable. The key independent variables are the interaction terms between post-election time dummies and Partisan Disagreement. Column (1) reports the results with full sample. Column (2) - (4) reports the results with subsamples of market beta quantiles. t-statistics in parentheses. * p<0.1, ** p<0.05, *** p<0.01

	Effective Spread			
	Full Sample	Beta_Q = 1	Beta_Q = 2	Beta_Q = 3
After0	0.0001911***	0.0001627***	0.0002105***	0.0002016**
	(4.43)	(3.28)	(2.83)	(2.21)
After1	0.0003364***	0.0003369***	0.0003970***	0.0002923***
	(7.37)	(4.20)	(4.59)	(4.03)
After2	0.00001445	0.000006427	0.00005554	-0.00001653
	(0.60)	(0.30)	(1.02)	(-0.40)
Partisan Disagreement	5.055e-07	-0.00002331	0.00004377	-0.00001885
	(0.02)	(-0.76)	(1.28)	(-0.65)
After0 × Partisan Disagreement	-0.00002327	-0.00001025	-0.00004261	-0.00001064
	(-0.69)	(-0.25)	(-0.64)	(-0.21)
After1 × Partisan Disagreement	-0.00006708*	-0.00008297	-0.00002406	-0.0001065**
	(-1.93)	(-1.21)	(-0.47)	(-2.10)
After2 × Partisan Disagreement	-0.00004075*	-0.00003187	-0.00006390	-0.00001893
	(-1.66)	(-0.82)	(-1.61)	(-0.52)
Overall Disagreement	0.00002780*	0.00004493	9.082e-07	0.00003100
	(1.81)	(1.47)	(0.04)	(1.34)
N	46232	15067	15735	15262
adj. R^2	0.216	0.183	0.217	0.226

2 Investment Decision During the Pandemic: Evidence from Net Flows of Locally Headquartered Mutual Funds

2.1 Introduction

The existing literature (Guiso et al., 2018; Kuhnen & Knuston, 2011; Tulbadi,2020; Wang & Young, 2020) has focused on how changes in external pressure shape investment decisions. Since the outbreak of the Covid-19 pandemic in early 2020, residents in the US have experienced drastic shifts in social norms and living styles in response to the spread of the virus. The emergence of this highly infectious virus leads to heightened public health concerns, which are accompanied by worsening economic conditions induced by the public lockdown. As a relatively exogenous economic and emotional shock to investors, the Covid pandemic provides a good setting for studying changes in economic behaviors under external pressure. In this setting of the Covid pandemic, one interesting question to be answered is how the local spread of viruses leads to changes in investment decisions. Does the spread of Covid lead to changes in risk-taking behavior? If so, how do Covid-induced changes in risk-taking behavior lead to changes in the investment decision? This study attempts to approach these questions by empirically examining the impact of state-level Covid spreads on the net flows of locally headquartered mutual funds.

Home bias literature (Coval & Moskowitz, 2002; Zhu (2002)) documents investors' preference toward financial assets with a more local presence. Following Tulbadi (2020), a fundamental assumption in the rationale behind the identification strategy is that local investors made up a portion of the demand for locally headquartered mutual funds due to home bias in investment preference. Since Covid spread is associated with heightened public health concerns during the pandemic, I hypothesized that local Covid spread induces visceral feelings, such as fear and anxiety, from local investors, leading to increased risk aversion and lower investment participation in the financial market. The central hypothesis to be tested in this study is that local Covid spread leads to higher fund outflows from locally headquartered mutual funds, as local investors demand liquidity in response to the pandemic.

I used the New York Times (NYT) Covid database and the CRSP Survivor-Bias-Free US Mutual Fund database to examine this hypothesis empirically. The baseline empirical results indicate that state-level Covid spreads have a reducing effect on the net flow in locally headquartered mutual funds, which is consistent with the hypothesis that local Covid spreads reduce investment participation. Further, the baseline results indicate that Covid cases and deaths impose heterogeneous effects on the liquidity demand from local mutual funds, as the former induce more pronounced outflows than the latter. One possible explanation behind this heterogeneous effect on fund outflows is the difference in public concerns over Covid cases and deaths.

In addition to the baseline results, I conducted several further empirical analyses to help interpret the economic mechanisms behind the Covid-induced mutual fund outflows. The first key question in interpreting the baseline results is regarding the heterogenous effect of local Covid spread on liquidity demand across different pandemic stages. Specifically, did the Covid spread induce more liquidity demand in the early stage of the pandemic when public concerns over the virus were at their peak? Did local Covid spread continue driving mutual fund outflows as public concerns subsided after the early stage of the pandemic? To answer these questions, I empirically examine the baseline specification in the subsamples of the Covid crisis period (02/2020 –04/2020) and the post-Covid crisis period (05/2020 –12/ 2021), respectively. The empirical results from the subsample regressions indicate that Covid-induced mutual fund outflows were much more pronounced during the early Covid crisis period. In addition, even after the Covid crisis period, state-level Covid spread continues to reduce the net flows of locally headquartered mutual funds. However, the effect is much less pronounced when compared to the Covid crisis subsample.

Further, I empirically examine the potentially heterogeneous effect of local Covid spread on investment decisions across different types of investors. An essential question is whether Covid spread triggered more liquidity demand from retail investors than institutional investors. When compared to institutional investors, retail investors are more prone to the pressure caused by the Covid pandemic, either economically or mentally. Their investment decisions are more likely to be biased by visceral feelings induced by the virus. Therefore, I hypothesize that Covid spread leads to more liquidity demand from retail fund share classes. Consistent with this hypothesis, the empirical results show that Covid-induced mutual fund outflows are more pronounced for retail fund share classes.

Another critical aspect to be examined to interpret the baseline empirical results is the interplay between the Covid pandemic and heightened risk aversion. One interesting question to be answered is whether investors experience heightened risk aversion during the pandemic and therefore demand more liquidity from risky investments. To examine this economic channel, I empirically examine the heterogenous effect of local Covid spread on the net flows of mutual funds with different levels of risk. Under this channel of explanation, the Covid-induced fund outflows are expected to be more pronounced for share classes associated with higher fund-level risk. Specifically, I examine this question from the fund investment style, return volatility, and fund market beta perspective. Consistent with the risk aversion channel of explanation, the empirical findings show that Covid-induced mutual fund outflows are less pronounced for funds with safer investment objectives while more pronounced for funds that concentrated in high volatility and market beta quantiles. Overall, the empirical findings imply that heightened risk aversion during the pandemic plays a crucial role in shaping liquidity demand.

Finally, in addition to the visceral response to the emergence of the virus, a competing explanation behind the Covid-induced mutual fund outflows is the worsening local economic conditions associated with the pandemic. By inducing an economic shutdown, the local Covid spread imposes a negative shock on the economic conditions of residents. Such a negative income or wealth shock may force local investors to demand liquidity from mutual fund investment. To help disentangle these two channels of explanation, I further included controls of local economic conditions in the empirical specification. In contrast to the negative economic shock channel of explanation, controlling local economic conditions does not significantly alter the coefficients on Covid spread for all specifications, which lends more support to the notion that visceral feelings are a more plausible driving force behind Covid-induced mutual fund outflows.

To my knowledge, this study is the first to examine the impact of the Covid spread on the liquidity demand from locally headquartered mutual funds. The results emphasize that Covid spread leads to higher liquidity demand from local mutual funds through the channel of heightened risk aversion. A visceral response seems a more plausible explanation since controlling economic conditions does not alter the results. This study contributes to several strands of literature. Firstly, the findings in this study add to the new strand of emerging studies focusing on the economic impact of the Covid pandemic on the financial market. One strand of literature focuses on the impact of the Covid crash on the financial market at the onset of the pandemic, particularly in terms of liquidity demand and asset pricing. The onset of the pandemic in early 2020 led to heightened market sell off in the financial market as investors became pessimistic about the economic impact of the pandemic. Li et al. (2021) documented that during the Covid crisis in March 2020, liquidity restrictions, such as gates and liquidity fees by monetary market funds, further exacerbated fund outflows. Haddad et al. (2021) documented that during the Covid collapse period, investment-grade corporate bonds were traded at a discount to credit default swaps, and investment-grade ETFs were traded at a discount to their underlying bonds, which can be explained by liquidity demand from bonds investors such as mutual funds. Kargar et al. (2021) found that the onset of the Covid crisis led to a significant increase in corporate bond trading costs, which were partially subdued after Fed intervention. Fahlenbrach et al. (2021) documented that firms with higher financial flexibility experienced smaller price drops during the Covid collapse, especially for firms with higher Covid exposure. Pastor and Vorsatz (2020) document that investors are demanding funds with higher sustainability scores during the Covid crisis period. In addition to the findings of the existing studies, this study documents that local Covid spread leads to increased liquidity demand from locally headquartered mutual funds. Additionally, the empirical results indicate the local Covid spread drives fund outflows even after the initial stage of the pandemic.

Another strand of the Covid literature focuses on the interplay between sentiment induced by the pandemic and financial or risk-taking decisions. By using county-level data on Covid spread and bank branch deposits, Levine et al. (2021) documented that higher county-level Covid infection rate leads to higher deposit growth and lower deposit rates at bank branches, which can be explained by pandemic-induced anxiety about job and income losses. Additionally, further studies have indicated that demographic groups with heterogeneous views on the virus may react differently to the spread of the virus in terms of risk-taking behavior and financial decision-making. Barrios and Hochberg (2021) documented that given Covid cases or deaths, individuals in counties with a higher Trump vote in 2016 paid less attention to Covid-related information. The authors argued that partisan opinions on viruses play a role in shaping risk aversion during the pandemic. Vorsatz (2021) finds that non-partisan mutual fund teams have higher fund flow when compared to partisan mutual fund teams, which the investment preference of partisan clients can explain.

This study also adds to the literature strand that focuses on visceral emotions' impact on risk-taking behaviors. Wang and Young (2020) documented that terrorists attack induce fund flows from equity mutual funds into government bond funds by shifting risk aversion among retail investors. Kuhnen and Knuston (2011) found that a positive emotional state induces people to engage more in risk-taking activities, and a negative emotional state does the opposite. Using mass shootings as a natural experiment, Borgan and Yonker (2021) documented that mutual funds near the mass shooting reduce the risk-taking behavior of nearby mutual funds, which the visceral response can explain. Bharath and Cho (2023) documented that households experiencing natural disasters tend to allocate fewer assets in the stock market. The results of this study imply that Covid spreads liquidity demand from mutual funds, which can be potentially explained by the visceral response to the spread of the virus.

The rest of the paper is organized as follows. Section 2 describes the data source and construction of the sample. Section 3 describes the baseline empirical strategies. Section 4 presents and discusses the baseline results. Section 5 presents the analysis of the hydrogenous effect across different stages of the pandemic. Section 6 presents the analysis of the heterogeneous Covid-induced outflows on retail and institutional fund share class. Section 7 presents the analysis of the risk aversion channel of explanation for the baseline results. Section 8 attempts to disentangle the effect of visceral response and worsening economic conditions on fund outflows. Section 9 draws the final analysis and conclusion.

2.2 Data and Sample Construction

To examine the hypothesis that local Covid spread leads to mutual fund outflow, I constructed a data set that merges data from several sources. In this section, I will discuss the data source and sample construction.

2.2.1 Mutual Fund Data Source

The primary source of mutual fund data is the CRSP mutual fund database. I collected data from the CRSP mutual fund database on monthly mutual fund share class return, total net asset (TNA), and net asset per share. Additionally, from the CRSP mutual fund summary database, I collected data on quarterly expense ratio and turnover rate, Lipper fund investment styles, fund founding year, and fund type, which indicates whether the fund class share is retail or institutional. From the CRSP mutual fund summary, I gather information on the location of the management company, particularly the state of headquartered, which is to be matched with state-level Covid spread data. The mutual fund sample covers the period of the Covid pandemic from February 2020 to December 2021.

2.2.2 State Level Covid Spread Data Source

The key independent variable examined in this study is local Covid spread. The data of state-level daily cumulative Covid cases and deaths are collected from the Covid dataset provided by the New York Times (NYT). The NYT Covid dataset is comprehensive and informative, as it provides daily updates of Covid cases and cases for each state. The dataset provides information on cumulative Covid cases and deaths reported for each state-date observation in the sample. Specifically, the dataset starts with the first Covid case in the United States on 1/21/2020 in Washington. There are missing records for state-date at the early stage of the pandemic, namely 1/2020 – 03/2020, mainly because states report the first Covid cases or deaths at various time horizons during this period. For example, while the first Covid case appeared in January for Washington state, it was until mid-March that all 50 states, including the District of Columbia, reported their first Covid cases. For all missing state-month records of Covid cases or deaths in early 2020, I confirm that the missing values are all before the month when the first Covid cases were reported in the corresponding state. Therefore, I filled the sample by replacing these missing records with zero. To measure monthly Covid cases or deaths for each state, I computed the difference in cumulative state-level Covid cases or deaths between the end dates of consecutive months. The dataset of Covid records consists of 50 states, including the District of Columbia, and covers the Covid pandemic period from January 2020 to November 2021.

2.2.3 Other Data Sources

I collected data from several other datasets to construct additional control variables in the study. I collected monthly data on Fama-French factors from Kenneth French's website, including market excessive returns, size factor returns, value factor returns, momentum factor returns, and risk-free rates. I collected data on the state-level projected population in 2019 from the Census website. Data on monthly state-level unemployment rates were collected from the Bureau of Labor Statistics (BLS) website. Additionally, I gather data on the state-level Google search volume index of “Unemployment,” “Job Loss,” and “Income Loss” from Google Search Volume API to be included as control variables in some of the specifications.

2.2.4 Variable Construction

The key dependent variable in this study is the monthly mutual fund share class net flow, which is constructed with data from the CRSP dataset. To construct measures of monthly mutual fund share class net flow, I use monthly individual fund share class level returns and TNA as the basis of my calculation following academic standards. There are sufficient variations among the net flows of fund share class within one fund, and therefore double-counting is of less concern in this design. For each fund-month combination, I computed monthly net flow as the percentage change of TNA into the fund between consecutive month ends. The calculation can be demonstrated as the following equation.:

$$NetFlowi,t = \frac{TNA_{i,t} - TNA_{i,t-1} \times (1 + Return_{i,t})}{TNA_{i,t-1}} \tag{7}$$

Where $TNA_{i,t}$ refers to the net total asset of fund i at the end of month t, and $Return_{i,t}$ refers to the monthly return of fund i at month t. Fund net flows are winsorized at the 99% level. I dropped all fund observations with missing returns or net assets from the sample.

State-level Covid cases or deaths are the independent variables of key interest when examining the impact of Covid spread on local mutual fund net flows. As mentioned previously, monthly new Covid cases or deaths are computed as the difference between cumulative Covid cases and deaths by the end date of consecutive months. To remove the possible skewness of state-level Covid cases or deaths imposed by the state population, I further scale the new Covid cases or deaths by the 2019 projected state population. The measurements are constructed as a logarithm of one plus monthly Covid cases or deaths per 100,000 people in each state to control for the potential issue of skewness in the variable, as well as the state-level zero Covid cases or deaths at early stages of the pandemic. I also include monthly cumulative Covid cases and deaths scaled by population as alternative measures of Covid spread in the specification.

Similarly, these two variables are constructed as logarithms of one plus monthly state-level cumulative Covid cases or deaths scaled by state population. Thus, I have four measures of state-level Covid spread to be separately examined as the key independent variable:

- log of monthly Covid cases per capita
- log of monthly Covid deaths per capita
- log of cumulative Covid cases per capita
- log of cumulative deaths per capita

In the empirical specification, they are represented by Log Monthly Covid Cases, Log Monthly Covid Deaths, Log Total Covid Cases, and Log Total Covid Deaths.

I included other individual fund share class-level characteristics as control variables in the specification. I control for lagged TNA and net assets per share in the empirical specification. The lagged values of fund TNA and net assets per share are monthly values prior to the monthly observation of net flow. To remove skewness in TNA and net asset per share, I take the log of these two variables. In the specification, I also include a one-quarter lagged expense ratio and turnover rate as additional control variables. I further included the log of fund age, which is calculated as the difference between the year of observation and the year when the fund was first offered.

A key concern in the empirical specification is the potential endogeneity caused by the interplay between Covid spread and fund performance. As documented by existing studies (Barber et al., (2016); Cashman et al., (2012); Ippolito, 1992; Ivkovic and Weisbenner, (2009;), Surri and Tufano, (1998)), investors' responses to fund performance, and there exists a robust positive relationship between fund performance and net flows. Therefore, a key concern in interpreting the empirical results with local Covid spreads is whether investors are liquidating their assets in response to worsening fund performance rather than the impact of Covid on investors themselves. It is possible that state-level Covid may be correlated with individual fund performance by damaging the local economy by inducing economic lockdowns and shortages in the labor market. This possible association between fund performance and local economic conditions can be channeled from mutual funds' holdings in local firms. Existing studies in the home bias literature (Coval & Moskowitz (2002); Craft (2006); Kang and & Sultz (1997), Ke et al.,(2009), Coval and Moskowitz(2002)) have documented that mutual funds tend to allocate more of their capital into firms with a local presence. Coval and Moskowitz (2002) found that US domestic mutual fund managers put a higher weight on locally headquartered firms, which can be explained by information asymmetry between local and nonlocal investors. Covid spread affected local economics and, therefore, the operation condition of local firms by deteriorating local economic conditions. In this scenario, this detrimental effect on the local economy may further transmit into the performance of local mutual funds through their holdings of firms with a local presence. To help reduce the endogeneity concerns of fund performance on the Covid-induced mutual fund outflows, I include a one-month lagged fund share class return in the specification.

For similar reasons, controls of fund-level risk are included in the empirical specification. For each fund-month pair, I computed the monthly fund return volatility as the standard deviation of the fund daily return in the month; then, I took the log of one-month lagged volatility in the specification as a control variable. In the specification, I also included individual fund share class market beta as an additional control for fund-level risk. Following Wang and Young (2020), for each fund share class month pair, I computed the fund's market beta by running a rolling 36-month regression of the Fama-French 3-factor model before the current month. As alternative measures of market beta in some models, I calculate the Fama-French 4 -factor model market beta and CAPM beta by running a rolling 36-month regression.

2.2.5 Summary Statistics

Table 2.1 presents the summary statistics for fund-level variables. The sample has a total number of 457949 monthly fund share class observations. Monthly fund net flow has a mean of -0.000739 and a standard deviation of 0.0907. Table 2.2 presents the summary statistics for the state-level variables. There are a total number of 1150 state-month combinations in the sample, which covers 50 states during the sample period from 1/2021 to 11/2021. The key independent variables are monthly state-level Covid cases and deaths per capita. Monthly cases or deaths per 100,000 people at the state level mean 635.89 and 9.67, and standard deviation of 677.65 and 11.1456, respectively. The statistics suggest significant skewness exists in monthly Covid cases or deaths even after scaling for the state-level population. One reason could be the uneven distribution of Covid cases across different pandemic stages. At the early stage, newly emerged cases or deaths may be relatively low, while the number of infections increases exponentially as the pandemic continues into a later stage. The log of one plus monthly Covid cases and deaths per capita have means of 5.49 and 1.88 and standard deviation of 2.038 and 1.0276, respectively, which indicates that taking log reduces the skewness in the key independent variables. Similarly, taking log significantly reduced the skewness in the variables for cumulative cases or deaths per capita, according to the summary statistics.

