Essays in Financial Market and Information Acquisition

by

Jiatao Liu

This thesis is submitted for the degree of *Doctor of Philosophy*

Faculty of Finance
Bayes Business School
City, University of London

December 2021

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I am extremely grateful to my supervisor, Professor Ian Marsh, for his invaluable advice, endless support and paternal guidance. His immense knowledge and bountiful experience have encouraged me to become a fully fledged academic researcher. In these five years of my Ph.D. studies, he also educated me on becoming a better person and making a positive impact on society.

I am deeply indebted to Professor Giovanni Cespa and Professor Thierry Foucault for their inspiration in my research and training me to grow as an economist to explore important economics and finance questions. Meanwhile, I am also deeply grateful to Professor Alessandro Rebucci, Professor Jack Bao, and Dr. Zipeng Wang for their mentoring during my Ph.D. studies.

Additionally, I would like to thank Richard Peterson, the Managing Director of Thomson Reuters MarketPsych, for supporting my research with his generous provision of the TRMI news dataset.

My family's help has been considerable. I appreciate them for encouraging and giving me the confidence to pursue my dreams and objectives. My father, Shanxiang Liu, empowered me to become a "strong" man; my mother, Yali Chen, gave me wings with which to fly and discover the world; my sister, Jiaqi Liu, taught me how to embrace people with a kind heart. This thesis is for you!

I would also like to thank my Ph.D. colleagues, Nan Zhao and Robin Tietz, for their knowledge sharing, insightful comments, and suggestions for my work. I would also like to express my gratitude to my friends, Haiming Bao, Shuzhi Huang, Difan Hong, Xianze Hong, Xingyi Li, David Smith, Jiaojiao Xia, and Qiyan Li, for the encouragement and cherished time I spent with them. I cannot imagine how I would have spent my time outside of my studies without them.

Finally, I would like to give special thanks to Ginger J. Xing for her data consulting and encouraging me to keep going.

Declaration of Authorship

I, Jiatao Liu, declare that this thesis titled, 'Essays in Financial Market and Information Acquisition' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself. Chapter 1 is co-authored with Ian Marsh. Chapter 3 is co-authored with Ian Marsh, Paolo Mazza, and Mikael Petitjean.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: Jiatao Liu

Date: 23/12/2021

Abstract

This thesis explores how information is incorporated into the valuation of assets, including classical securities - such as stocks and the innovative financial instruments - such as cryptocurrencies. On the one hand, I study the stock market by analyzing information acquisition theories and testing the fundamental *Homo economicus* tenets of neoclassical economics with novel textual data from online media and newswires. On the other hand, I document stylized facts for the nine most liquid cryptocurrencies and investigate whether the cryptocurrencies' pricing behaviors are explained by information in their own factor structure rather than information in the traditional financial market.

Chapter 1 studies mood, measured by Twitter messages, which causes investors' insufficient acquisition of information about assets and the implications of asset pricing. Using a Twitter-based mood measure, the study finds that mood swings are negatively predictive of investors' acquisition of earnings-related information when seeking to learn about companies' performance. Therefore, this study argues that this bias effect contributes to the explanation of classical (unconditional) pricing models' failures. Conducting tests on cross-sectional stock returns, the empirical results show that stocks that are more sensitive to mood earn a higher expected excess return than less mood-sensitive stocks. Sorting stocks to construct the risk factor portfolio based on mood betas as sensitivity to mood risk, this study is the first to quantify the risk premium (0.56% per month) by holding stocks subject to mood risk. The results are consistent with the theoretical prediction that investors mistakenly use mood as information rather than acquiring sufficient fundamental information about assets, thereby inducing mispricing in asset valuation.

Chapter 2 studies investors who use biased information from news media, with as subsequent tendency to make irrational decisions about acquiring firm-specific information compared to rational expectations. A static model of information acquisition by introducing a new irrationality channel in the form of biased information transmission yields testable predictions that are verified by using a novel dataset of news stories. First, when sentiment in news articles, as a proxy for biased public information, is more optimistic, investors tend to acquire less earnings-relevant information before the earnings announcement and vice versa. Second, the return predictability from firm-specific news sentiment confirms that it contributes to variations in asset information risk due, in a biased belief equilibrium, to the proportion of informed investors deviating from rational expectations. Overall, these findings suggest that biased public information

inherent in news sentiment serves to irrationalize investors' acquisition of firm-specific information through a biased perception of uncertainties in the risky asset payoff.