Table 2.2 also presents other state-level variables, including the unemployment rate, Google search indexes for "job loss," "income loss," and "unemployment." During the sample period, the unemployment rate has a mean of 6.66%, a standard deviation of 0.0334, and a maximum of 27.5%. The Google search volume index of "income loss," "job loss," and "unemployment" have means of 29.65, 13.20, and 34.64, and standard deviation of 21.30, 10.40, and 23.74. Notably, the maxima of these Google search volume indexes are 100, 75, and 100, which implies that economic hardship is a significant concern during the pandemic. Hence, in addition to a visceral response to the spread of the virus, economic hardship can be another potential driving force behind the Covid-induced fund outflows.

2.3 Empirical Specification

To examine the impact of Covid spread on local mutual fund net flows, I use the following baseline regression model:

$$NetFlows_{i,s,t} = \beta_0 + \beta_1 \times Covid_{s,t-1} + \gamma \times X_{i,t-1} + \alpha_s + \alpha_t + \epsilon_{i,s,t} \tag{8}$$

Where i, s, t, index mutual fund share class, state, year-month, respectively. The dependent variable $NetFlow_{i,s,t}$ represents the monthly net flows at fund share class i, state s, and month t. The key independent variables $Covid_{s,t-1}$ refer to the state-level Covid spread measures at state s in month t - 1. The Covid spread measures can be log of monthly cases per capita, a log of monthly deaths per capita, log of cumulative cases per capita, or log of cumulative deaths per capita. $X_{i,t-1}$ represents other control variables, including TNA, TNA per share, log of volatility, expense ratio, turnover ratio, monthly return, monthly market return, fund age, and fund market beta. To control for autocorrelation in fund net flows, I include one - and two-month lagged fund net flows in the control variables.

Additionally, I include state and year-month fixed effects in the specification, which are represented by α_s and α_t, respectively. State fixed effects are included to control for state-level invariants that potentially correlate with both the mutual fund flow and local Covid spread. For example, existing studies (Borris & Hochberg, 2021; Gollwitzer et al., (2020;), Khalatabari-Soltani, 2020; Rabin & and Dutra, (2020), Borris and Hochberg (2021), Khalatabari-Soltani (2020)) document that demographic characteristics such as political propensity, education, and religious or cultural beliefs are correlated with the Covid infection rate. Such characteristics (Cole & Shastry, 2009; Ke, (2019), Cole and Shastry (2009)) may determine financial decisions across demographic groups. These unobservables vary across states due to differences in demographic distributions and may potentially confound with the effect of Covid spread on mutual fund flows.

More importantly, including the year-month fixed effect is essential in interpreting the empirical results since the local Covid spread is likely to be correlated with time-invariant unobservables during the pandemic. Without controlling for aggregate level time-invariants, it is hard to interpret the empirical results of the specification due to the difficulty of discerning whether the results are driven by Covid spread at local level or time-invariant confounders. Specifically, the local level of Covid spread is correlated with the pandemic stage at the national level, which essentially determines the overall economic and financial market conditions and, therefore, shapes financial decisions and liquidity demand during the pandemic. Covid spread may also be correlated with other time-shifting national-level unobservables such as policy shifts or public opinion that may further determine investment decisions. For example, the release of Federal relief checks and changes in Fed monetary policy during the pandemic may shift investment decisions by alleviating the worsening economic conditions and boosting investor confidence. National-level political sentiment, which likely correlates with the pandemic in the election year, may potentially confound the effect of Covid on investment decisions. By including year-month fixed effects in the specification, endogeneity concerns with these time-varying aggregate-level unobservables are reduced to some extent, and the coefficients on local Covid spread measures are easier to interpret.

β_1 measures the coefficient on the main independent variable, Covid spread. Under my hypothesis, local Covid spread may change risk attitudes among investors, either through worsening economic conditions or inducing visceral responses from local investors. Changes in risk attitude by Covid spread may further lead to lower financial participation and, therefore, outflows from locally headquartered mutual funds. Therefore, I expect β_1 to be negative and statistically significant under this hypothesis. γ is a vector of coefficients on the other control variables. β_0 and $\epsilon_{i,s,t}$ refer to the constant intercept and error term, respectively. I estimate the empirical model by using ordinary least squares (OLS) and report the standard errors two-way clustered at month and fund level to control for within-group correlation of error terms.

2.4 Empirical Results

2.4.1 Baseline Results

In this section, I present the empirical results estimated with the model specified in Equation (2). Table 2.3 displays the baseline empirical results. Columns (1) – (4) display the empirical results with four measures of state-level Covid spread. For all measures of state-level Covid spread except for cumulative Covid deaths per capita, the coefficients are negative and statistically significant, which is consistent with the hypothesis that local Covid spread induces outflows from mutual funds. The coefficients are -0.001634, -0.008038, and -0.001751 log of monthly Covid cases per capita, log of monthly Covid deaths per capita, and log of cumulative Covid cases per capita, respectively. The coefficient on the log of cumulative Covid deaths per capita is negative but statistically insignificant. For the log of monthly Covid cases per capita, one standard deviation (2.04) increase in the independent variable leads to a decrease of 0.33 percent in mutual fund net flow ($2.04 \times 0.001634 = 0.0033$). Similarly, a one standard deviation increases in the log of cumulative Covid cases per capita leads to a decrease of 0.49 percent in mutual fund net flow ($2.78 \times 0.001751 = 0.0049$). However, for a log of monthly Covid deaths per capita, the magnitude is relatively smaller. One standard deviation (1.0276) increase in the log of monthly Covid deaths per capita leads only to a decrease of 0.08 percent in mutual fund net flows ($1.0276 \times 0.008038 = 0.0008$).

The results are generally consistent with the hypothesis that state-level Covid spread induces outflow from local mutual funds, and the effects are heterogeneous across different measures of Covid spread. According to the magnitude of the coefficients, Covid cases have a much more significant economic impact on mutual fund net flows when compared to Covid deaths. One potential explanation behind the heterogeneous effect of Covid cases and deaths on fund flow is that Covid cases have imposed more public concerns over the virus since the subsequent infection rate in the local area may be more correlated with newly emerged or cumulative cases rather than deaths. As such, investors may react more pronounced to liquidating their assets from mutual funds in response to Covid cases rather than deaths. Additionally, among the two measurements of Covid deaths, only monthly new Covid deaths per capita have a statistically significant impact on mutual fund net flows. One possible explanation is that newly emerged deaths may have a more salient effect on the mental status of investors and therefore trigger a pronounced reaction in asset liquidation.

2.5 Further Analysis: Covid Crash Period

To help understand the baseline results, I conducted further analysis to examine the economic mechanisms behind the Covid-induced fund outflows. The first question is whether Covid-induced mutual fund outflows are mainly concentrated in the Covid crisis period, specifically during the first four months of 2021. During this early stage of the pandemic, Covid-19, as a newly emerging airborne deadly virus, triggered enormous fear and anxiety in the public, who had limited knowledge of the prospect of the pandemic. At the same time, the financial market overreacted to the spread of Covid in the United States, leading to a massive sell-off followed by a historic dip in major stock market indexes. Subsequently, the economic shutdown followed by the emergence of the virus led to a heightened unemployment rate and economic loss. After the Covid crash period, the public fear of the virus gradually subsided, and the economic condition and financial market rebounded from the bottom.

Several existing studies (Barrios and Hochberg, (2021;), Levine et al., (2021)) focus on the Covid crash period to examine the economic impact of the pandemic. Indeed, the economic impact of the pandemic may concentrate in this early stage, when heightened public concern over the virus was mixed with economic shutdown and financial market crash. As the public uncertainty evolves around the prospect of the pandemic, concerns over Covid spread peaked during the early stage. As such, heightened external pressure may press investors to liquidate more of their financial assets during this period. Thus, to understand the baseline empirical results, a natural question to ask is whether the Covid-induced mutual fund outflows concentrated during the Covid crisis period.

To examine this question empirically, I divide the sample into subsamples based on the stage of the pandemic, precisely the Covid crash period and the post-crash period. I run the baseline regression separately on the subsamples to examine the heterogenous effect of Covid spread on mutual fund net flows. Since the concerns over the virus are more concentrated during the early stage, the reducing effect of Covid spread on mutual fund flows is more pronounced during the Covid crash period. Therefore, I expect the coefficients to be more pronounced in terms of magnitude and statistical significance for the subsample of the Covid crash period. Another uncertain question is whether the Covid-induced fund outflows are solely driven by the Covid crash period when investors sell off their financial assets in response to concerns over the virus and its economic impact. If the baseline results are solely driven by the sell-off during the Covid crash period, then the coefficients are expected to be significant only for the Covid crash subsample. On the other hand, if the Covid spread continues driving fund outflows after the crash period, then I expect the coefficients on Covid spread to remain negative and statistically significant in the subsample of the post-crash period.

In Table 2.4, panels A and B display the during and after the Covid crash period, respectively. As shown by Panel A, in the subsample of the Covid crash period, the coefficients on all four measures of Covid spread are negative and statistically significant. The coefficients on Covid cases measures also remain negative and significant in the subsample of the post-Covid crash period, which implies that the local Covid spread continues driving fund outflows even after public fear of the virus gradually subsided from its peak after the early stage of the pandemic. Furthermore, in terms of the magnitude of the coefficients, the results indicate that Covid spread imposes a more economically significant impact on mutual fund outflows during the crash period.

For example, the coefficients are -0.001187 and -0.003161 on monthly new Covid cases during the crisis and post-crisis periods, implying that the impact during the Covid crash period is 2.6 times that in the subsequent period of the pandemic. A similar difference in magnitude can be observed for cumulative Covid cases. Notably, for measures of Covid deaths, the coefficients are statistically significant only for the subsample of the Covid crash period, which lends some support to the notion that public fear over the virus is driving fund outflows. Covid deaths may impose a significant level of public fear during the early stage of the pandemic, leading to the liquidation of mutual fund shares. However, the empirical results indicate that the concern over Covid deaths subsides during the post-crash period, which may result from public learning about the low mortality of the virus.

2.6 Further Analysis: Retail Fund

Investors may have experienced different hardships during the pandemic, hold different views of the virus and its economic impact, and therefore react differently in their financial decisions to the local spread of Covid. Hence, the reducing effect of local Covid spread on net flow may be heterogeneous across mutual funds with different investment objectives and clientless targets. One interesting question is whether the pandemic imposes a heterogenous effect on retail and institutional investors since the two types of investors may react differently to the spread of viruses when making financial decisions. In this section, I empirically examine the effect of Covid spread on liquidity demand across heterogeneous investor types.

Compared to institutional investors, retail investors are generally more inclined to be under the negative shock of the pandemic, either economically or emotionally. From an economic perspective, retail investors are more likely to experience worsening living conditions, employment status, and health conditions during the pandemic. Under such economic hardship, retail investors are more inclined to demand liquidity in response to the spread of the virus. Their decisions on the financial market may reflect the reactions of the common public to the pandemic. From the perspective of visceral response to the virus, when compared to institutional investors, retail investors are more prone to be biased by visceral feelings when making financial decisions. They may also lack the financial or economic sophistication to analyze the impact of Covid spread on the economics and financial market and therefore are more inclined to demand liquidity from mutual funds due to sentimental rather than rational reasoning. As such, I hypothesize that the reducing effect of local Covid spread on fund flows will be more pronounced for retail fund class shares since their clientless are more likely to be under the influence of negative shock from the pandemic.

To examine this hypothesis, I categorize fund share class based on the fund type information provided by the CRSP mutual fund summary dataset. Precisely, I classify the mutual fund share class in the sample into retail and institution funds according to fund type information. I included a dummy variable, Retail, in the specification, which equals one if the fund is a retail fund and 0 if the fund is an institution fund. Further, for each of the four measures of state-level Covid spread, I interact with the dummy variable Retail. Under my hypothesis that retail investors have a stronger reaction to Covid spread, I expect the coefficients on the interaction term to be negative and statistically significant.

Table 2.5 reports the results with the interaction term included in the specification. The results are generally consistent with the hypothesis that retail investors demand more liquidity in response to the virus, as the coefficients on the interaction term are negative and statistically significant for all four measures of Covid spread. Notably, for measures of Covid deaths, the coefficients are statistically significant only for the interaction term, suggesting that retail investors are reacting to Covid deaths by demanding liquidity from mutual funds. One explanation is that Covid deaths trigger a more visceral response from retail investors, therefore resulting in more liquidity demand in retail fund class shares. Under this explanation, the results lend more support to visceral response as one of the driving forces behind the Covid-induced mutual fund net flows.

2.7 Further Analysis: Fund Level Risk

In this section, I discuss the risk aversion channel of explanation for Covid-induced mutual fund outflows. The negative shock of the Covid pandemic may shift investors' risk aversion, resulting in a heterogeneous impact on liquidity demand across financial assets. Existing studies document that mental status may play a key role in determining risk aversion. Through this channel, visceral feelings triggered by the emergence of a virus may lead to a shift in risk-taking behaviors. On the other hand, a negative shock on income or wealth induced by the pandemic may also lead to heightened risk aversion. Risk aversion heightened by the pandemic may pressure investors to liquidate their investments in risky financial assets.

If heightened risk aversion is a major economic channel that explains the Covid-induced fund outflows, a natural question to ask is whether this reducing effect on net flows is more pronounced for funds associated with a higher level of risk. Under this risk aversion channel of explanation, I hypothesized that the Covid-induced fund outflows are more pronounced for funds associated with higher risk. In this section, I examined this question from three perspectives of fund-level risk: fund investment style, fund market beta, and fund volatility.

2.7.1 Fund Lipper Investment Style

The first measure of fund-level risk to be examined is fund investment style. The overall risk level of the fund class share is fundamentally determined by the type of asset held in the portfolio. From the perspective of investors, the most direct way to assess the risk level of mutual funds is through the type of asset the fund mainly holds. If the Covid-induced mutual fund outflows can be explained from the channel of heightened risk aversion, investors should demand more liquidity from funds with a riskier investment style. Wang and Young (2020) documented that terrorism-induced net flows are negative for funds with risky investment objectives (equity) and positive for those with relatively safe investment objectives (bond). An interesting question to examine is whether Covid-induced fund net flows follow similar patterns, particularly if a heterogeneous effect exists across funds with different investment styles. To help further understand the risk aversion channel of explanation, I empirically examine the interplay between fund investment style and Covid-induced net fund net flows. Funds that invest heavily in corporate, municipal bonds, debt, or money markets are generally considered less risky than those that invest in the equity market. Following Wang and

Young (2020), I am particularly interested in sorting individual funds based on the risk level of their investment style.

With data from the CRSP database, I collected information on the Lipper fund investment style for each fund share class. With a description of the Lipper fund style, I further categorized the fund share class based on their investment objectives. There are 123 unique Lipper objective descriptions in the sample, and I manually categorized fund share classes based on their investment objectives. If the words like "debt," bond," money market," and "fixed income" appear in the description of the fund Lipper investment style, I categorized the fund share class investment style into the safe group. Under the risk aversion channel of explanation, I hypothesized that the reducing effect of Covid spread on fund net flow is more concentrated in funds that invest in risky assets and, therefore, less pronounced for funds that focus on safe assets. To examine this hypothesis, I further constructed a dummy variable Safe, which equals one if the fund falls into the safe category and equals zero if the fund falls into the risky category. I interact the dummy variable Safe with all four measures of state-level Covid spread. Under the hypothesis, I expected the coefficients on the interaction term to be positive and statistically significant.

Table 2.6 reports the results when the dummy variable $Safe$ and the interaction term are included in the specification. Consistent with the hypothesis, the coefficients on the interaction term are positive and statistically significant for all four measures of state-level Covid spread, indicating that the reducing effects of local Covid spread on mutual fund net flows are less pronounced for fund class share with safe investment styles. For monthly Covid deaths, the magnitude of the interaction term exceeds that of the independent variable. For cumulative Covid deaths, the coefficients are statistically significant only for the interaction term. These results potentially imply that local Covid deaths potentially trigger inflows into fund share class with a safe investment style. Overall, the results are consistent with the risk aversion channel of explanation.

2.7.2 Fund Volatility

Another perspective to examine the risk aversion channel is through the interactive effect between fund volatility and local Covid spread on fund net flows. I hypothesize that the Covid- induced mutual fund net flows are more concentrated on funds with high return volatility if investors are demanding liquidity due to heightened risk aversion. For each month, I categorize fund class shares into three quantiles based on the volatility of daily return in the previous month. Then, I create a dummy variable Vol_H that indicates whether the fund share class falls into the high volatility quantile. Specifically, Vol_H equals one if the fund share class falls into the third quantile of return volatility and 0 otherwise. I further interacted with the dummy variable Vol_H with all four measures of the local Covid spread. Under my hypothesis, I expect the coefficients on the interaction term to be negative and statistically significant.

Table 2.7 reports the results with the dummy Vol_H and its interaction term with the Covid spread included in the specification. Since the coefficients on the interaction term are negative and statistically significant for all four measures of local Covid spread, the results support the hypothesis that Covid-induced outflows are more concentrated in funds with higher levels of risk. The empirical evidence further supports the risk aversion channel of Covid-induced fund outflows.

I also use fund daily return volatility in a longer time range as alternative measures of fund-level risk. Specifically, for each fund share class and month combination, I measure return volatility as the standard deviation of fund share class daily return in the prior 3 -and 12 -months windows to create a dummy variable indicating high volatility following the same procedures mentioned above. Table 2.8 and 2.9 report the results with alternative measurements of fund volatility. Overall, the results are consistent with the ones displayed in Table 2.7, indicating that the reducing effect of the local Covid spread is more pronounced for funds with higher volatility.

2.7.3 Fund Beta

I further examined the risk aversion channel of Covid-induced fund outflow from the perspective of fund betas. Mutual fund share classes with higher market beta have a higher exposure to market risk. Suppose the economic mechanisms behind Covid-induced fund outflows are risk aversion heightened by the pandemic. In that case, I hypothesize the reducing effect on fund net flows to be more pronounced for funds with higher betas. As mentioned previously, I estimated the market beta for each fund by running 36 -months of rolling regressions at a monthly frequency as a measurement of fund share class exposure to market risk. To examine the interactive effect between Covid spread and market beta, I further categorize mutual funds into three quantiles based on their market beta for each year- month. I construct a dummy variable, β_H, indicating high market beta, which equals one if the fund falls into the third quantile of market beta and 0 otherwise, to interact with measures of local Covid spread. Under my hypothesis, I expect the coefficients on the interaction term to be negative and statistically significant.

Table 2.10 reports the results with the dummy variable β_H and its interaction term with Covid measures included in the specification. The coefficients on the interaction term are all negative and statistically significant, implying that Covid-induced mutual fund net flows are more pronounced for fund share classes with higher market risk exposure.

I use an alternative measurement of fund share class beta to examine the hypothesis. I measure the fund market beta by running a 36 -months rolling regression of the Fama-French 4 -factor model and CAPM model for each fund share class. Table 2.11 and 2.12 show the results with alternative measures of fund market beta. According to the results displayed, the coefficients on the interaction term are all negative and statistically significant for the alternative measurements of fund share class measures, which are consistent with the findings displayed in Table 2.10. Overall, the results indicate that local Covid spread leads to more outflows from the fund share class with higher market risk exposure, consistent with the risk aversion channel of explanation.