Lastly, Chapter 3 studies stylized facts on the return and volatility dynamics of the nine most liquid cryptocurrencies by using high-frequency tick data. Factor structures exist in both returns and volatility, but the explanatory power from the common factor is much stronger for volatility. The factor structures do not relate strongly to fundamental economic factors, and Bitcoin – which this study proposes is a "crypto market factor" – has only weak explanatory power. Dating the bubble in Bitcoin pricing allows the analysis to split the sample into pre-bubble, bubble and post-bubble periods. The importance of these different periods is clear, revealing shifting relationships between the nine cryptocurrencies and Bitcoin. Model-free realized cryptocurrency betas with Bitcoin increase during the bubble and the explained fraction of cryptocurrency variance remains at an elevated level after the bubble burst. In sum, the results show that information in the factor structure explains variations of returns and volatilities in the cryptocurrency market.

To my family and the people I've loved.

Chapter 1

Mood Swings and Insufficient Information Acquisition: A Study on Cross-Section of Stock Returns

1.1 Introduction

A stock's mood beta is its sensitivity to variations in the mood of the public. As noted by Hirshleifer et al. (2020), mood can be viewed as a special case of investor sentiment, and as in their paper, our focus is on the effects of emotional valence - whether the public's mood is happy or sad.[1] We draw inspiration from the psychological studies by Schwarz and Clore (2007) and Storbeck and Clore (2008), who propose that agents apply their feeling or affects as information for decision-making judgment. An early psychological study by Easterbrook (1959) builds the foundation to discuss the effect of emotion on agents' behaviors that the number of cues and information utilized in tasks tends to decrease with emotions raised. On the one hand, careless decision-making with less fundamental information incorporated has been addressed in the classic psychological study by Tversky and Kahneman (1974). They state that people with positive or optimistic mood tend to use their heuristic thinking in the decision-making process. For example, in economics, overconfident or optimistic investors may conduct trading or investment decisions based on non-asset fundamental information. Studies in psychology elaborate the "mood-as-input" theory to argue

[1] Sentiment is a much broader term encompassing both affective concepts such as emotions - including mood - and non-affective concepts such as attention or heuristic beliefs.

that positive mood induces subjects to make fewer attentions and cognitive efforts due to enjoyable or easy satisfied feelings (Meeten and Davey, 2011). On the other hand, the negative relation between negative mood and decreased amount of fundamental information used in decision-making is mainly implied by the impact of depression in classical psychological studies through narrowing or impairment of attentional captivity. (Hasher and Zacks, 1979).[2] They elaborate that negative moods such as depression or stress reduce individuals' attention capacity; as a consequence, individuals' are unsophisticated to perform complex tasks that require substantial efforts. For instance, a study published on *Nature* by Keller et al. (2019) states that low mood induces impairment of attention and influences overall cognitive styles regardless of consciously or unconsciously. As investors' choice to acquire information is *active learning* that requires efforts or costs such as allocating limited attention (Veldkamp, 2011), we seek to study whether affective states such as mood debilitate investors performance on information learning about assets.[3] Inspired by Loewenstein (2000)'s study that argues incorporating affective states into the utility function to enrich the normative analysis, we are motivated to take mood as a conditional factor to investigate its impact on the average investor's decision on the choice of information acquisition.

Instead of using weather as the customary proxy for mood,[4] our study is inspired by the growing body of literature in the field of textual analysis in finance and economics. Essentially, the measure of mood we used is from the Hedonometer project run by the University of Vermont Complex Systems Center. Hedonometer constructs a daily happiness index based on the analysis of the words used in messages posted on Twitter. A random sampling of approximately 50 million messages posted to the system (representing around 10% of the total number of messages posted each day) is then analyzed. The words from the English language messages are pooled, and this pool of words is assigned a happiness score based on the average happiness score of the words it contains.[5] As is immediately apparent, this index is not designed to be

[2] Conway and Giannopoulos (1993) finds the evidence to indicate that depressed people use less of the available and relevant information when they conduct complex tasks. More specifically, Conway and Giannopoulos propose that the causation of insufficient information demand is mainly due to the problem of reduced attentional resources not from the motivational deficits. See related studies by Dobson and Dobson (1981) and Silberman et al. (1983), who argue that negative mood from depression causes less usage of fundamental information in learning and decision making.