2.8 Alternative Mechanism: Local Economic Condition

There are two competing economic mechanisms behind the reducing effect of local Covid spreads on mutual fund outflows used to understand the results discussed previously. The first channel is that Covid spread induces a visceral response from local investors, and this change in mental status subsequently leads to higher risk aversion and more liquidity demand from the mutual fund share class. The second channel is that Covid induces demand liquidity through adverse shocks on investor income and wealth through worsening local economic conditions. Levine et al. (2021) document that local Covid spread has increased search volume for terms related to economic hardship, such as "income loss" or "unemployment." The economic damage of Covid spread may press investors to liquidate their assets in mutual funds. Therefore, a question to address is which channel mainly drives the Covid-induced mutual fund outflows. In this section, I attempt to disentangle these two potential economic channels by controlling for local economic conditions in the specification.

I collected data on state-level monthly unemployment rates from the BLS website to measure local economic conditions. Following Levine et al. (2021), I also collected data from the monthly Google search volume index for "income loss," "job loss," and "unemployment" as measurements of residents' attention to economic hardship. If the Covid-induced fund net flows can be explained by the worsening local economic condition during the pandemic, then controlling for economic factors in the specification is expected to absorb some of the effects of Covid spread on mutual fund net flows. Therefore, under the economic channel of explanation, I expect the coefficients on local Covid spread measurements to be less pronounced after including these measures of local economic conditions in the specification. If including economic factors does not alter the main coefficients, then the visceral response is a more plausible explanation for Covid-induced fund outflows.

Table 2.13 displays the results with measures of local economic factors included in the specification. The inclusion of controls of local economic factors does not significantly alter the coefficients on measures of local Covid spread when compared to the baseline results in Table 2.3. The coefficients are not statistically significant for any of these controls for local economic factors, implying that worsening economic conditions may not be the main economic channel behind Covid-induced mutual fund outflows.

In the appendix, Table 2.A.1 – 2.A.4 display the results with specifications that include interaction terms with dummy variables introduced in previous sections. The results indicate that controlling for local economic factors does not significantly alter the magnitude of coefficients of key interest, further implying that worsening economic conditions is not the main driving force behind Cvid-induced outflows. Therefore, it is more plausible that investors demand liquidity from mutual funds as a visceral response to the spread of Covid rather than financial decisions under economic hardship.

2.9 Alternative Measurement: State Level Google Search Volume

To lend further support that the Covid-induced outflows come from concerns over the pandemic in the local area, I use an alternative measurement in the specification. Attention to Covid, as measured by Google search volume, can be used as a proxy for investors' concern over the virus. Specifically, I replace the measurement of local Covid spread with a logarithm of one plus Google search volume of" "Covid" at the state monthly level. Table 2.14 displays the results. In general, the findings indicate that attention to the Covid spread leads to more liquidity demand from locally headquartered mutual funds. The coefficients on the interaction term are generally consistent with the findings in previous sections, except for the Safe dummy interaction, which becomes insignificant. Overall, the results corroborate concerns over the virus as a major driving force behind the Covid-induced mutual fund outflows.

2.10 Conclusion and Discussion

In this paper, I empirically examine the impact of local Covid spreads on the net flows of locally headquarter mutual funds. The results indicate that Covid spread reduces the net flows of locally headquartered mutual funds. This reducing effect was more pronounced during the early stage of the pandemic when public fear of the virus peaked. Nonetheless, even after the Covid crash period, Covid spread still induces outflows from locally headquartered mutual funds though the magnitude is more minor. The further empirical analysis also indicated that the reducing effect is more pronounced for retail fund share classes, which implies that Covid spread has a more salient effect on inducing liquidity demand from retail investors.

This study also attempts to examine the risk aversion channel of explanation behind Covid- induced mutual fund net flows from the perspective of fund-level risk. The empirical findings indicate that local Covid spread has a heterogeneous effect on inducing liquidity demand from funds with different levels of risk. Notably, a higher level of fund risk is associated with more pronounced outflows triggered by Covid spread. The empirical evidence is generally consistent with the risk aversion channel of explanation.

Further evidence lends support to the visceral response economic channel of explanation. Particularly, controlling for local economic conditions does not significantly alter the main coefficients, which indicates that economic hardship may not be the main driving force behind Covid-induced net flows. With Google search volume of "Covid" as an alternative measurement of local concerns over the pandemic, the results are generally consistent with the baselines. Overall, the results lend support to the visceral response channel of explanation.

Overall, the findings in this study have important implications for existing literature and future works. First, the study further corroborates the existing studies on the impact of mental status on financial decision-making and risk-taking behavior. Using the Covid pandemic as an exogenous shock, this study implies that the visceral response induced by the pandemic leads to asset liquidation, which can be potentially explained by heightened risk aversion. The findings of changes in risk aversion during the pandemic can offer further implications to the literature on time-varying risk aversion. Concerns over the virus may explain the demand for financial assets during the pandemic, which may have further implications for asset pricing. Future studies can be built upon the findings in this study by using more micro-level data, such as brokerage-level data of investor transactions, particularly for retail investors.

References

Barber, B. M., Huang, X., & Odean, T. (2015). Which risk factors matter to investors? Evidence from mutual fund flows. *Working Pape*r.

Bharath, S. T., & Cho, D. (2023). Do natural disaster experiences limit stock market participation?. *Journal of Financial and Quantitative Analysis*, 58(1), 29-70.

Bogan, V. L., & Yonker, S. E. (2021). Fear and Risk: Do Visceral Factors Affect Risk Taking?. *Working Paper.*

Cashman, G. D., Nardari, F., Deli, D. N., & Villupuram, S. V. (2014). Investor behavior in the mutual fund industry: evidence from gross flows. *Journal of Economics and Finance*, 38, 541-567.

Craft, T. M. (2006). Home bias makes sense for US pension plans. *Journal of Portfolio Managementt*, 32(3), 26.

Cole, S. A., & Shastry, G. K. (2009). *Smart money: The effect of education, cognitive ability, and financial literacy on financial market participation* (pp. 09-071). Boston, MA: Harvard Business School.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6), 2045-2073.

Fahlenbrach, R., Rageth, K., & Stulz, R. M. (2021). How valuable is financial flexibility when revenue stops? Evidence from the COVID-19 crisis. *The Review of Financial Studies*, 34(11), 5474-5521.

Gollwitzer, A., Martel, C., Brady, W. J., Pärnamets, P., Freedman, I. G., Knowles, E. D., & Van Bavel, J. J. (2020). Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nature Human Behaviour*, 4(11), 1186-1197.

Guiso, L., Sapienza, P., & Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, 128(3), 403-421.

Haddad, V., Moreira, A., & Muir, T. (2021). When selling becomes viral: Disruptions in debt markets in the COVID-19 crisis and the Fed's response. *The Review of Financial Studies*, 34(11), 5309-5351.

Ippolito, R. A. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *The Journal of Law and Economics*, 35(1), 45-70.

Ivković, Z., & Weisbenner, S. (2009). Individual investor mutual fund flows. *Journal of Financial Economics*, 92(2), 223-237.

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

Kargar, M., Lester, B., Lindsay, D., Liu, S., Weill, P. O., & Zúñiga, D. (2021). Corporate bond liquidity during the COVID-19 crisis. *The Review of Financial Studies*, 34(11), 5352-5401.

Ke, D. (2022). Left behind: Partisan identity and wealth inequality. *Working Paper*.

Ke, D., Ng, L., & Wang, Q. (2010). Home bias in foreign investment decisions. *Journal of International Business Studies*, 41, 960-979.

Khalatbari-Soltani, S., Cumming, R. C., Delpierre, C., & Kelly-Irving, M. (2020). Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *Journal of Epidemiol Community Health*, 74(8), 620-623.

Kuhnen, C. M., & Knutson, B. (2011). The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis*, 46(3), 605-626.

Levine, R., Lin, C., Tai, M., & Xie, W. (2021). How did depositors respond to COVID-19?. *The Review of Financial Studies*, 34(11), 5438-5473.

Li, L., Li, Y., Macchiavelli, M., & Zhou, X. (2021). Liquidity restrictions, runs, and central bank interventions: Evidence from money market funds. *The Review of Financial Studies*, 34(11), 5402-5437.

Rabin, C., & Dutra, S. (2022). Predicting engagement in behaviors to reduce the spread of COVID-19: the roles of the health belief model and political party affiliation. *Psychology, Health & Medicine*, 27(2), 379-388.

Seasholes, M. S., & Zhu, N. (2010). Individual investors and local bias. *The Journal of Finance*, 65(5), 1987-2010.

Sirri, E. R., & Tufano, P. (1998). Costly search and mutual fund flows. *The Journal of Finance*, 53(5), 1589-1622.

Tubaldi, R. (2020). Mutual funds' fire sales and the real economy: Evidence from hurricanes. *Working Paper.*

Vorsatz, M. B. (2022). Costs of political polarization: Evidence from mutual fund managers during covid-19. *Working Paper.*

Wang, A. Y., & Young, M. (2020). Terrorist attacks and investor risk preference: Evidence from mutual fund flows. *Journal of Financial Economics*, 137(2), 491-514.

Tables

Table 2.1: Summary Statistics for Fund Level Variables

This table reports the summary statistics for fund share class level variables. Lag Ln(TNA) refers to the one month lagged total net asset. Lag Ln(TNA per Share) refers to one month lagged total net asset per share. Lag Ln(Volatility) refers to log of one month lagged volatility. Lag Ln(TNA per Share) refers to one month lagged total net asset per share. Lag Return refers to one month lagged return. $NetFlow_{i,t-1}$ and $NetFlow_{i,t-2}$ are one and two month lagged net flow.

	N.of.Ob	Mean	Standard Deviation	Min	Max
NetFlow	454637	-0.000737	0.0904	-0.4206	0.7124
Lag Ln(TNA)	454637	3.926722	2.7358	-2.3026	12.9915
Lag Ln(TNA per Share)	454637	2.843840	0.7258	-0.7765	7.6143
Lag Ln(Volatility)	454637	-5.168082	1.1994	-19.9499	0.5252
Log Volatility_i, t	454547	-5.161357	1.2080	-19.9499	0.5252
Expense Ratio	454637	0.009506	0.0056	-0.0010	0.1179
Turnover Ratio	454637	0.854433	2.0921	0.0000	110.6600
Lag Return	454637	0.009746	0.0578	-0.9564	7.6087
NetFlow_i,t-1	454637	-0.000336	0.0904	-0.4211	0.7081
NetFlow_i,t-2	454637	-0.000263	0.0910	-0.4173	0.7196
beta	454637	0.607569	0.4269	-6.8115	5.2119

Table 2.2: Summary Statistics for State Level Variables

This table reports the summary statistics for state level variables. Monthly Cases PC and Monthly Deaths PC are state level monthly Covid cases or deaths per 10,000 people. Log Monthly Cases PC and Log Month Deaths PC are log of state level monthly cases or deaths. Total cases PC and Total Deaths PC are cumulative cases or deaths per 10,000 people at state level. Log Total Cases PC and Log Total Deaths PC are log of cumulative cases or deaths per 10,000 people at state level. Unemployment is monthly state level unemployment rate. Ln(income loss index) is the log of state level Google search volume index for "income loss". Income loss is the state level Google search volume index for "income loss". Ln(Job Loss index) is the log of state level Google search volume index for "Job Loss". Job Loss index is the state level Google search volume index for "Job Loss". Ln(Unemployment index) is the log of state level Google search volume index for "Unemployment". Unemployment index is the state level Google search volume index for "Unemployment".

	N.of.Ob	Mean	Standard Deviation	Min	Max
Monthly Cases PC	1150	635.892922	677.6563	0.0000	4626.9331
Monthly Deaths PC	1150	9.667740	11.1456	0.0000	111.4261
Log Monthly Cases PC_s, t-1	1150	5.485336	2.0382	0.0000	8.4399
Log Monthly Deaths PC_s,t-1	1150	1.887512	1.0276	0.0000	4.7223
Total Cases PC	1150	5885.691177	5302.5794	0.0000	21255.3223
Total Deaths PC	1150	101.902710	87.2349	0.0000	328.4253
Log Total Cases PC_s,t-1	1150	7.309210	2.7853	0.0000	9.9644
Log Total Deaths PC_s,t-1	1150	3.800810	1.7562	0.0000	5.7973
Unemployment	1002	0.064235	0.0334	0.0180	0.2750
Ln(Income Loss Index)	1002	3.038875	1.0867	0.0000	4.6151
Income Loss	1002	29.651697	21.3031	0.0000	100.0000
Ln(Job Loss Index)	1002	2.339955	0.9057	0.0000	4.3307
Job Loss Index	1002	13.202595	10.3968	0.0000	75.0000
Ln(Unemployment Index)	1002	3.308675	0.7978	1.0986	4.6151
Unemployment Index	1002	34.647705	23.7393	2.0000	100.0000

Table 2.3: Baseline Empirical Results:

This table reports the baseline empirical results with OLS estimation. The dependent variable is fund share class level net flow. The key independent variables are log of state level monthly Covid cases per 10,000 people, log of state level monthly Covid deaths per 10,000 people, log of state level cumulative Covid cases per 10,000 people and log of state level cumulative Covid deaths per 10,000 people.* p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001634***			
	(-5.93)			
Log Monthly Deaths PC_s,t-1		-0.0008038***		
		(-2.82)		
Log Total Cases PC_s,t-1			-0.001751***	
			(-3.97)	
Log Total Deaths PC_s,t-1				-0.00002657
				(-0.06)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.4: Results with Subsamples during and after the Covid Crash Period:

This table reports the empirical results with OLS estimation. Panel A reports the estimation results in the subsample during Covid crash period, 01/2020 - 04/2020. Panel B reports the estimation results in the subsample after the Covid crash period, 05/2020-12/2021. The dependent variable is fund share class level net flow. The key independent variables are log of state level monthly Covid cases per 10,000 people, log of state level monthly Covid deaths per 10,000 people, log of state level cumulative Covid cases per 10,000 people and log of state level cumulative Covid deaths per 10,000. * p<0.1, ** p<0.05, *** p<0.01

Panel A: Subsample in the Covid Crash Period

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.0031612**			
	(-3.33)			
Log Monthly Deaths PC_s,t-1		-0.0040438**		
		(-3.02)		
Log Total Cases PC_s,t-1			-0.0031619**	
			(-3.33)	
Log Total Deaths PC_s,t-1				-0.004044**
				(-3.02)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	63162	63162	63162	63162
Adjusted R^2	0.0225	0.0225	0.0225	0.0225

Panel B: Subsample in the Post Covid Crash Period

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001187***			
	(-4.57)			
Log Monthly Deaths PC_s,t-1		-0.0005095*		
		(-1.95)		
Log Total Cases PC_s,t-1			-0.0009142**	
			(-2.10)	
Log Total Deaths PC_s,t-1				0.0002658
				(0.66)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	394787	394787	394787	394787
Adjusted R^2	0.032	0.032	0.032	0.032

Table 2.5: Results with Retail Fund Share Indicator

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and retail fund share class. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. *Retail* is a dummy variable which equals to 1 if the fund share class is retail and 0 otherwise. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001445*** (-5.05)			
Retail × Log Monthly Cases PC_s, t-1	-0.0004312*** (-2.82)			
Log Monthly Deaths PC_s,t-1		-0.0001706 (-0.53)		
Retail × Log Monthly Deaths PC_s,t-1		-0.001463*** (-5.13)		
Log Total Cases PC_s,t-1			-0.001616*** (-3.64)	
Retail × Log Total Cases PC_s,t-1			-0.0002807** (-2.47)	
Log Total Deaths PC_s,t-1				0.0002862 (0.67)
Retail × Log Total Deaths PC_s,t-1				-0.0006803*** (-3.83)
Retail	0.002298** (2.49)	0.002779*** (4.22)	0.002028** (2.16)	0.002769*** (3.28)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457943	457943	457943	457943
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.6: Results with Safe Investment Style Indicator

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund Lipper investment style. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. *Safe* is a dummy variable which equals to one if the fund focuses on safe investment asset (fixed income, money market etc) * $p<0.1$, ** $p<0.05$, *** $p<0.01$

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001843***			
	(-6.58)			
Safe × Log Monthly Cases PC_s, t-1	0.0009392***			
	(4.37)			
Log Monthly Deaths PC_s,t-1		-0.001211***		
		(-4.07)		
Safe × Log Monthly Deaths PC_s,t-1		0.001859***		
		(4.75)		
Log Total Cases PC_s,t-1			-0.001850***	
			(-4.18)	
Safe × Log Total Cases PC_s,t-1			0.0004141***	
			(2.66)	
Log Total Deaths PC_s,t-1				-0.0001440
				(-0.34)
Safe × Log Total Deaths PC_s,t-1				0.0005281**
				(2.20)
Safe	-0.0007183	0.0007997	0.001316	0.002187*
	(-0.51)	(0.78)	(0.94)	(1.73)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.7: Results with High One Month Volatility Quantile Indicator

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level volatility measures. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. Vol_H is a dummy variable which equals to 1 if the onemonth volatility of the fund share class falls into the 3rd quintile and 0 otherwise. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001191*** (-4.10)			
Vol_H × Log Monthly Cases PC_s, t-1	-0.0006181*** (-3.35)			
Log Monthly Deaths PC_s,t-1		-0.0004179 (-1.33)		
Vol_H × Log Monthly Deaths PC_s,t-1		-0.001036*** (-3.05)		
Log Total Cases PC_s,t-1			-0.001392*** (-2.98)	
Vol_H × Log Total Cases PC_s,t-1			-0.0004473*** (-3.26)	
Log Total Deaths PC_s,t-1				0.0002386 (0.54)
Vol_H × Log Total Deaths PC_s,t-1				-0.0008821*** (-4.06)
Vol_H	0.006984*** (6.09)	0.005633*** (6.79)	0.006909*** (5.93)	0.007267*** (6.78)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.8: Results with High Three Months Volatility Quantile Indicator

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level volatility measures. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. Vol_H is a dummy variable which equals to 1 if the three month volatility of the fund share class falls into the 3rd quintile and 0 otherwise. p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001409*** (-5.02)			
Vol_H × Log Monthly Cases PC_s, t-1	-0.0006529*** (-3.62)			
Log Monthly Deaths PC_s,t-1		-0.0004080 (-1.33)		
Vol_H × Log Monthly Deaths PC_s,t-1		-0.001176*** (-3.48)		
Log Total Cases PC_s,t-1			-0.001604*** (-3.63)	
Vol_H × Log Total Cases PC_s,t-1			-0.0004607*** (-3.45)	
Log Total Deaths PC_s,t-1				0.0002657 (0.63)
Vol_H × Log Total Deaths PC_s,t-1				-0.0009306*** (-4.43)
Vol_H	0.007748*** (6.88)	0.006480*** (7.86)	0.007582*** (6.67)	0.008041*** (7.74)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.9: Results with High Twelve Month Volatility Quantile Indicator

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level volatility measures. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. Vol_H is a dummy variable which equals to 1 if the 12 month volatility of the fund share class falls into the 3rd quintile and 0 otherwise. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001191*** (-4.10)			
Vol_H × Log Monthly Cases PC_s, t-1	-0.0006181*** (-3.35)			
Log Monthly Deaths PC_s,t-1		-0.0004179 (-1.33)		
Vol_H × Log Monthly Deaths PC_s,t-1		-0.001036*** (-3.05)		
Log Total Cases PC_s,t-1			-0.001392*** (-2.98)	
Vol_H × Log Total Cases PC_s,t-1			-0.0004473*** (-3.26)	
Log Total Deaths PC_s,t-1				0.0002386 (0.54)
Vol_H × Log Total Deaths PC_s,t-1				-0.0008821*** (-4.06)
Vol_H	0.006984*** (6.09)	0.005633*** (6.79)	0.006909*** (5.93)	0.007267*** (6.78)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.10: Results with High Market Beta Quantile Indicator – Fama French Three Factor Model

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level Fama French 3 factor beta. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. $beta_H$ is a dummy variable which equals to 1 if the Fama-French 3 factor beta of the fund share class falls into the 3rd quintile and 0 otherwise. p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001406***			
	(-5.02)			
beta_H × Log Monthly Cases PC_s, t-1	-0.0006689***			
	(-4.12)			
Log Monthly Deaths PC_s,t-1		-0.0004149		
		(-1.36)		
beta_H × Log Monthly Deaths PC_s,t-1		-0.001130***		
		(-3.57)		
Log Total Cases PC_s,t-1			-0.001569***	
			(-3.56)	
beta_H × Log Total Cases PC_s,t-1			-0.0005176***	
			(-4.40)	
Log Total Deaths PC_s,t-1				0.0002877
				(0.68)
beta_H × Log Total Deaths PC_s,t-1				-0.0008918***
				(-4.78)
beta_H	0.007628***	0.006164***	0.007804***	0.007698***
	(7.23)	(7.45)	(7.50)	(7.95)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.11: Results with High Market Beta Quantile Indicator – Fama French Carhart Four Factor Model

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level Fama French 4 factor beta. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. $beta_H$ is a dummy variable which equals to 1 if the Fama-French 4 factor beta of the fund share class falls into the 3rd quintile and 0 otherwise. p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001402*** (-5.02)			
beta_H × Log Monthly Cases PC_s, t-1	-0.0006962*** (-4.27)			
Log Monthly Deaths PC_s,t-1		-0.0004029 (-1.32)		
beta_H × Log Monthly Deaths PC_s,t-1		-0.001181*** (-3.71)		
Log Total Cases PC_s,t-1			-0.001563*** (-3.54)	
beta_H × Log Total Cases PC_s,t-1			-0.0005339*** (-4.52)	
Log Total Deaths PC_s,t-1				0.0003018 (0.72)
beta_H × Log Total Deaths PC_s,t-1				-0.0009231*** (-4.92)
beta_H	0.008017*** (7.49)	0.006488*** (7.75)	0.008160*** (7.71)	0.008069*** (8.18)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.12: Results with High Market Beta Quantile Indicator – CAPM Model

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and fund level CAPM beta. The sample covers the period 02/2020 - 12/2020. The dependent variable is monthly net flow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. $beta_H$ is a dummy variable which equals to 1 if the Fama-French 4 factor beta of the fund share class falls into the 3rd quintile and 0 otherwise. p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001235*** (-4.26)			
beta_H × Log Monthly Cases PC_s, t-1	-0.0004911*** (-2.93)			
Log Monthly Deaths PC_s,t-1		-0.0005237* (-1.67)		
beta_H × Log Monthly Deaths PC_s,t-1		-0.0006959** (-2.16)		
Log Total Cases PC_s,t-1			-0.001406*** (-3.00)	
beta_H × Log Total Cases PC_s,t-1			-0.0003207*** (-2.62)	
Log Total Deaths PC_s,t-1				0.0001682 (0.38)
beta_H × Log Total Deaths PC_s,t-1				-0.0005161*** (-2.64)
beta_H	0.005360*** (4.98)	0.004024*** (4.84)	0.005044*** (4.71)	0.004831*** (4.81)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.13: Results with Controls of Local Economic Factors

This table reports the empirical results of OLS estimation with controls of local economic conditions. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. The regression includes additional controls for local economic conditions. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001675*** (-6.05)			
Log Monthly Deaths PC_s,t-1		-0.0008086*** (-2.81)		
Log Total Cases PC_s,t-1			-0.001761*** (-3.97)	
Log Total Deaths PC_s,t-1				0.00004840 (0.11)
Unemployment	-0.01889 (-1.20)	-0.005562 (-0.35)	-0.006132 (-0.39)	-0.01008 (-0.61)
Unemployment Index	-0.003090* (-1.88)	-0.002498 (-1.52)	-0.002788* (-1.70)	-0.002617 (-1.60)
Job Loss Index	0.001084 (0.46)	0.0007737 (0.33)	0.001240 (0.53)	0.0007945 (0.34)
Income Loss Index	-0.001522 (-1.08)	-0.002096 (-1.50)	-0.002033 (-1.45)	-0.002146 (-1.53)
Controls	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.14: Results with State-Level Google Search Index for "Covid" as an Alternative Measurement

This table reports the empirical results of OLS estimation with state-level Google search index as an alternative measurement for Covid spread. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow	(5) NetFlow
Covid_index	-0.004803***	-0.003272*	-0.004631***	-0.004044**	-0.003970**
	(-2.80)	(-1.84)	(-2.66)	(-2.31)	(-2.28)
Retail		0.001284***			
		(2.71)			
Retail × Covid_index		-0.003374***			
		(-4.07)			
Safe			0.004529***		
			(5.77)		
Safe × Covid_index			-0.0005787		
			(-0.52)		
Vol_H				0.004585***	
				(6.97)	
Vol_H × Covid_index				-0.002308**	
				(-2.22)	
beta_H					0.004538***
					(6.70)
beta_H × Covid_index					-0.002266**
					(-2.28)
Controls	Yes	Yes	Yes	Yes	Yes
Economic Factor	Yes	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949	457949
Adjusted R^2	0.029	0.029	0.029	0.029	0.029

Appendix

Table 2.A.1: Results with Retail Fund Share Indicator and Controls for Local Economic Conditions This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and retail fund share class. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. *Retail* is a dummy variable which equals to 1 if the fund share class is retail and 0 otherwise. Local economic conditions are included as control variables in the specification. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001488***			
	(-5.18)			
Retail × Log Monthly Cases PC_s, t-1	-0.0004321***			
	(-2.82)			
Log Monthly Deaths PC_s,t-1		-0.0001656		
		(-0.51)		
Retail × Log Monthly Deaths PC_s,t-1		-0.001482***		
		(-5.19)		
Log Total Cases PC_s,t-1			-0.001628***	
			(-3.64)	
Retail × Log Total Cases PC_s,t-1			-0.0002768**	
			(-2.43)	
Log Total Deaths PC_s,t-1				0.0003636
				(0.80)
Retail × Log Total Deaths PC_s,t-1				-0.0006749***
				(-3.79)
Retail	0.002322**	0.002835***	0.002019**	0.002766***
	(2.51)	(4.30)	(2.15)	(3.28)
Controls	Yes	Yes	Yes	Yes
Economic Factors	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.A.2: Results with Safe Investment Style Indicator and Controls for Local Economic Conditions

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and a dummy variable $Safe$ indicating fund investment style. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. $beta_H$ is a dummy variable which equals to 1 if the Fama French 3 factor model beta falls into the 3rd quintile. Local economic conditions are included as control variables in the specification. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001887***			
	(-6.71)			
Safe × Log Monthly Cases PC_s, t-1	0.0009432***			
	(4.39)			
Log Monthly Deaths PC_s,t-1		-0.001216***		
		(-4.05)		
Safe × Log Monthly Deaths PC_s,t-1		0.001870***		
		(4.77)		
Log Total Cases PC_s,t-1			-0.001861***	
			(-4.17)	
Safe × Log Total Cases PC_s,t-1			0.0004128***	
			(2.65)	
Log Total Deaths PC_s,t-1				-0.00006924
				(-0.16)
Safe × Log Total Deaths PC_s,t-1				0.0005267**
				(2.19)
Safe	-0.0007187	0.0007980	0.001346	0.002214*
	(-0.51)	(0.77)	(0.96)	(1.75)
Constant	0.04467***	0.04430***	0.04360***	0.04334***
	(10.52)	(10.47)	(10.29)	(10.23)
Controls	Yes	Yes	Yes	Yes
Economic Factors	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.A.3: Results with High Volatility Indicator and Controls for Local Economic Conditions

This table reports the empirical results of OLS estimation with controls of local economic conditions. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. Vol_H is a dummy variable which equals to 1 if the 12 month volatility of the fund share class falls into the 3rd quintile and 0 otherwise. p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001543*** (-5.47)			
Vol_H × Log Monthly Cases PC_s, t-1	-0.0003986** (-2.16)			
Log Monthly Deaths PC_s,t-1		-0.0006208** (-2.02)		
Vol_H × Log Monthly Deaths PC_s,t-1		-0.0005860* (-1.68)		
Log Total Cases PC_s,t-1			-0.001669*** (-3.76)	
Vol_H × Log Total Cases PC_s,t-1			-0.0003490*** (-2.61)	
Log Total Deaths PC_s,t-1				0.0002654 (0.60)
Vol_H × Log Total Deaths PC_s,t-1				-0.0006504*** (-3.09)
Vol_H	0.008386*** (7.35)	0.007355*** (8.81)	0.008796*** (7.80)	0.008910*** (8.64)
Controls	Yes	Yes	Yes	Yes
Economic Factors	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

Table 2.A.4: Results with High Market Beta Quantile Indicator and Controls for Local Economic Conditions

This table reports the empirical results of OLS estimation, with interaction term between Covid spread measures and high market beta indicator. The sample covers the period 02/2020 - 12/2020. The dependent variable is Netflow at fund share class level. There are four separate regressions with log of monthly Covid cases per capita , log of monthly Covid deaths per capita, log of cumulative Covid cases per capita and log of cumulative Covid deaths per capita at state level. $beta_H$ is a dummy variable which equals to 1 if the Fama French 3 factor model beta falls into the 3rd quintile. Local economic conditions are included as control variables in the specification. * p<0.1, ** p<0.05, *** p<0.01

	(1) NetFlow	(2) NetFlow	(3) NetFlow	(4) NetFlow
Log Monthly Cases PC_s, t-1	-0.001447*** (-5.15)			
beta_H × Log Monthly Cases PC_s, t-1	-0.0006756*** (-4.16)			
Log Monthly Deaths PC_s,t-1		-0.0004216 (-1.37)		
beta_H × Log Monthly Deaths PC_s,t-1		-0.001119*** (-3.53)		
Log Total Cases PC_s,t-1			-0.001575*** (-3.55)	
beta_H × Log Total Cases PC_s,t-1			-0.0005241*** (-4.46)	
Log Total Deaths PC_s,t-1				0.0003750 (0.84)
beta_H × Log Total Deaths PC_s,t-1				-0.0009023*** (-4.83)
beta_H	0.007670*** (7.26)	0.006149*** (7.42)	0.007857*** (7.55)	0.007747*** (7.99)
Controls	Yes	Yes	Yes	Yes
Economic Factors	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes	Yes
Observations	457949	457949	457949	457949
Adjusted R^2	0.025	0.025	0.025	0.025

3 The Impact of Social Media Sentiment on IPO Pricings

3.1 Introduction

Social media has become an increasingly important platform for information exchange and opinion posting. Integrated with advanced information technology and the internet, social media has become more effective in gathering information from various sources and shaping investor sentiment in some respects. Compared to traditional news media, information flow on social media is updated much faster and, therefore, can reflect the most up-to-date sentiment on the internet. Additionally, compared to traditional media, social media has various sources of information flow, particularly those from the ordinary people in the public. With social media as a new platform for information sharing and gathering, ordinary people have acquired more power to shape public opinion, participate in political activities, or even change the trend in the financial market. As a source of information, social media has been playing an increasingly important role in the financial market, where the spread of information fundamentally determines the price movement of financial assets. According to Wu (2019), Greenwich Associates report that in interviews with 256 institutional investors, 80% of them have frequently used social media platforms to facilitate their work. Additionally, among investors who frequently use social media at work, nearly 1 out of 3 confirmed that the information they gathered from social media has an impact on their investment decision. As such, the quickly updated information flow on social media platforms may have imposed a significant impact on the asset pricing process.

Though social media posts may reflect the most up-to-date information flow on the internet, the content and sentiment on social media of this information flow contains much noise and bias and, therefore, can often be misleading—the lack of censorship or editorial rules results in decreased accuracy of information on social media platforms. Much of the information does not represent fact but rather misinformation or sentimental noise. As such, regarding the impact on the financial market, whether information on social media harms or enhances price efficiency remains a question to be answered. In early 2021, Retail investors gathered on the social media platform Reddit.com to push up stock prices for AMC and GameStop, even though the fundamentals of these two stocks have suffered severely from the Covid-19 pandemic. In this event, social media gathering has harmed the price efficiency of the financial market by causing large deviations in the stock prices of certain firms from their fundamentals.

This study empirically examines the impact of social media attention and sentiment on IPO pricing. The tested hypothesis comes from the prediction of prior theoretical frameworks (Ljungqvist et al. (2006), Cornelli, Goldreich, and Ljungqvist (2006), and Derrien (2005)) that over-optimism of retail investors results in overpricing of IPO shares in the short term after trading starts, which is followed by a long-term reversal. Using social media sentiment as a measure of retail investors' valuation of IPO shares, I hypothesize that a more optimistic social media sentiment during the pre-IPO period leads to a higher first-day return and a worse long-term performance.

The data on social media posts is collected from Stocktwits, one of the largest online social media platforms for investors in the United States. Specifically, for each IPO in the US market during 2016 – 2019, I collect all Stocktwits posts that contain its ticker symbol during a 90-day window prior to its first trading day. This data collection process results in a sample of 387 firms with a non-zero number of posts on Stocktwits during the pre-IPO period. The number of posts on Stocktwits during the pre-IPO period is used to measure investor attention toward a given IPO. To measure investors' sentiment toward a given IPO, I construct bullish and bearish sentiment measurements based on the ratio of "Bullish" and "Bearish" posts tagged by the user in the total number of posts, respectively.

Generally, the regression results show that greater attention or a more optimistic sentiment on social media predicts a higher first-day price run-up. Regarding the two sentiment measurements, the "Bearish" posts ratio tends to have a more pronounced impact on first-day price run-up than the "Bullish" posts ratio in terms of both magnitude and statistical significance. Furthermore, the empirical results suggest that the long-term reversal is mainly driven by sentiment rather than attention. The number of pre-IPO posts generally has no statistically significant impact on the long-term return of IPOs. The empirical results suggest that both bearish and bullish tend to result in significant reversals in the long run. The reversal caused by bearish sentiment tends to be larger in magnitude. However, the reversal caused by bullish sentiment tends to be more consistent, as indicated by a larger number of significant coefficients with various time horizons after the first trading day. When compared to the coefficients on the first trading day, the coefficients on long-term reversal are considerable in terms of magnitude. This suggests that social media sentiment has caused a significant level of overpricing in the first trading day. These results are robust after using market benchmarked returns as dependent variables or restricting the sample to firms with a relatively smaller number of posts. This robustness check indicates that the results of social media sentiment are not driven by market conditions or only firms that attract enormous volume of investor attention.

Moreover, most of the posts in the sample are tagged as neither “Bullish” nor “Bearish”. As such, I attempt to classify these untagged posts with natural language process (NLP) techniques to see if the previous results still hold when these untagged posts are included in the construction of sentiment measurements. Specifically, I have employed a word-to-vector algorithm FastText, which is developed by Facebook AI Research, to train a machine learning model on posts tagged by users. After that, I used the trained model to classify all the untagged posts and used these classified posts to construct sentiment measurements. Consistent with the previous results, the empirical results generally indicate that a more optimistic social media sentiment leads to a higher first-day price run-up and a worse long-term performance.

Lastly, I attempt to examine whether social media attention or sentiment leads to a higher turnover rate in the short run after IPO. If the overvalued demand from retail investors drives the above-described results, then a more optimistic social media sentiment leads to a higher turnover rate after the IPO. With machine learning classified posts, the empirical results show that more positive pre-IPO sentiment leads to a higher turnover rate on the first trading day or week after the IPO. This suggests that the frenzied demand from retail investors drives the results of short-term price run-ups and long-term reversals. Informed investors may have sold their IPO shares at an overvalued price to overoptimistic retail investors shortly after the IPO.

This paper is organized as the following. Section 2 introduces the background and hypothesis development, including related literature. Section 3 presents the main empirical specification employed in this study. Section 4 describes the data source and variable construction and presents summary statistics. Section 5 presents the main empirical results. Section 6 presents some robustness checks to the main empirical results. Section 7 presents further analysis on classifying untagged posts and examining the initial turnover rate. Section 8 makes the conclusion.

3.2 Background and Hypothesis Development

3.2.1 Related Literatures

This study contributes to several distinct streams of literature in finance. Firstly, it contributes to the literature that explores the impact of online platforms on the financial market. With the support of advanced information technology, these emerging online platforms, such as social media, provide an essential source of information to the financial market. Compared to traditional news media, user posts on the online platform represent the opinion of ordinary people from the public and therefore reflect the crowd's wisdom. A critical feature of these online platforms is the lack of editorial censorship, which guarantees various sources of information on the platforms. However, this lack of censorship also leads to massive amount of biased information on these online platforms. Therefore, whether they helped enhance information efficiency in the financial market remains to be determined.

In recent years, rich literature has emerged that focuses on user posts and reviews on these online platforms and on how such information flow may influence the asset pricing process or corporate governance. There are mainly two stands of literature focusing on this field. The first strand mainly examines how firms employ social media platforms to interact with investors. Blankespoor, Miller, and White (2014) show that firms use Twitter to disseminate their disclosure information among market participants further, resulting in lower information asymmetry and higher liquidity. Jung, Naughton, Tahoun, and Wang (2018) show that firms strategically disseminate their disclosure information on Twitter to shape their informational environment. Specifically, they find that firms are less likely to disseminate via Twitter when the quarterly earnings announcement misses analysts' forecasts. Lee, Hutton, and Shu (2015) find that firms employ corporate social media to communicate with investors to attenuate the adverse price impact of product recall announcements. Chen, Hwang, and Liu (2019) find that corporate executives use social media to communicate with market participants, which improves the firm's informational environment and liquidity.

The second strand of studies focuses on how information flow from online platforms helps predict disclosure information or asset price movement. This paper is more closely related to this strand of study. Mao, Wei, Wang, and Liu (2012) find that a daily number of tweets mentioning S&P 500 stocks has significant predictive power on the subsequent price change of the S&P 500 index. Curtis, Richardson, and Schmardebeck (2014) find that a higher level of investor attention on social media leads to a more significant price reaction to the earnings announcement and a less pronounced post-earnings announcement drift. Bartov, Faurel, and Mohanram (2018) find that aggregate sentiment from individual tweets on Twitter can positively predict the company's quarterly earnings announcement and its price reaction. Using textual analysis of the articles published on Seeking Alpha, Chen, De, Hu, and Hwang (2014) find that sentiment from both the articles and commentaries positively predicts subsequent stock return and earnings surprise. By employing machine learning techniques to classify posts on Stocktwits, McGurk, Nowak, and Hall (2019) and Leung, Wong, and Wong (2020) find that social media sentiment positively predicts future stock returns. Cookson and Niessner (2020) find that a higher level of disagreement on social media (Stocktwits) leads to higher trading volume, and this effect is more pronounced for disagreement among investors with the same investment philosophy.