[3] Investors' attention, demand or learning to assets' fundamental information falls within the ambit of the information acquisition theory. In subsequent context, information demand, acquisition or learning are used interchangeably.

[4] Weather as a classical proxy for mood has been comprehensively addressed in the literature on the effect of mood on financial market and investors' trading behavior. See related studies by Saunders (1993), Hirshleifer and Shumway (2003) and Chang et al. (2008).

[5] A more detailed discussion of the fundamental work on the Twitter mood index can be found in Dodds and Danforth (2010), Cody et al. (2016), Reagan et al. (2016) and Reece et al. (2017).

finance-oriented.

We use this happiness score as our proxy for the public's mood. Compared to the weather or sporting results - both exogenous shocks that are assumed to affect people's mood - the happiness score is an endogenous measure that reflects mood. A recent study by Edmans et al. (2021) uses Spotify music data in a similar manner - albeit over a shorter sample - to argue that their endogenous music measured sentiment captures information about mood swings. Nevertheless, the main discussion in the study by Edmans et al. (2021) follows the path of sentiment mispricing effect on market return in the literature, in that the theoretical foundation of the mood-biasing effect on information acquisition in our study is distinct from extant behavioral studies in sentiment or mood. Clearly, the appeal of Twitter differs across demographic groups. A recent survey by Wojcik and Hughes (2019) concludes that Twitter users are younger, more educated, have higher incomes and are more likely to identify as Democrats than the U.S. adult population as a whole.[6] On the other hand, Twitter is a close match to the population in terms of gender and ethnicity. As such, we acknowledge that the Twitter-based happiness score is an imperfect proxy for the public's mood. We would expect this to make it more difficult to find empirical support for the relationships we hypothesize. We discuss the Twitter mood data in more detail in section 1.2.

By virtue of textual analysis on Twitter messages, we are able to measure the public's mood and test the effect of changes in mood on investors' acquisition of information about assets. We follow the study by Weller (2018) and estimate a price jump ratio around earnings announcements as a firm-specific measure of information acquisition. This jump ratio is the post-announcement absolute cumulative abnormal return (ACAR) divided by the total ACAR, including the pre-announcement period. As investors acquire more (less) information about earnings before the announcement, we expect this jump ratio to be lower (higher). As an alternative and more direct measure, we also directly calculate the average number of SEC EDGAR file downloads. Finally, we estimate mood swings by calculating the average absolute change of the daily Twitter mood index or its average volatility in the most recent month before the firm's earnings announcement.

Empirically, we find evidence that mood swings predict lower levels of information acquisition by investors. When mood becomes either more positive or negative, investors decide to acquire less value-relevant information (regarding firm earnings)

[6] See the link for the article. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

about assets before the information is released. As a consequence, investors inadequately learn fundamental information that ought to be incorporated into asset prices. In a seminal study, Van Nieuwerburgh and Veldkamp (2009) state that an asset's return and risk fall as investors learn about it. Therefore, an asset that the average investor understands well is expected to have a lower standard deviation of its return. In line with the importance of investors' information acquisition choice, Veldkamp (2011) proposes that conditional betas on information that the average investor knows must differ from the unconditional betas estimated by classical pricing models such as the single-factor CAPM or Fama-French multifactor models. As investors' learning about assets is *ad hoc*, another interesting question arises as to what the implications of the mood biasing effect on investors' information learning are for asset pricing.