Additionally, this study adds to the literature on IPO underpricing, particularly with the strand that emphasizes the impact of pre-IPO information flow. Prior empirical studies have documented the first-day price run-up of IPO as underpricing, which is measured as the percentage difference between the first-day closing price and offer price. Existing studies offer various explanations for IPO underpricing. One strand of studies (Beaty and Ritter (1986), Rock (1986)) hypothesizes that IPO underpricing is caused by information asymmetry between various parties involved in the IPO process. Notably, a higher valuation uncertainty is associated with higher expected underpricing to compensate investors for such uncertainty. As such, IPO issues with greater risk or information uncertainty are underpriced more. Lowry and Shu (2002) propose that IPO issues with a higher litigation risk are associated with higher underpricing to compensate for such risk. Michael and Shaw (1994) find that IPOs underwritten by repuTable 3.underwriters experience less underpricing and better long-term performance. Similarly, Megginson and Weiss (1991) indicate that IPOs certificated by venture capital tend to have lower first-day underpricing.

Another strand of studies argues that IPO underpricing results from fads or over-optimism. Theoretical literature has attempted to model the role of irrational sentiment investors in determining the level of IPO underpricing and long-term performance. Ljungqvist, Nanda, and Sigh (2006) develop a theoretical model to explain IPO first-day underpricing and long-term underperformance. In the presence of irrational sentiment investors, IPO issuers compensate regular institutional investors with underpricing of the offer price to compensate for the risk of holding inventory when the hot market may end soon. However, this arrangement is at the expense of sentiment investors who purchased shares after the IPO, resulting in a long-term price reversal. Aggarwal, Krigman, and Womack (2002) develop a theoretical model in which managers of the issuing firm use first-day underpricing to build momentum to attract sentiment investors so that managers can later sell shares at a higher price. Cornelli, Goldreich, and Ljungqvist (2006) present a model that uses pre-IPO grey market prices to measure the sentiment of retail investors. Their theoretical framework indicates sentiment investors in the pre-IPO grey market will overweigh their private signal. As such, high demand from these sentimental investors causes the initial aftermarket price to exceed the fundamental value, followed by a long-term reversal. Derrien (2005) develops a theoretical framework, which indicates that a higher bullishness of sentiment investors will lead to an overpriced offer price and a higher initial return that are deviated from the fundamental value. In general, these theoretical frameworks imply that the existence of exuberant sentiment investors leads to a higher IPO first-day return, followed by a worse long-term performance.

Consistent with these theoretical frameworks, empirical evidence from existing studies have documented that over-optimism sentiment leads to higher IPO first-day underpricing followed by worse long-term performance. Ritter (1991) shows that firms with higher first-day returns tend to underperform in the long run compared to matched counterparts. Ritter and Welch (2002) find that IPO underpricing is more pronounced during the hot market period and follows a worse long-term performance. Additionally, they argue that nonrational or agency problems, rather than information asymmetries, could be the primary factor in explaining IPO first-day underpricing. Using pre-IPO grey market price as a proxy for retail investor sentiment, Cornelli et al. (2006) find that a higher pre-IPO grey market price leads to higher first-day return followed by long-term reversal for European IPOs. Dorn (2009) finds that German IPOs overpaid more aggressively by retail investors in the pre-IPO grey market tends to have higher initial return and worse aftermarket performance. Using data from French IPOs, Derrien (2005) finds that a higher fraction of pre-IPO demand from individual investors leads to a higher first-day return and worse long-term performance. Using Indian IPO data, Clarke, Khurshed, Pande, and Sigh (2016) find that demand from individual investors explains a much more significant portion of IPO underpricing than voluntary underpricing set by underwriters. Other studies empirically examine how pre-IPO information that may influence retail investors' sentiment impacts IPO pricing. Bajo and Raimondo (2017) find that the tone of news media coverage, a significant source of information for retail investors, tends to positively impact the IPO's first-day return. Green and Hwang (2011) find that IPOs in industries with higher expected skewness tend to attract more retail investors, earning a higher first-day return and a worse long-term performance.

Notably, some of these studies use pre-IPO trading patterns, such as grey market prices, to measure the sentiment of retail investors. However, in the US, one needs to be a professional investor or have a special relationship with the IPO managers to purchase IPO shares before the first trading day legally. As such, it is impossible to measure the pre-IPO sentiment of retail investors with premarket trading patterns. As such, I turn to another source of information that may proxy for retail investors' valuation during the pre-IPO period, social media posts. I will use opinions posted on social media platforms to construct attention and sentiment measures and examine their impact on IPO pricing.

3.2.2 Stocktwits.com

StockTwits is a social media platform that facilitates idea sharing between investors. Launched in 2008, this website has become one of the most popular and fast-growing online platforms for investors. Notably, the active user base has grown from around 230,000 in 2012 to around 2 million in 2019. An important feature introduced by Stocktwis is to use a hashtag with a Stock ticker symbol for users to share ideas regarding a specific company more conveniently. One can easily find all recent posts regarding a specific company by typing in the stock ticker symbol in the search bar. As shown by Figure 3.1, one can view all recent posts related to Apple Inc. by typing in its ticker symbol "APPL" into the search bar. Twitter later adopted this innovative idea of hashtags. Different from platforms like Seeking Alpha, opinions posted on Stocktwits are not subject to any editorial review or changes. As such, posts on this platform can vastly reflect the opinions of unprofessional retail investors, even though these opinions are often biased and contain much misinformation.

Another essential feature of Stocktwits is that it allows users to tag their posts as "Bullish" or "Bearish." As shown by Figure 3.2, before making a post, the user can tag the post as "Bullish" or "Bearish" by clicking the buttons on the bottom right. The user can choose not to tag his or her post. Since information on social media platforms contains much noise, this tagging feature enables researchers to target bullish or bearish sentiments more easily. Additionally, these tagged posts can be used as benchmarks for natural language processing (NLP) analysis to classify untagged posts.

Based on the theoretical framework of prior studies, I hypothesize that greater investor attention, as measured by the number of posts during the pre-IPO period, induces a higher IPO first-day return and a long-term reversal. This is because higher investor attention is expected to be associated with a higher demand from retail investors. A more optimistic social media sentiment also induces higher initial returns and worse long-term performance. A more optimistic sentiment is associated with a higher valuation of retail investors. Therefore it is more likely that informed investors will sell overvalued IPO shares to retail investors when the sentiment is overoptimistic.

3.3 Empirical Specification

The sample in this study consists of 387 IPOs from 2016 – 2019. To provide a concise way of estimating the impact of investor attention and sentiment on IPO pricing, I start with an empirical specification as the following:

$$Y_i = \beta_0 + \beta_1 \, Posts\_Numberi + \beta_2 \, Bullish\_Ratio_i + \beta_3 \, Bearish\_Ratio_i + X_i + \epsilon_i \tag{9}$$

where the dependent variable Y is (1) IPO underpricing at IPO day, as measured by either the percentage difference between the offer price and first-day opening (closing) price for firm i or (2) the long-term return t days since the first trading day, as measured by the percentage difference between the closing price for firm i at first trading day and the closing price t days after. The main independent variables are $Posts\_Number$, $Bullish\_Ratio$, and $Bearish\_Ratio$. Measured by the number of posts containing the tick symbol of firm i in the 90 days window before the first trading day, $Posts\_Number$ measures the investor attention grabbed by the IPO of firm i. $Bullish\_Ratio$ and $Bearish\_Ratio$ measure the bullish and bearish sentiment conveyed by their social media posts on Stocktwits. These sentiment measurements are calculated as the ratio of posts tagged as "Bullish" or "Bearish" by the investor in the total posts prior to the first trading day.

Moreover, X represents several control variables commonly used to explain IPO underpricing. This set of control variables includes several firm-level variables that measure the characteristics of issuing firm, including the log number of proceeds, log number of issued shares, age of the issuing firm, venture capital dummy, and high-tech industry dummy. The variable IPO proceeds measure the amount of money collected from the offering, a proxy for the size of the IPO. The number of issued shares measures the supply of IPO offerings and, therefore, will be an essential determinant of IPO pricing. Previous studies (Chambers and Dimson (2009) and Ritter (1991)) have shown that younger firms have a higher IPO underpricing to compensate for their higher risk and uncertainty. Existing studies (Lee and Wahal (2003) and Gompers (1996)) show that IPOs backed by venture capital (VC) tend to exhibit higher underpricing on the first trading day since VCs are willing to bear the cost of underpricing to establish a reputation of being able to make firms public. Due to their higher technological and valuation uncertainty, high-Tech IPOs are involved with higher risk and therefore need a larger expected underpricing to compensate for this risk (Loughran and Ritter (2004). Moreover, existing studies (Carter and Manaster (1990), Megginson and Weiss (1991), Carter, Dark, and Singh (1998) and Loughran and Ritter (2004)) have documented that underwriter prestige has played an essential role in IPO pricing while documented different empirical evidence on this impact. Carter and Manaster (1990), Megginson and Weiss (1991), and Carter et al. (1998) argue that IPOs with more prestigious underwriters have a higher reputational capital commitment. Therefore, investors demand lower expected underpricing for compensating risk. Loughran and Ritter (2004) argue that firms seeking higher analyst coverage with prestigious underwriters increase the market power of these underwriters to increase IPO underpricing. Another essential control variable is price revision. However, due to the massively missing data in SDC, it is practically impossible to construct it for this sample of IPOs.

Moreover, I include an exchange dummy that equals one if the IPO is listed on NYSE and 0 otherwise. Traditional firms that mainly focus on the bottom line and pay dividends tend to list on NYSE, while more growth-oriented firms tend to list on Nasdaq. This difference in the preferred exchange between different types of firms may have resulted in different IPO outcomes. To control for hot market condition, the empirical specification includes a control variable that measures the 3-week cumulative market return before each IPO's first trading day. A better market condition, as measured by higher cumulative market return, may attract more investors into the first trading day of IPO, resulting in a higher first-day price run-up.

Moreover, the specification includes industry fixed effect to control for any unobservable industry-invariant patterns of IPO pricing. Year fixed effect is included in the specification since IPO patterns may vary by year due to changing economic and financial market conditions. More importantly, the year-fixed effect helps control the increasing number of posts for IPOs in recent years as Stocktwits gain more popularity among investors. Lastly, ϵ refers to the error term.

According to the hypothesis developed previously, I expect investor attention and sentiment to be positively correlated with IPO first-day return and negatively correlated with long-term performance. That said, the number of posts, which measures investor attention, is expected to positively impact the IPO first-day price run-up and negatively impact long-term return. Similarly, the "Bullish" ("Bearish") tagged posts ratio, which measures bullish (bearish) investor sentiment, is expected to have a positive (negative) impact on first-day return and a negative (positive) impact on long-term performance. Therefore, when the dependent variable measures the first-day price run-up, the coefficient β_1 and β_3 are expected to be positive, while the coefficient β_2 is expected to be negative. However, when the dependent variable measures the long-term return, the coefficient β_1 and β_3 are expected to be negative, while the coefficient β_2 is expected to be positive.

3.4 Data and Summary Satistics

3.4.1 Data and Variable Construction

Data on IPO information is collected from SDC Platinum. For each IPO, I collect data on the offer price, proceeds, number of shares, lead underwriter, firm industry, and IPO date. Additionally, from SDC, I acquired information on whether the firm is backed by venture capital or belongs to the high-tech industry to construct VC and high-tech dummies. Data on IPO date and firm industry are used to construct the year and industry dummies, respectively. Data on the underwriter's reputation is collected from the spreadsheet provided on the personal website of Jay Ritter and matched with the lead underwriter for each IPO. Data of firm age is also acquired from the personal website of Jay Ritter.

The daily stock price data is collected from CRSP. For each IPO, I collect the first-day opening and closing price and match them with the offer price collected from SDC to calculate the first-day price run-up. Additionally, from CRSP, I collect closing price data t days after the first trading day to measure the long-term return, which is calculated as the percentage between the closing prices at day one and day t. Moreover, daily market return is collected from CRSP to calculate the cumulative return in the 3-week window prior to the first trading day.

Data of social media posts is collected by web-scraping posts on Stocktwits. For each IPO, I collect all posts on Stocktwits that contain its ticker symbol in the 90 days window prior to its first trading day, according to the IPO date collected from SDC. Then for each ticker symbol, the number of posts is calculated as the sum of all posts during the 90 days window. For each IPO, the ratio of "Bullish" ("Bearish") tagged posts is calculated as the number of posts tagged as “Bullish” (“Bearish”) divided by the total number of posts that contain the ticker symbol during the 90 days window.

3.4.2 Summary Statistics

Table 3.1 presents the summary statistics of the constructed variables. The first-day price run-ups are generally positive, as indicated by the positive mean of the opening return and first-day return (0.148 and 0.174, respectively). The critical dependent variables show a very uneven distribution. The number of pre-IPO posts has a mean of 29, a standard deviation of 128, and a median of 9. This suggests that most IPOs have gained relatively low investor attention during the pre-IPO period. To accommodate this uneven distribution of number of posts, I take the log of it in the main empirical specification. Several noTable 3.IPOs have gained a relatively significant level of investor attention on social media. The maximum number of pre-IPO posts in the sample was for SNAP in 2017, with a total post of 1720. The IPOs of UBER and LYFT in 2019 gained 1581 and 816 pre-IPO posts during the pre-IPO period, respectively. In total, 15 firms have gained over 100 posts on Stocktwits during the 90 days window before their first trading day.

The sentiment measurements are also unevenly distributed. The “Bearish” posts ratio has a mean of 0.01, a standard deviation of 0.04, a median of 0, and a maximum of 0.33. Additionally, the “Bullish” posts ratio has a mean of 0.10, a standard deviation of 0.14, a median of 0, and a maximum of 0.80. The summary statistics of these two variables suggest that most of the posts during the pre-IPO period are tagged as neither “Bullish” nor “Bearish.” As such, the main regression expressed by equation (1) can include both sentiment measurements since they do not completely complement each other. Additionally, the mean of “Bullish” posts is ten times that of “Bearish” posts, suggesting that investors are more willing to share optimism than pessimism on online platforms. This is consistent with the notion that investors have a behavioral bias toward broadcasting positive information, as noted in Cookson and Niessner (2020).

3.5 Empirical Results and Analysis

3.5.1 Baseline Results

Table 3.2 shows the baseline results of the main specification expressed by equation (1). Column (1) and column (2) present the empirical results for the first day price run up. The dependent variables are the first day opening return and first day closing return, respectively. As mentioned previously, opening return is calculated as the percentage difference of the first day opening price from the IPO offer price. First day closing return is calculated as the percentage difference of the first day closing price from the IPO offer price. The coefficients on the control variables are mostly consistent with those in existing studies. Notably, underwriter reputation has a negative though insignificant coefficient. This suggests that in recent years, the impact of underwriter reputation is more consistent with certification hypothesis that IPO with more repuTable 3.underwriters bear less risk to be compensated with underpricing. Consistent with the hypothesis, empirical results show that number of posts during pre-IPO period, which measures investor attention to the IPO, tend to increase the first day price run up. For the two measurements of first day price run up, the coefficients on the number of posts are 0.03934 and 0.06382 respectively, and both are significant at 1% of significance level. The magnitudes of coefficients suggest that a 10% increase in the number of pre-IPO posts on Stocktwits lead to an increase of 0.16% and 0.26% on the first day opening return and first day closing return, respectively. The results on number of pre-IPO posts indicate that a higher investor attention on social media lead to a higher first day price run up, which is consistent with the prediction of my hypothesis.

Consistent with the hypothesis, the coefficients on the two sentiment measurements indicate that a more optimistic investor sentiment leads to a greater IPO underpricing. For both the first day opening return and closing return, the coefficients on the “Bearish” post ratio are negative and significant at 1% significance level, suggesting that a higher portion of pessimistic sentiment leads to lower first day IPO price run up. The magnitude of the coefficients on “Bearish” posts ratio are -0.8243 and -1.2971 for the two measurement of first day price run up. This indicates that a one standard deviation (0.04) increases in the “Bearish” post ratio during pre-IPO period leads to a 3.3% and 5.2% decrease in the opening and closing return at the first trading day.

The coefficients on “Bullish” post ratio are positive for both measures of first day price run up, significant at 5% significance level for first day opening return, yet only significant at 10% level for first day closing return. Consistent with the prediction of hypothesis, the sign of the coefficients indicates that a higher portion of optimistic sentiment leads to a higher first day price run up. Admittedly, the evidence is relatively weak in significance level for the result with first day closing return as dependent variable. The magnitudes of coefficients are 0.1634 and 0.2172 respectively for first day opening and closing return, respectively. This indicates that a one standard deviation (0.14) increases in the “Bullish” posts ratio during the pre-IPO period leads to a 2.2% and 3.1% increase in the first day opening return and closing return, respectively. Notably, the magnitude of coefficients with “Bullish” posts ratio are significantly smaller than those for “Bearish” posts ratio (0.1634 vs 0.8243 and 0.2172 vs 1.2971). The magnitude of coefficients for “Bullish” posts ratio are around 1/5 to 1/6 of that associated with “Bearish” posts ratio.

Column (3), (4) and (5) represent the empirical results where the dependent variables are the return 30, 60 and 90 days after the first trading day. The dependent variables are calculated as the percentage difference between the closing price at 30th, 60th and 90th trading day and the closing price at the first trading day. The purpose of these specifications with long term return as dependent variable is to examine whether social media attention and sentiment have caused overpricing followed by reversal afterward. If the sign of coefficients is in contrast with those for the first day price run up, then the evidence supports the notion that social media activities result in overpricing followed by reversal. For number of posts, although the coefficients become negative for return at 30th and 90th day after first trading day, they are statistically insignificant. As such, for returns in these relatively short periods after the first trading day, there is no significant empirical evidence suggesting pre-IPO attention have caused overpricing in the first trading day. For “Bearish” post ratio, the coefficients on all three specifications become positive, which are in contrast with the coefficients for the first day price run up. Moreover, the coefficients for returns 60 and 90 days after the first trading day are statistically significant at 10% and 1% significance level. Specifically, the magnitudes of these coefficients are 1.2778 and 1.8596, respectively. This suggests that a one standard deviation (0.04) increase in “Bearish” post ratio results in a 5.12% and a 7.44% decrease in the returns 60 and 90 days after the first trading day, respectively. Notably, the reversal in the long run has a similar or greater magnitude than the price decrease in the first trading day (5.12 % vs 5.2%, 5.12% vs 7.44%), indicating that social media sentiment has caused overpricing with significant economic magnitude. For “Bullish” post ratio, the coefficients are negative but lack statistical significance. As such, I have found no statistically significant evidence suggesting that optimistic sentiment on social media have caused reversal during the 30-, 60-, and 90-days window after the first trading day.

In general, the baseline results indicate that a higher volume of investor attention or more optimistic sentiment on social media leads to a higher IPO first day price run up. Specifically, the results suggest that negative sentiment has a more pronounced impact on the first day price run up, since the coefficients on “Bearish” posts ratio have a larger magnitude and statistical significance than those on “Bullish” posts ratio. Particularly, the coefficients for “Bearish” posts ratio are around 4-5 times larger than that of “Bullish” posts ratio in terms of magnitude. There are several explanations for this difference of impact between “Bullish” and “Bearish” posts. Firstly, as mentioned previously, investors generally have the tendency to broadcast optimistic opinions rather than pessimistic opinions, as suggested by the much larger fraction of “Bullish” posts than “Bearish” posts. Therefore, it is likely that they share pessimistic opinions only if such emotions are strong and therefore these “Bearish” posts reflect a much stronger negative sentiment than the positive sentiment reflected by “Bullish” posts. Another possible explanation is that negative comments, though in a smaller portion than positive comments, are more likely to grab the attention of other investors than positive comments and therefore have a stronger sentiment spillover on the platform. These two reasons may explain why "Bearish" posts ratio imposes a more pronounced impact on first trading day price run up than "Bullish" posts, despite their relatively smaller fraction in total number of posts. Similarly, in the relatively short terms after the first trading day (30/60/90 days), "Bearish" posts ratio tends to significantly predict the price reversal during these periods while "Bullish" posts ratio has no statistically significant impact. Notably, though investor attention measured by number of posts positively predicts first day price run up, it has no statistically significant predictive power on the price reversal in the short term after the first trading day. Da, Engelberg and Gao (2011) finds that investor attention, as measured by Google search volume leads to higher IPO first day return and long-term reversal. The empirical findings in this study suggests that though investor attention increases IPO first day price run up, the overpricing and the following price reversal is mostly driven by sentiment rather than attention.

3.5.2 Long-Term Return

To further examine whether social media sentiment causes overpricing followed by price reversal, I have run the regression with the dependent variable replaced with return in a long-time horizon after the first trading day. Specifically, the dependent variables are returns 120-,150-,180-,240-, and 360 days after the first trading day. Similarly, return t days after the first trading day is calculated as the percentage difference between the closing price at the first trading day and the closing price t days after. Table 3.3 presents the result with dependent variables replaced as returns in these longer time horizons.

Generally, the results indicate that investor sentiment leads to price reversal in the long run, which implies that the initial price run-up was the result of sentimental overpricing. Like the results in the relatively short term after the first trading day, there is no significant coefficient with a negative sign for the number of posts. Rather, when the dependent variables are returned 240 and 360 days after the first trading day, the positive coefficients on the number of posts become weakly significant. Therefore, there is no empirical evidence suggesting that investor attention has caused any overpricing followed by reversal in the long run.

Notably, in contrast to the price reversal in the relatively short term after IPO, the reversals, in the long run, are caused mainly by optimistic sentiment rather than pessimistic sentiment. On the one hand, the coefficients on the "Bearish" posts ratio are statistically significant only for return on the 180th day after the first trading day, with a magnitude of 2.7887 and significant at a 1% significance level. This implies that one standard deviation (0.04) increase in the ratio of "Bearish" posts leads to an 11.2% increase in the long-term return 180 days after the first trading day. The magnitude (11.2%) of this price reversal is with considerable economic significance, given that the first-day closing return has an average of 17.2% and one standard deviation (0.04) increase in "Bearish" posts ratio leads to a 5.2% decrease in first-day closing return.

On the other hand, the "Bullish" post ratio tends to have a strong negative impact on the long-term return of IPOs, despite their relatively weaker coefficients on the first-day price run-up. The coefficients are negative and statistically significant for all long-term return measures except for the return on the 360th day after the first trading day. Specifically, the coefficients have magnitudes of -0.5960, -0.6340, -0.7773, and -0.6978 for returns on 120, 150, 180, and 240 days after the first trading day and are statistically significant at 1% level for the first three columns (120, 150, 180 days) and 5% significance level for the fourth measures (240 days). Consistent with the results for the first-day price run-up, coefficients on the "Bullish" posts ratio are significantly smaller than that of the "Bearish" posts ratio in terms of magnitude (-0.60, -0.63, -0.78, -0.70 vs. 2.79). Nonetheless, the coefficients on the "Bullish" posts ratio have increased by around two times bigger in magnitude (-0.60, -0.63, -0.78, -0.70 vs. 0.22) compared to the coefficient when the first-day closing return is the dependent variable. As such, one standard deviation (0.14) increase in "Bullish" posts ratio leads to an 8.4% (8.8%, 11%, and 9.8%) decrease in long-term returns 120 (150, 180, and 240) days after the first trading day. When compared to the increase in the first-day run-up (8.4%, 8.8%, 11% and 9.8% vs. 3.1%) caused by a one standard deviation increase in "Bullish" posts ratio, the reversal in long-term return is of significant economic magnitude.

In general, the results for the long-term return after the first trading day are consistent with the results for the relatively short-term return, except that the price reversal is mainly driven by optimistic rather than pessimistic sentiment. The results show no empirical evidence that investor attention leads to any price reversal in the long run, indicating that the overpricing is mainly driven by sentiment rather than attention.

3.6 Robustness Check

3.6.1 Results for IPOs with Less than 100 Posts

As indicated by the summary statistics, the number of posts during the pre-IPO period is very unevenly distributed across firms. The variable number of posts has a median of 9, a mean of 29.13, a standard deviation of 127.64, and a maximum of 1720. This wildly uneven distribution of the number of posts raises the concern that the result may be driven by firms with an unusually high number of posts, i.e., firms that grab an extraordinarily high volume of investor attention during the pre-IPO period. Several noTable 3.IPOs with an unusually high number of pre-IPO posts are UBER (1581 posts), SNAP (1720 posts), and LYFT (816 posts). It is possible that these IPOs may cause the result of sentiment-driven overpricing with a high volume of investor attention. To examine the generality of the empirical results, I run the regression on a more restricted sample that consists of only firms with less than 100 pre-IPO posts. This results in a restricted sample of 373 firms.

Table 3.4 and 5 show the empirical results with this restricted sample for short-run and long-run performance, respectively. Generally, the results for this restricted sample are consistent with the ones conducted in the full sample. The empirical results in Table

3.4 show that the “Bearish” (“Bullish”) posts ratio tends to decrease (increase) the first-day return. The magnitudes of the coefficients have decreased for both the “Bearish” and “Bullish” posts ratio, but not significantly (-0.65 vs -0.82, -1.18 vs -1.30, 0.16 vs 0.15, 0.22 vs 0.21). The statistical significance of the coefficients has decreased. Notably, the coefficient on the “Bullish” posts ratio for first-day opening return becomes significant only at a 10%

Table 3.5 shows the result for long-term performance with this restricted sample. Consistent with the results with the whole sample, the results imply that a higher “Bearish” (“Bullish”) posts ratio leads to a lower (higher) long-term performance. Notably, though the coefficients on the “Bearish” posts ratio are statistically significant only for returns 180 days after the first trading day for the full sample, they become statistically significant at a 10% level for returns 120, 150, and 240 days after the IPO for this restricted sample. Moreover, the magnitudes of these coefficients on the “Bearish” posts ratio have increased compared to the ones with the full sample. For the “Bullish” posts ratio, the coefficients are generally similar to the ones with a full sample, while the coefficients’ magnitude has increased to some extent.

Overall, the empirical results with this restricted sample of firms with less than 100 pre-IPO posts reject the possibility that firms with high volumes of investor attention mainly drive the results. Consistent with the full sample results, the “Bearish” (“Bullish”) posts ratio tends to decrease (increase) the first-day price run-up while increasing (decreasing) the long-term performance. While the coefficients are somewhat less pronounced for the first day and short-term return after IPO, they are generally more substantial for long-term performance.

3.6.2 Market Benchmarked Return

Another concern is that market movement may have driven the pricing of IPOs, especially the long-term performance after the first trading day. Additionally, market conditions may substantially influence investor attention and sentiment. Since market price movements are likely to be highly serially correlated before and after the first trading day, there is the concern that the results of investor attention or sentiment on IPO pricing or long-term performance may be confounded with market price movement. To address this issue, I have used returns benchmarked by market return as dependent variables for the regression. Specifically, the benchmarked first-day closing return is calculated as the first-day closing return minus the first trading day market return. The benchmarked return t days after the first trading day is calculated as the return minus the market return t days afterward. Table 3.6 and Table 3.7 present the results with market benchmarked return as the dependent variables. Generally, the results with the market benchmarked return are consistent with the baseline results presented in Table 3.2 and 3. The empirical results imply that pessimistic (optimistic) sentiment leads to a lower (higher) first-day return and a higher (lower) long-return. The coefficients did not change significantly in terms of magnitude and statistical significance.

3.7 Further Analysis

3.7.1 Machine Learning Approach for Untagged Posts

Another issue for this study is that there is a significant portion of posts are tagged as neither "Bullish" nor "Bearish." If the users did not specifically tag their posts as "Bearish" or "Bullish," then it is likely that their sentiment on these posts is not very strong. However, Cookson and Niessner (2020) indicate that these untagged posts on Stocktwits are, in fact, very "Bullish" or "Bearish" based on their textual content. As such, even though the emotion contained in these posts may be less pronounced, they may still reflect a significant portion of investor sentiment. Therefore, I intend to examine the potential impact of these untagged posts on IPO pricing. Following prior studies (Bartov et al., (2018), Cookson and Niessener (2020) and Leung et al., (2020)) that utilize machine learning techniques to classify untagged tweets, I attempt to measure the sentiment conveyed by these untagged posts and classify them as "Bullish" or "Bearish" through Natural Language Processing (NLP) techniques. Following Leung et al. (2020), I attempt to use FastText, a Word-to-Vector algorithm developed by Facebook AI Research, to classify these untagged posts as "Bullish" or "Bearish."

Specifically, word-to-vector methodology treats each word and multiword phrase as a series of vectors, which store their information as a numerical value and learn the meaning of each single word or word phrase in association with their neighboring words. A critical feature of FastText is that the algorithm can automatically correct irregularly formatted or misspelled words. For example, Tweets may contain typos such as "amplify" when referring to "amplify." Or in other cases, Stocktwits users may misspell "another" as "another." In these cases, FastText can automatically correct these typos by using sub-word information.

To implement the FastText algorithm to learn about the textual information contained in Stocktwits posts, I have included all Stocktwits posts labeled as "Bullish" or "Bearish" by the users in the training sample. I must acknowledge that the training sample is larger than the sample I used to construct sentiment measures in the baseline regression. The training sample contains 32585 Stocktwits posts that are explicitly tagged as "Bullish" or "Bearish" by the users, which is much larger than the sample used to construct the sentiment measures (11223 posts in total, with 449 "Bearish" posts and 1572 "Bullish" posts). I selected this more extensive sample of posts since the sample used for variable construction is too small to conduct meaningful textual training. The training sample consists of 24566 "Bullish" posts and 8019 "Bearish" posts. To resolve the issue of the unbalanced sample, I have randomly undersampled the number of "Bullish" posts so that there are equal numbers of "Bullish" and "Bearish" posts in the training sample.

Furthermore, the training sample is randomly split into a training set (70%) and a validation set (30%). The training set is used to conduct a supervised training of the multinomial logistic model, while the latter is used to test the accuracy of the trained model. Specifically, I set up the dimension of the training vector as 500, the size of the context window (neighboring) as 10, and the epoch as 25. After applying the trained model on the validation set, I have achieved an accuracy of 72%, with an accuracy of 72% for both "Bullish" and "Bearish" posts. The accuracy in the training set is 93%.

After the training of the FastText model, I applied the trained model to the untagged posts in the sample used to construct sentiment measures. As all the posts are classified as either “Bullish” or “Bearish,” I construct the sentiment measures following the previous procedures. For each IPO, the pessimistic (optimistic) sentiment is measured by the ratio of “Bearish” (“Bullish”) posts in the total pre-IPO posts. Table 3.8 shows the summary statistics for the newly constructed sentiment measures. Notably, even after all untagged posts are classified, the “Bullish” posts ratio still dominates the “Bearish” posts ratio, with an average of 0.72 vs. 0.28 and a median of 0.75 vs 0.25. Since all posts are classified, the ratios of “Bearish” and “Bullish” posts are linearly correlated, and therefore, only one of them needs to stay in the regression. I have included only the “Bearish” posts ratio in the regression.

Table 3.9 and 10 show the regression results with short-term and long-term returns, respectively. A higher “Bearish” posts ratio leads to a lower first-day return, as indicated by the negative and significant coefficients in columns (1) and (2) of Table 3.9. Unlike previous findings, no significant coefficients were found for returns 30, 60, and 90 days after the first trading day. However, in Table 3.10, the positive and significant coefficients imply that a more optimistic sentiment leads to underperformance on the 120th, 150th, and 180th day after the first trading day. Notably, the coefficients are much smaller in magnitude when compared to the previous findings, where only posts inexplicably tagged by the users are used to construct sentiment measures (-0.1038 vs. -0.8243, -0.1394 vs. -1.2971 and 0.2737 vs. 2.7887). This decrease in coefficient magnitude may suggest that sentiments conveyed by untagged posts are less strong than those inexplicably tagged by users. Generally, the empirical results are consistent with the previous findings that a more optimistic social media sentiment during the pre-IPO period leads to a higher first-day price run-up and a worse long-term performance. Please see Table 3.A.1-3.A.4 in Appendix for results with a restricted sample of firms with no more than 100 posts and market benchmarked returns. In general, the results are similar to the ones in Table 3.9 and 10.

3.7.2 Turnover Rate

If the IPO's first-day price run-up caused by social media sentiment results from overoptimistic retail investor demand, then informed institutional investors are likely to have sold a large number of IPO shares to retail investors in the short period after the IPO. As such, I expect a more optimistic pre-IPO social media sentiment to be followed by a higher turnover rate shortly after the IPO. Following Cornelli et al. (2006), the turnover rate is calculated as the number of shares traded divided by the total number of issued shares. Table 3.11 shows the result when the log of the first-day and first-week turnover rate are the dependent variables. The number of pre-IPO posts positively predicts a higher turnover rate in the first trading day and the first week after IPO, as indicated by the positive and significant coefficients. However, though the sign of coefficients is the same as my expectation, the two sentiment measures tend to have no statistically significant effect on the turnover rate.

Table 3.12 shows the result with machine learning classified sentiment measurement. The number of posts still has a positive and significant coefficient. Sentiment measurement, as measured by the ratio of "Bearish" posts, tends to decrease the turnover rate on the first trading day and the first week after IPO, as indicated by the negative and significant coefficients. The results with machine learning classified sentiment measurement indicate that a more optimistic sentiment leads to a higher turnover rate shortly after IPO, implying that the demand of overoptimistic retail investors has driven the overpricing.

3.8 Conclusion

In this paper, I empirically examine the impact of social media attention and sentiment on IPO pricing. I collect pre-IPO user post data from Stocktwits for IPOs during 2016 -2019 and use user-tagged posts to construct bullish and bearish sentiment measurements. Using these social media sentiment measurements as proxies for retail investors' valuation for IPO shares, I attempt to empirically examine the theoretical prediction proposed by prior studies that over-optimism of exuberant sentiment investors tends to cause initial overpricing of IPO shares that deviates from the fundamental values and is then followed by a long-term reversal. In general, the empirical results are consistent with the predictions of prior theoretical frameworks. A more bullish (bearish) pre-IPO sentiment predicts a higher (lower) IPO first-day price run-up and a worse (better) long-term performance. However, the empirical results suggest that while investor attention tends to push up first-day return, it has no significant impact on long-term reversal. This suggests that the overpricing issue is mainly driven by sentiment rather than attention. The results are robust to using market benchmarked return as dependent variables or restricting the sample to firms with a relatively smaller number of posts.

Additionally, I have attempted to classify untagged posts with NLP machine learning techniques since prior study indicates that these posts are, in fact, very bullish or bearish. Including these classified posts into the construction of sentiment measurements, empirical results are generally similar to the main results, where more optimistic social media sentiment leads to higher first-day price run-up and worse long-term returns. However, the magnitude of sentiment impact becomes less pronounced since the sentiment contained in these untagged posts may be weaker than the tagged ones. Moreover, using sentiment measurements constructed by these classified posts, I find that more optimistic social media sentiment leads to a higher turnover rate during the short time after IPO. This result suggests that over-optimism may have enabled informed investors to sell more overpriced shares to exuberant retail investors. This implies that the first-day price run-up and long-term reversal result from retail investor demand.

This study has added to the existing literature on social media and the financial market by examining the impact of social media information flow on IPO pricing. Additionally, by using social media sentiment as a proxy for retail investor valuations, this study empirically examines the impact of sentiment investors on IPO overpricing in the US financial market, where pre-IPO trading is limited to professional investors. More importantly, it provides insight into using online platform data to predict IPO pricing patterns, which can help provide investment guidance for practitioners.

References

Aggarwal, R. K., Krigman, L., & Womack, K. L. (2002). Strategic IPO underpricing, information momentum, and lockup expiration selling. *Journal of Financial Economics*, 66(1), 105-137.

Bajo, E., & Raimondo, C. (2017). Media sentiment and IPO underpricing. *Journal of Corporate Finance*, 46, 139-153.

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter help predict firm-level earnings and stock returns?. *The Accounting Review*, 93(3), 25-57..

Beatty, R. P., & Ritter, J. R. (1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15(1-2), 213-232.

Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *The Accounting Review*, 89(1), 79-112.

Carter, R. B., Dark, F. H., & Singh, A. K. (1998). Underwriter reputation, initial returns, and the long-run performance of IPO stocks. *The Journal of Finance*, 53(1), 285-311.

Carter, R., & Manaster, S. (1990). Initial public offerings and underwriter reputation. *The Journal of Finance*, 45(4), 1045-1067.

Chambers, D., & Dimson, E. (2009). IPO underpricing over the very long run. *The Journal of Finance*, 64(3), 1407-1443.

Chen, H., Hwang, B. H., & Liu, B. (2019). The Emergence of'Social Executives' and Its Consequences for Financial Markets. *Working Paper.*

Chen, H., De, P., Hu, Y., & 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.

Clarke, J., Khurshed, A., Pande, A., & Singh, A. K. (2016). Sentiment traders & IPO initial returns: The Indian evidence. *Journal of Corporate Finance*, 37, 24-37.

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.

Cornelli, F., Goldreich, D., & Ljungqvist, A. (2006). Investor sentiment and pre-IPO markets. *The Journal of finance*, 61(3), 1187-1216.

Curtis, A., Richardson, V. J., & Schmardebeck, R. (2014). Investor attention and the pricing of earnings news. *Handbook of Sentiment Analysis in Finance*, Forthcoming.

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

Derrien, F. (2005). IPO pricing in "hot" market conditions: Who leaves money on the table?. *The Journal of Finance*, 60(1), 487-521.

Green, T. C., & Hwang, B.-H. (2012). Initial Public Offerings as Lotteries: Skewness Preference and First-Day Returns. *Management Science*, 58(2), 432–444.

Gompers, P. A. (1996). Grandstanding in the venture capital industry. *Journal of Financial Economics*, 42(1), 133-156.

Jung, M. J., Naughton, J. P., Tahoun, A., & Wang, C. (2018). Do firms strategically disseminate? Evidence from corporate use of social media. *The Accounting Review*, 93(4), 225-252.

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, 53(2), 367-404.

Lee, P. M., & Wahal, S. (2004). Grandstanding, certification and the underpricing of venture capital backed IPOs. *Journal of Financial economics*, 73(2), 375-407.

Leung, W. S., Wong, G., & Wong, W. K. (2019). Social-media sentiment, portfolio complexity, and stock returns. *Working Paper.*

Ljungqvist, A., Nanda, V., & Singh, R. (2006). Hot markets, investor sentiment, and IPO pricing. *The Journal of Business*, 79(4), 1667-1702.

Loughran, T., & Ritter, J. (2004). Why has IPO underpricing changed over time?. *Financial Management*, 5-37.

Lowry, M., & Shu, S. (2002). Litigation risk and IPO underpricing. *Journal of Financial Economics*, 65(3), 309-335.

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012, August). Correlating S&P 500 stocks with Twitter data. In *Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research* (pp. 69-72).

McGurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance*, 44, 458-485.

Megginson, W. L., & Weiss, K. A. (1991). Venture capitalist certification in initial public offerings. *The Journal of Finance*, 46(3), 879-903.

Michaely, R., & Shaw, W. H. (1994). The pricing of initial public offerings: Tests of adverse-selection and signaling theories. *The Review of Financial Studies*, 7(2), 279-319.

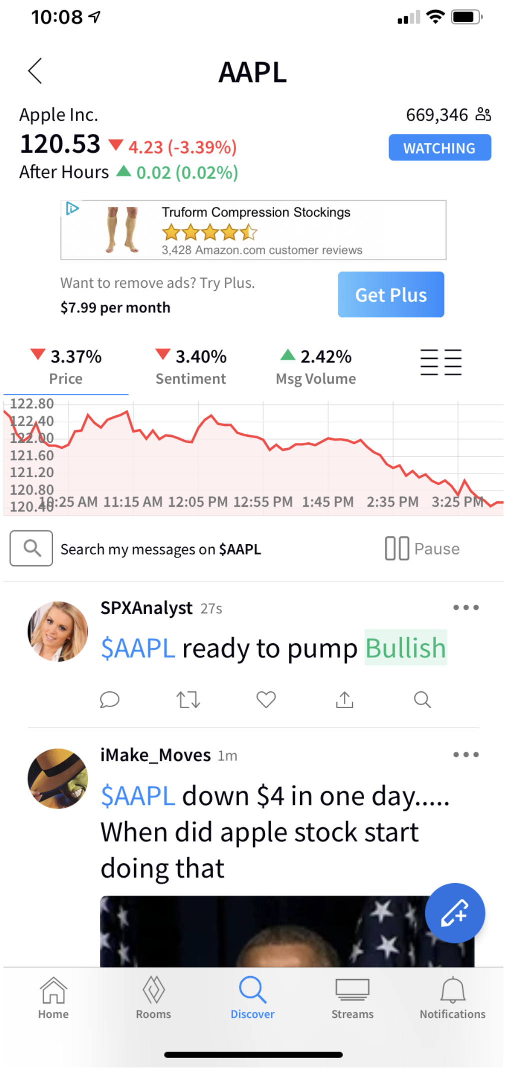
Ritter, J. R. (1991). The long-run performance of initial public offerings. *The Journal of Finance*, 46(1), 3-27.

Ritter, J. R., & Welch, I. (2002). A review of IPO activity, pricing, and allocations. *The Journal of Finance*, 57(4), 1795-1828.

Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics*, 15(1-2), 187-212.

Wu, D. (2019). Does social media get your attention?. *Journal of Behavioral Finance*, 20(2), 213-226.

Tables and Figures



10:08

AAPL

Apple Inc.

120.53 ▼ 4.23 (-3.39%)

After Hours ▲ 0.02 (0.02%)

669,346

Truform Compression Stockings

3,428 Amazon.com customer reviews

Want to remove ads? Try Plus.

$7.99 per month

▼ 3.37% Price

▼ 3.40% Sentiment

▲ 2.42% Msg Volume

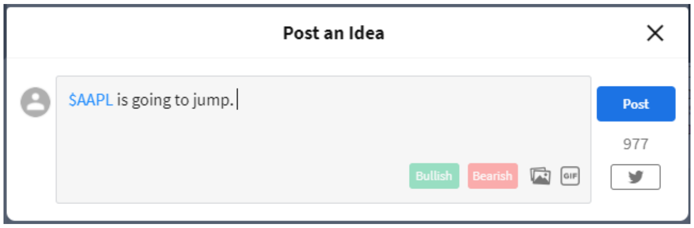
SPXAnalyst 27s

$AAPL ready to pump Bullish

iMake_Moves 1m

$AAPL down $4 in one day..... When did apple stock start doing that

Figure 3.1: Search Stock by Cashtag on Stocktwits



Post an Idea

$AAPL is going to jump.

Figure 3.2: Customize Post with "Bullish" or "Bearish" Tag Option

Table 3.1: Summary Statistics

This table reports the summary statistics of input variables for regression analysis

	N	mean	sd	min	max	p50
Opening Return	387	0.15	0.22	-0.36	0.94	0.08
First Day Return	379	0.17	0.31	-0.41	2.31	0.09
Number of Posts	387	29.13	127.64	1.00	1720.00	9.00
Bearish Tag Ratio	387	0.01	0.04	0.00	0.33	0.00
Bullish Tag Ratio	387	0.10	0.14	0.00	0.80	0.00
Age	387	14.61	21.85	0.00	166.00	9.00
Proceeds	387	250.08	542.79	3.50	8100.00	108.50
3_Week_Mkt	387	0.02	0.03	-0.10	0.13	0.02
Num_Share	387	14264411.08	23205669.55	350000.00	2.07e+08	7350000.00
Reputation	387	7.41	3.68	-9.00	9.00	9.00
VC Dummy	387	0.54	0.50	0.00	1.00	1.00
Exchange Dummy	387	0.72	0.45	0.00	1.00	1.00
Tech Dummy	387	0.53	0.50	0.00	1.00	1.00

Table 3.2: Baseline Empirical Results

This table reports the baseline empirical results with OLS estimation of Equation (1). The dependent variables are opening return, first day return, 60 days return, and 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Opening_Return	(2) First_Day_Return	(3) Return_30_Days	(4) Return_60_Days	(5) Return_90_Days
Log Number of Posts	0.03934***	0.06382***	-0.009253	0.03763	-0.0005583
	(3.54)	(3.87)	(-0.34)	(1.28)	(-0.02)
Bearish Tag Ratio	-0.8243***	-1.2971***	0.4591	1.2778*	1.8596***
	(-2.84)	(-3.12)	(0.67)	(1.72)	(2.70)
Bullish Tag Ratio	0.1634**	0.2172*	-0.1492	-0.2412	-0.2742
	(2.00)	(1.84)	(-0.77)	(-1.14)	(-1.40)
Age	-0.0003817	-0.0009543	0.0001240	0.0001225	0.0002617
	(-0.67)	(-1.03)	(0.08)	(0.07)	(0.17)
Proceeds	0.1430***	0.1546***	0.02342	0.04346	0.1157
	(4.41)	(3.32)	(0.31)	(0.52)	(1.50)
3_Week_Mkt	0.8062**	1.4034***	-0.6390	-1.3880	-2.4109***
	(2.20)	(2.64)	(-0.74)	(-1.46)	(-2.74)
Shares Number	-0.1491***	-0.1882***	-0.01689	-0.05217	-0.1277
	(-3.98)	(-3.49)	(-0.19)	(-0.54)	(-1.43)
Reputation	-0.003932	-0.007450	-0.02127**	0.002975	0.003608
	(-1.08)	(-1.41)	(-2.46)	(0.31)	(0.41)
VC Dummy	0.03494	0.06480*	-0.005187	-0.02615	0.003546
	(1.29)	(1.65)	(-0.08)	(-0.37)	(0.05)
Exchange Dummy	-0.007399	-0.03663	0.002607	0.04882	0.1033
	(-0.26)	(-0.87)	(0.04)	(0.65)	(1.49)
Tech Dummy	0.04048	-0.05628	0.01654	0.1059	0.1300*
	(1.37)	(-1.32)	(0.24)	(1.39)	(1.84)
Constant	1.7606***	2.3406***	0.2844	0.5664	1.4767
	(3.72)	(3.45)	(0.26)	(0.47)	(1.31)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	387	379	379	379	379
Adjusted R^2	0.201	0.132	-0.031	-0.024	0.038

Table 3.3: Long-Term Return Empirical Results

This table reports the baseline empirical results with OLS estimation of Equation (1). The dependent variables are long-term returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$

	(1) Return_120_Days	(2) Return_150_Days	(3) Return_180_Days	(4) Return_240_Days	(5) Return_360_Days
Log Number of Posts	0.01355	0.01729	0.02171	0.07331*	0.1136*
	(0.49)	(0.54)	(0.57)	(1.66)	(1.89)
Bearish Tag Ratio	1.0851	0.9470	2.7887***	1.3899	0.7978
	(1.55)	(1.18)	(2.89)	(1.25)	(0.53)
Bullish Tag Ratio	-0.5960***	-0.6340***	-0.7773***	-0.6978**	-0.6304
	(-3.00)	(-2.79)	(-2.84)	(-2.21)	(-1.37)
Age	0.0006396	0.0008683	0.001472	-0.0008079	-0.002484
	(0.41)	(0.49)	(0.69)	(-0.33)	(-0.75)
Proceeds	0.01221	0.01832	0.04293	0.1441	0.2861
	(0.16)	(0.20)	(0.40)	(1.16)	(1.65)
3_Week_Mkt	-2.9670***	-2.4656**	-4.4510***	-2.9657**	-3.9219**
	(-3.31)	(-2.40)	(-3.60)	(-2.08)	(-2.01)
Shares Number	-0.003301	-0.02174	-0.06841	-0.1418	-0.3043
	(-0.04)	(-0.21)	(-0.55)	(-0.99)	(-1.54)
Reputation	0.01472*	0.01617	0.01357	0.009210	0.03378*
	(1.66)	(1.59)	(1.11)	(0.65)	(1.73)
VC Dummy	0.003399	-0.02238	0.05872	0.01225	0.02588
	(0.05)	(-0.30)	(0.64)	(0.12)	(0.18)
Exchange Dummy	0.1469**	0.1159	0.1353	0.2216**	0.3218**
	(2.08)	(1.44)	(1.39)	(1.98)	(2.09)
Tech Dummy	0.1353*	0.1196	0.1224	0.02530	-0.1073
	(1.89)	(1.46)	(1.24)	(0.22)	(-0.66)
Constant	-0.2242	0.0004329	0.5523	1.2736	3.2449
	(-0.20)	(0.00)	(0.35)	(0.70)	(1.30)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	378	378	378	378	337
Adjusted R^2	0.050	0.023	0.043	0.052	0.103

Table 3.4: Baseline Empirical Results with IPOs recevied less than 100 Posts

This table reports the baseline empirical results with OLS estimation of Equation (1) in the restricted subsample of IPOs receved less than 100 posts during the pre-IPO period. The dependent variables are opening return, first day return, 60 days return, and 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Opening_Return	(2) First_Day_Return	(3) Return_30_Days	(4) Return_60_Days	(5) Return_90_Days
Log Number of Posts	0.04020***	0.06196***	-0.01343	0.04023	0.008826
	(3.39)	(3.56)	(-0.46)	(1.28)	(0.30)
Bearish Tag Ratio	-0.6466**	-1.1811***	0.3908	1.2291	1.9791***
	(-2.21)	(-2.79)	(0.55)	(1.61)	(2.76)
Bullish Tag Ratio	0.1531*	0.2064*	-0.1615	-0.2820	-0.3179
	(1.91)	(1.77)	(-0.82)	(-1.35)	(-1.61)
Age	-0.0007068	-0.001139	0.00001044	-0.0001389	0.00005017
	(-1.19)	(-1.24)	(0.01)	(-0.08)	(0.03)
Proceeds	0.1588***	0.1768***	0.02305	0.03963	0.1265
	(4.81)	(3.70)	(0.29)	(0.46)	(1.56)
3_Week_Mkt	0.7882**	1.2260**	-0.8100	-1.6772*	-2.6351***
	(2.20)	(2.35)	(-0.92)	(-1.78)	(-2.98)
Shares Number	-0.1568***	-0.1959***	-0.005816	-0.02702	-0.1253
	(-4.16)	(-3.59)	(-0.06)	(-0.28)	(-1.36)
Reputation	-0.004964	-0.009974*	-0.02234**	0.0009964	0.001913
	(-1.38)	(-1.91)	(-2.54)	(0.11)	(0.22)
VC Dummy	0.03114	0.06105	-0.01372	-0.03919	-0.004600
	(1.18)	(1.59)	(-0.21)	(-0.57)	(-0.07)
Exchange Dummy	-0.005354	-0.04320	-0.003696	0.03262	0.09915
	(-0.19)	(-1.03)	(-0.05)	(0.43)	(1.40)
Tech Dummy	0.03736	-0.05436	0.01902	0.1247	0.1459**
	(1.29)	(-1.29)	(0.27)	(1.65)	(2.05)
Constant	1.8118***	2.3922***	0.1473	0.2415	1.4124
	(3.84)	(3.50)	(0.13)	(0.20)	(1.22)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	373	367	367	367	367
Adjusted R^2	0.175	0.115	-0.033	-0.018	0.051

Table 3.5: Long-Term Returns Empirical Results with IPOs recevied less than 100 Posts

This table reports the baseline empirical results with OLS estimation of Equation (1) in the restricted subsample of IPOs receved less than 100 posts during the pre-IPO period. The dependent variables are long-term returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Return_120_Days	(2) Return_150_Days	(3) Return_180_Days	(4) Return_240_Days	(5) Return_360_Days
Log Number of Posts	0.03564	0.03914	0.05245	0.1091**	0.1340**
	(1.18)	(1.13)	(1.26)	(2.32)	(2.20)
Bearish Tag Ratio	1.4166*	1.4115*	3.3049***	2.0580*	1.9858
	(1.94)	(1.68)	(3.28)	(1.80)	(1.37)
Bullish Tag Ratio	-0.6455***	-0.6947***	-0.8588***	-0.7960**	-0.8602**
	(-3.21)	(-3.00)	(-3.10)	(-2.53)	(-1.99)
Age	0.0001891	0.0004305	0.0009119	-0.001096	-0.002688
	(0.12)	(0.24)	(0.42)	(-0.44)	(-0.86)
Proceeds	0.04848	0.06157	0.09051	0.1980	0.2809*
	(0.59)	(0.65)	(0.80)	(1.54)	(1.66)
3_Week_Mkt	-3.1589***	-2.6598**	-4.6699***	-3.1817**	-4.6380**
	(-3.48)	(-2.55)	(-3.74)	(-2.25)	(-2.52)
Shares Number	-0.02263	-0.05091	-0.09320	-0.1810	-0.2766
	(-0.24)	(-0.47)	(-0.72)	(-1.23)	(-1.44)
Reputation	0.01198	0.01317	0.009427	0.004989	0.03190*
	(1.33)	(1.27)	(0.76)	(0.35)	(1.72)
VC Dummy	-0.001061	-0.02305	0.05692	0.01464	0.03278
	(-0.02)	(-0.30)	(0.62)	(0.14)	(0.24)
Exchange Dummy	0.1524**	0.1165	0.1347	0.2226**	0.2803*
	(2.11)	(1.40)	(1.35)	(1.97)	(1.92)
Tech Dummy	0.1471**	0.1339	0.1539	0.08314	-0.05382
	(2.02)	(1.60)	(1.54)	(0.73)	(-0.35)
Constant	-0.08467	0.2609	0.7252	1.6281	2.8653
	(-0.07)	(0.19)	(0.45)	(0.88)	(1.20)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	366	366	366	366	327
Adjusted R^2	0.065	0.034	0.058	0.067	0.126

Table 3.6: Baseline Empirical Results with Market Benchmarked Return

This table reports the baseline empirical results with OLS estimation of Equation (1) with market benchmarked returns as dependent variables. The dependent variables are opening return, benchmarked first day return, benchmarked 60 days return, and benchmarked 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$

	(1) Opening_Return	(2) return_1_b	(3) return_30_b	(4) return_60_b	(5) return_90_b
Log Number of Posts	0.03934***	0.06287***	-0.0008537	0.05022*	0.007090
	(3.54)	(3.82)	(-0.03)	(1.73)	(0.26)
Bearish Tag Ratio	-0.8243***	-1.3005***	0.3620	1.0854	1.6507**
	(-2.84)	(-3.13)	(0.54)	(1.48)	(2.42)
Bullish Tag Ratio	0.1634**	0.2193*	-0.1412	-0.2393	-0.2627
	(2.00)	(1.86)	(-0.74)	(-1.15)	(-1.36)
Age	-0.0003817	-0.0009705	0.0001615	0.00005690	0.0002669
	(-0.67)	(-1.05)	(0.11)	(0.03)	(0.18)
Proceeds	0.1430***	0.1556***	0.02090	0.03435	0.09571
	(4.41)	(3.34)	(0.28)	(0.42)	(1.25)
3_Week_Mkt	0.8062**	1.3835***	-0.4505	-1.0578	-1.6143*
	(2.20)	(2.61)	(-0.52)	(-1.13)	(-1.86)
Shares Number	-0.1491***	-0.1883***	-0.01621	-0.04499	-0.1177
	(-3.98)	(-3.50)	(-0.19)	(-0.47)	(-1.33)
Reputation	-0.003932	-0.007458	-0.02100**	0.004025	0.005840
	(-1.08)	(-1.41)	(-2.45)	(0.43)	(0.67)
VC Dummy	0.03494	0.06386	-0.008824	-0.03446	0.0002973
	(1.29)	(1.63)	(-0.14)	(-0.50)	(0.00)
Exchange Dummy	-0.007399	-0.03720	-0.004908	0.04354	0.1006
	(-0.26)	(-0.89)	(-0.07)	(0.59)	(1.47)
Tech Dummy	0.04048	-0.05657	0.01578	0.09420	0.1161*
	(1.37)	(-1.33)	(0.23)	(1.25)	(1.67)
Constant	1.7606***	2.3433***	0.2716	0.4454	1.2941
	(3.72)	(3.45)	(0.25)	(0.37)	(1.16)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	387	379	379	379	379
Adjusted R^2	0.201	0.132	-0.033	-0.024	0.015

Table 3.7: Baseline Empirical Results with Long-Term Market Benchmarked Return

This table reports the baseline empirical results with OLS estimation of Equation (1) with market benchmarked returns as dependent variables. The dependent variables are long term benchmarked returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) return_120_b	(2) return_150_b	(3) return_180_b	(4) return_240_b	(5) return_360_b
Log Number of Posts	0.02011	0.01767	0.02510	0.07615*	0.1224**
	(0.72)	(0.56)	(0.66)	(1.75)	(1.98)
Bearish Tag Ratio	1.0555	0.9425	2.5744***	1.2543	0.7601
	(1.50)	(1.18)	(2.68)	(1.15)	(0.50)
Bullish Tag Ratio	-0.6516***	-0.6487***	-0.7449***	-0.6333**	-0.5968
	(-3.27)	(-2.85)	(-2.73)	(-2.03)	(-1.27)
Age	0.0005954	0.0007154	0.001086	-0.0005151	-0.002866
	(0.38)	(0.40)	(0.51)	(-0.21)	(-0.86)
Proceeds	0.01521	0.02262	0.04342	0.1382	0.2771
	(0.19)	(0.25)	(0.40)	(1.12)	(1.54)
3_Week_Mkt	-2.2391**	-1.9708*	-3.7022***	-2.4512*	-3.6531*
	(-2.49)	(-1.92)	(-3.00)	(-1.75)	(-1.85)
Shares Number	-0.009875	-0.02226	-0.06182	-0.1377	-0.2899
	(-0.11)	(-0.21)	(-0.50)	(-0.96)	(-1.41)
Reputation	0.01520*	0.01595	0.01456	0.007845	0.03463*
	(1.70)	(1.57)	(1.19)	(0.56)	(1.74)
VC Dummy	0.006691	-0.03070	0.04532	0.003172	-0.006995
	(0.10)	(-0.41)	(0.50)	(0.03)	(-0.05)
Exchange Dummy	0.1508**	0.1230	0.1390	0.2207**	0.3526**
	(2.13)	(1.53)	(1.43)	(2.01)	(2.22)
Tech Dummy	0.1339*	0.1272	0.1249	0.002793	-0.1309
	(1.86)	(1.55)	(1.27)	(0.02)	(-0.79)
Constant	-0.2951	-0.1738	0.2908	1.0442	2.7821
	(-0.26)	(-0.13)	(0.19)	(0.58)	(1.08)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	378	378	378	376	325
Adjusted R^2	0.039	0.010	0.022	0.041	0.064

Table 3.8: Summary Statistics with Classified Untagged Posts

This table reports the summary statistics of input variables for regression analysis after all untagged posts are classified by the Fasttext NLP tool.

	N	mean	sd	min	max	p50
Open_Offer_Return	383	0.15	0.22	-0.36	0.94	0.08
First Day Return	375	0.18	0.31	-0.41	2.31	0.09
Number of Posts	383	26.86	112.50	1.00	1597.00	9.00
Bearish Tag Ratio	383	0.28	0.24	0.00	1.00	0.25
Bullish Tag Ratio	383	0.72	0.24	0.00	1.00	0.75
Age	383	14.67	21.94	0.00	166.00	9.00
Proceeds	383	251.25	545.12	3.50	8100.00	110.50
3_Week_Mkt	383	0.02	0.03	-0.10	0.13	0.02
Num_Share	383	14324535.48	23306070.32	350000.00	2.07e+08	7350000.00
Reputation	383	7.48	3.59	-9.00	9.00	9.00
VC Dummy	383	0.55	0.50	0.00	1.00	1.00
Exchange Dummy	383	0.72	0.45	0.00	1.00	1.00
Tech Dummy	383	0.53	0.50	0.00	1.00	1.00

Table 3.9: Baseline Empirical Results with Classified Posts

This table reports the baseline empirical results with OLS estimation of Equation (1) with sentiment measures constructed by tagged posts and classified untagged posts. The dependent variables are opening return, first day return, 60 days return, and 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Opening_Return	(2) First_Day_Return	(3) Return_30_Days	(4) Return_60_Days	(5) Return_90_Days
Log Number of Posts	0.04450***	0.06111***	-0.002472	0.04163	-0.003163
	(4.08)	(3.74)	(-0.09)	(1.44)	(-0.12)
Bearish Tag Ratio	-0.1038**	-0.1394**	0.06608	0.01946	0.07664
	(-2.21)	(-2.04)	(0.60)	(0.16)	(0.68)
Age	-0.0003318	-0.0009296	-0.0001594	-0.00001049	0.0004131
	(-0.57)	(-0.98)	(-0.10)	(-0.01)	(0.26)
Proceeds	0.1490***	0.1655***	0.01448	0.03446	0.09916
	(4.60)	(3.53)	(0.19)	(0.41)	(1.28)
3_Week_Mkt	0.6805*	1.1819**	-0.3138	-1.0539	-2.2918**
	(1.82)	(2.16)	(-0.36)	(-1.09)	(-2.54)
Shares Number	-0.1591***	-0.2006***	-0.009659	-0.03965	-0.1035
	(-4.27)	(-3.71)	(-0.11)	(-0.41)	(-1.16)
Reputation	-0.003769	-0.007744	-0.01701*	0.004627	0.0001150
	(-1.00)	(-1.40)	(-1.92)	(0.47)	(0.01)
VC Dummy	0.04061	0.07307*	-0.03018	-0.04045	0.009602
	(1.46)	(1.80)	(-0.46)	(-0.56)	(0.14)
Exchange Dummy	-0.01190	-0.04149	0.01573	0.05648	0.1012
	(-0.40)	(-0.96)	(0.23)	(0.74)	(1.42)
Tech Dummy	0.05435*	-0.03546	-0.01078	0.07125	0.1057
	(1.81)	(-0.82)	(-0.15)	(0.93)	(1.47)
Constant	1.9120***	2.5370***	0.1535	0.3962	1.1779
	(4.05)	(3.71)	(0.14)	(0.33)	(1.05)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	383	375	375	375	375
Adjusted R^2	0.190	0.105	-0.040	-0.036	0.015

Table 3.10: Long-Term Returns Empirical Results with Classified Posts

This table reports the baseline empirical results with OLS estimation of Equation (1) with sentiment measures constructed by tagged posts and classified untagged posts. The dependent variables are long-term returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Return_120_Days	(2) Return_150_Days	(3) Return_180_Days	(4) Return_240_Days	(5) Return_360_Days
Log Number of Posts	-0.01587	-0.01848	0.001220	0.04746	0.09052
	(-0.58)	(-0.59)	(0.03)	(1.09)	(1.54)
Bearish Tag Ratio	0.3222***	0.3239**	0.2737*	0.1781	-0.1055
	(2.83)	(2.49)	(1.72)	(0.98)	(-0.41)
Age	0.0007624	0.0009767	0.001814	-0.0006228	-0.002423
	(0.48)	(0.54)	(0.82)	(-0.25)	(-0.72)
Proceeds	0.02143	0.02800	0.03546	0.1554	0.2955*
	(0.27)	(0.31)	(0.33)	(1.25)	(1.69)
3_Week_Mkt	-2.7743***	-2.3075**	-4.1956***	-2.6897*	-3.5349*
	(-3.04)	(-2.21)	(-3.30)	(-1.85)	(-1.78)
Shares Number	-0.009506	-0.02725	-0.05612	-0.1518	-0.3127
	(-0.11)	(-0.27)	(-0.45)	(-1.06)	(-1.58)
Reputation	0.01230	0.01344	0.008444	0.007339	0.03386*
	(1.34)	(1.28)	(0.66)	(0.50)	(1.67)
VC Dummy	0.01033	-0.01442	0.07359	0.01956	0.02930
	(0.15)	(-0.19)	(0.78)	(0.18)	(0.20)
Exchange Dummy	0.1639**	0.1333	0.1393	0.2370**	0.3266**
	(2.27)	(1.62)	(1.38)	(2.06)	(2.07)
Tech Dummy	0.1089	0.09251	0.07721	-0.01390	-0.1510
	(1.50)	(1.12)	(0.76)	(-0.12)	(-0.92)
Constant	-0.2198	0.001242	0.3657	1.3607	3.3835
	(-0.19)	(0.00)	(0.23)	(0.75)	(1.36)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	374	374	374	374	333
Adjusted R^2	0.042	0.015	0.007	0.035	0.093

Table 3.11: Results with Turnover Rate

This table reports the results of OLS esimtation of Euqation (1) with turnover rate as the dependent variable. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Turnover Rate 1st Day	(2) Turnover Rate 1st Week
Log Number of Posts	0.1434***	0.2047***
	(3.75)	(5.77)
Bearish Tag Ratio	-0.2089	-0.1686
	(-0.21)	(-0.18)
Bullish Tag Ratio	0.2810	0.3840
	(1.00)	(1.47)
Age	-0.002591	-0.004382**
	(-1.18)	(-2.15)
Proceeds	0.6575***	0.2786***
	(6.39)	(2.89)
3_Week_Mkt	2.3352*	2.9298**
	(1.86)	(2.52)
Shares Number	-0.6020***	-0.2291*
	(-4.76)	(-1.94)
VC Dummy	-0.03350	-0.01163
	(-0.37)	(-0.14)
Exchange Dummy	-0.02403	-0.1445
	(-0.24)	(-1.56)
Tech Dummy	-0.2891***	-0.3951***
	(-2.85)	(-4.17)
Constant	3.2501**	-0.4072
	(2.04)	(-0.27)
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	381	380
Adjusted R^2	0.334	0.369

Table 3.12: Results with Turnover Rate

This table reports the results of OLS esimtation of Euqation (1) with turnover rate as the dependent variable. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. The sentiment meausres are constructed by tagged posts and classified untagged posts. *t* statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

	(1) Turnover Rate 1st Day	(2) Turnover Rate 1st Week
Log Number of Posts	0.1935***	0.2670***
	(5.49)	(8.39)
Bearish Tag Ratio	-0.4748***	-0.5935***
	(-3.11)	(-4.31)
Age	-0.003455	-0.005557***
	(-1.62)	(-2.90)
Proceeds	0.6141***	0.2578***
	(6.30)	(2.90)
3_Week_Mkt	2.4658**	3.2490***
	(2.04)	(2.98)
Shares Number	-0.5479***	-0.1910*
	(-4.60)	(-1.76)
VC Dummy	-0.05185	-0.03171
	(-0.59)	(-0.40)
Exchange Dummy	-0.03289	-0.1578*
	(-0.34)	(-1.80)
Tech Dummy	-0.3385***	-0.4562***
	(-3.48)	(-5.15)
Constant	2.6648*	-0.7975
	(1.77)	(-0.58)
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	377	376
Adjusted R^2	0.362	0.435

Appendix

Table 3.A.1: Baseline Empirical Results with Classified Posts in the Restricted Subsample of IPOs received less than 100 posts

This table reports the baseline empirical results with OLS estimation of Equation (1) in the restricted subsample of IPOs receved less than 100 posts during the pre-IPO period. The dependent variables are opening return, first day return, 60 days return, and 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. The sentiment meausres are constructed with tagged posts and classified untagged posts. *t* statistics in parentheses * p<0.1, ** p<0.05, *** p<0.01

	(1) Open_Offer_Return	(2) First_Day_Return	(3) Return_30_Days	(4) Return_60_Days	(5) Return_90_Days
Log Number of Posts	0.05060***	0.06533***	-0.006842	0.03792	-0.004417
	(4.31)	(3.72)	(-0.24)	(1.21)	(-0.15)
Bearish Tag Ratio	-0.09741**	-0.1338**	0.07026	0.02517	0.08590
	(-2.15)	(-2.01)	(0.64)	(0.21)	(0.76)
Age	-0.0005278	-0.001208	-0.0002720	-0.0001588	0.0004114
	(-0.93)	(-1.29)	(-0.17)	(-0.09)	(0.26)
Proceeds	0.1674***	0.1926***	0.01527	0.02660	0.09777
	(5.12)	(4.03)	(0.19)	(0.31)	(1.21)
3_Week_Mkt	0.6886*	1.0212*	-0.4794	-1.3342	-2.4971***
	(1.89)	(1.91)	(-0.54)	(-1.39)	(-2.75)
Shares Number	-0.1663***	-0.2097***	-0.0008443	-0.01585	-0.09667
	(-4.47)	(-3.86)	(-0.01)	(-0.16)	(-1.05)
Reputation	-0.005306	-0.01093**	-0.01791**	0.003449	-0.0003590
	(-1.43)	(-2.01)	(-1.99)	(0.35)	(-0.04)
VC Dummy	0.03854	0.07083*	-0.03890	-0.05460	-0.001188
	(1.43)	(1.79)	(-0.59)	(-0.77)	(-0.02)
Exchange Dummy	-0.009757	-0.04898	0.01118	0.04171	0.09711
	(-0.33)	(-1.13)	(0.16)	(0.54)	(1.33)
Tech Dummy	0.04877*	-0.03544	-0.008451	0.08867	0.1190
	(1.67)	(-0.83)	(-0.12)	(1.16)	(1.64)
Constant	1.9343***	2.5833***	0.04075	0.1081	1.1015
	(4.14)	(3.78)	(0.04)	(0.09)	(0.95)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	370	363	363	363	363
Adjusted R^2	0.171	0.094	-0.042	-0.030	0.025

Table 3.A.2: Long-Term Returns Empirical Results with Classified Posts in the Restricted Subsample of IPOs received less than 100 posts

This table reports the long-term returns empirical results with OLS estimation of Equation (1) in the restricted subsample of IPOs receved less than 100 posts during the pre-IPO period. The dependent variables are long-term returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. The sentiment meausres are constructed with tagged posts and classified untagged posts. *t* statistics in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$

	(1) Return_120_Days	(2) Return_150_Days	(3) Return_180_Days	(4) Return_240_Days	(5) Return_360_Days
Log Number of Posts	-0.001223	-0.004904	0.01642	0.07271	0.09780
	(-0.04)	(-0.14)	(0.39)	(1.52)	(1.60)
Bearish Tag Ratio	0.3262***	0.3277**	0.2726*	0.1718	-0.04679
	(2.84)	(2.48)	(1.69)	(0.95)	(-0.19)
Age	0.0005129	0.0007721	0.001645	-0.0005925	-0.002324
	(0.32)	(0.42)	(0.73)	(-0.23)	(-0.73)
Proceeds	0.04733	0.05882	0.06171	0.1956	0.2650
	(0.58)	(0.62)	(0.54)	(1.51)	(1.55)
3_Week_Mkt	-2.9059***	-2.4320**	-4.3365***	-2.8116*	-4.0295**
	(-3.14)	(-2.29)	(-3.35)	(-1.93)	(-2.15)
Shares Number	-0.02505	-0.05097	-0.07252	-0.1862	-0.2688
	(-0.27)	(-0.47)	(-0.55)	(-1.27)	(-1.40)
Reputation	0.01054	0.01140	0.006233	0.004181	0.03275*
	(1.12)	(1.06)	(0.48)	(0.28)	(1.69)
VC Dummy	0.003146	-0.01776	0.06724	0.02006	0.03138
	(0.05)	(-0.23)	(0.70)	(0.19)	(0.22)
Exchange Dummy	0.1713**	0.1348	0.1387	0.2378**	0.2850*
	(2.30)	(1.58)	(1.33)	(2.04)	(1.89)
Tech Dummy	0.1131	0.09786	0.09950	0.03254	-0.1235
	(1.53)	(1.16)	(0.96)	(0.28)	(-0.80)
Constant	-0.08989	0.2377	0.5140	1.7156	2.8611
	(-0.08)	(0.18)	(0.31)	(0.93)	(1.19)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	362	362	362	362	323
Adjusted R^2	0.050	0.017	0.010	0.040	0.104

Table 3.A.3: Baseline Empirical Results with Classified Posts Market Benchmarked Return

This table reports the baseline empirical results with OLS estimation of Equation (1) with market benchmarked returns as dependent variables. The dependent variables are opening return, benchmarked first day return, benchmarked 60 days return, and benchmarked 90 days return. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. The sentiment measures are constructed with tagged posts and classified untagged posts. t statistics in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$

	(1) Open_Offer_Return	(2) return_1_b	(3) return_30_b	(4) return_60_b	(5) return_90_b
Log Number of Posts	0.04450*** (4.08)	0.06012*** (3.68)	0.005094 (0.20)	0.05166* (1.81)	0.003263 (0.12)
Bearish Tag Ratio	-0.1038** (-2.21)	-0.1403** (-2.05)	0.07374 (0.68)	0.03648 (0.31)	0.06094 (0.55)
Age	-0.0003318 (-0.57)	-0.0009451 (-0.99)	-0.0001458 (-0.10)	-0.00009015 (-0.05)	0.0004313 (0.28)
Proceeds	0.1490*** (4.60)	0.1662*** (3.54)	0.01197 (0.16)	0.02726 (0.33)	0.08158 (1.07)
3_Week_Mkt	0.6805* (1.82)	1.1612** (2.13)	-0.1348 (-0.16)	-0.7492 (-0.79)	-1.5254* (-1.72)
Shares Number	-0.1591*** (-4.27)	-0.2004*** (-3.71)	-0.008423 (-0.10)	-0.03431 (-0.36)	-0.09659 (-1.10)
Reputation	-0.003769 (-1.00)	-0.007729 (-1.40)	-0.01684* (-1.92)	0.005537 (0.57)	0.002283 (0.25)
VC Dummy	0.04061 (1.46)	0.07192* (1.77)	-0.03385 (-0.53)	-0.04779 (-0.68)	0.007946 (0.12)
Exchange Dummy	-0.01190 (-0.40)	-0.04198 (-0.97)	0.009291 (0.13)	0.05219 (0.69)	0.09878 (1.40)
Tech Dummy	0.05435* (1.81)	-0.03561 (-0.82)	-0.01149 (-0.17)	0.06110 (0.80)	0.09443 (1.33)
Constant	1.9120*** (4.05)	2.5364*** (3.71)	0.1318 (0.12)	0.2957 (0.25)	1.0367 (0.93)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	383	375	375	375	375
Adjusted R^2	0.190	0.104	-0.040	-0.035	-0.005

Table 3.A.4: Long-Term Returns Empirical Results with Classified Posts Market Benchmarked Return

This table reports the long-term returns empirical results with OLS estimation of Equation (1) with market benchmarked returns as dependent variables. The dependent variables are benchmarked long-term returns in 120, 150, 180, 240, and 360 days after the first trading day. The key independent variables are Log Number of Posts, Bearish Tag Ratio, and Bullish Tag Ratio. The sentiment measures are constructed with tagged posts and classified untagged posts.

	(1) return_120_b	(2) return_150_b	(3) return_180_b	(4) return_240_b	(5) return_360_b
Log Number of Posts	-0.01179	-0.01723	0.004258	0.05300	0.09928
	(-0.43)	(-0.55)	(0.11)	(1.24)	(1.65)
Bearish Tag Ratio	0.3080***	0.3009**	0.2743*	0.1408	-0.1182
	(2.69)	(2.31)	(1.73)	(0.79)	(-0.44)
Age	0.0007645	0.0008378	0.001393	-0.0003301	-0.002873
	(0.48)	(0.46)	(0.63)	(-0.13)	(-0.84)
Proceeds	0.03026	0.03477	0.03620	0.1438	0.2797
	(0.38)	(0.39)	(0.33)	(1.16)	(1.55)
3_Week_Mkt	-2.0482**	-1.7984*	-3.4436***	-2.2218	-3.2722
	(-2.23)	(-1.72)	(-2.72)	(-1.56)	(-1.63)
Shares Number	-0.02283	-0.03114	-0.04994	-0.1416	-0.2869
	(-0.25)	(-0.30)	(-0.40)	(-0.99)	(-1.40)
Reputation	0.01247	0.01339	0.009999	0.006182	0.03451*
	(1.35)	(1.27)	(0.78)	(0.42)	(1.67)
VC Dummy	0.01763	-0.02210	0.05734	0.01493	-0.005760
	(0.26)	(-0.29)	(0.61)	(0.14)	(-0.04)
Exchange Dummy	0.1680**	0.1405*	0.1458	0.2320**	0.3595**
	(2.31)	(1.70)	(1.45)	(2.06)	(2.22)
Tech Dummy	0.1069	0.09936	0.08134	-0.03474	-0.1765
	(1.47)	(1.20)	(0.81)	(-0.30)	(-1.05)
Constant	-0.2057	-0.1295	0.1053	1.0695	2.7804
	(-0.18)	(-0.10)	(0.07)	(0.59)	(1.08)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	374	374	374	372	321
Adjusted R^2	0.023	-0.002	-0.011	0.026	0.054