Therefore, we argue that the effect of insufficient information learning caused by mood contributes to explain the failures of classical pricing models. When pricing assets that investors inadequately study, the beta-sensitivity to risk factors in unconditional pricing models loses effectiveness due to mood's causation of insufficient information incorporation on learning about the risks contained in the assets. The economic intuition is that mood causes investors' inadequate asset information acquisition; thus, the structure of covariance between stochastic discount factor (multi-factors) and the risky asset return tends to deviate from the unconditional pricing models implied. More specifically, assets subject to mood's effect on insufficient learning contain more risk (high standard deviation). The betas (quantity of risk), which are the sensitivity of asset return to risk factors conditional on insufficient information acquisition induced by mood, should be higher than the unconditional betas, leading to a higher expected return. Using the unconditional betas underestimates the risk in assets; thus, researchers find almost a thousand risk factors in empirical asset pricing studies to invoke in situations in which CAPM and Fama-French factor models fail to explain cross-sectional stock returns. We propose the effect of mood-inducing investors' lack of study on assets provides the "opportunity" in empirical asset pricing studies to find new risk factors.

Nevertheless, we argue that not all assets are subject to the mood effect. We test the implications of the mood pricing effect, namely mood beta, for the cross-section of U.S. stock returns. We show that the returns of a significant proportion of U.S. stocks are sensitive to changes in Twitter mood and that mood is a significantly priced risk factor. As investors mistakenly incorporate mood as information and do not learn or acquire as much fundamental information as they should to price assets, mood as a behavioral factor adds additional risks which are not explained by asset fundamentals.

We propose that stocks that are sensitive to vary with public mood (moody stocks) earn a higher expected return as a risk premium required by investors who hold these mood risky assets. To the best of our knowledge, we are the first to find that mood risk measured by the sensitivity of public mood is significantly priced in the cross-section stock returns.

In line with Lo et al. (2005) who propose that multi-factor asset pricing models can be enriched by considering the effect of emotional factors, we estimate the mood beta by using the method proposed by Bali et al. (2017). This involves the identification of mood-sensitive stocks by adding the Twitter mood index into the Fama-French five-factor model in line with the momentum factor.[7] By sorting stocks according to their sensitivity to changes in the Twitter mood index, we show that portfolios of negatively (positively) mood-sensitive stocks earn excess returns of 1.66% (1.65%) per month. Portfolios of non-moody stocks earn excess returns of just 1.1% per month. We construct a mood-mimicking portfolio by taking long positions in stocks that have a large (positive or negative) sensitivity to mood and shorting stocks which are mood-insensitive. The mimicking mood portfolio has an average return of 0.56% per month. Our empirical findings are consistent with the theoretical implication that stocks investors insufficiently learn about are riskier with a higher expected return.[8]

We apply standard cross-sectional asset pricing techniques to test whether these mood betas are priced in the U.S. equity market. We first sort stocks into ten portfolios according to rolling sensitivities to Twitter mood to identify which stocks are the most sensitive to either positive mood or negative mood. Stocks' sensitivity to mood is induced by its effect on investors' trading behaviors and risk sensitivity. In the *Homo economicus* paradigm, investors' valuation of assets should be rationally based on acquired or learned fundamental information. However, as we find empirically, investors incorporate less fundamental information when their mood is more volatile, in either the positive or negative directions. Therefore, when investors become moody, they mistakenly rely on their feelings as useful information with which to trade or price assets and do not acquire as much fundamental information as they should. On the one hand, when investors' mood is more positive, they are less risk-averse and tend to overprice assets (Bassi et al., 2013; Kaplanski et al., 2015), as a result of which they

[7] Bali et al. (2017) measure stocks' sensitivity to economic uncertainty by taking the regression coefficients on the uncertainty index in the time-series regression as the uncertainty beta.

[8] The empirical findings also strongly confirm the recent seminal studies of mood as a behavioral factor contributing to mispricing and risks to the financial markets (Goetzmann et al., 2015; Bushee and Friedman, 2016; Hirshleifer et al., 2020).

invest in more risky stocks, exerting buying pressure on these stocks.[9] On the other hand, when investors' mood is more negative, their pessimistic feeling causes them to be more risk-averse and perceive higher risk. Raghunathan and Pham (1999) argue that agents with a sad mood are biased in favor of high risk with high reward, on the grounds that investors seek stocks which they believe to generate high returns in negative mood days to compensate for the high risk entailed by those stocks.[10] As a consequence, positive moody stocks' returns increase as mood is more optimistic and negative moody stocks' returns increases as mood is more pessimistic. In our mood beta estimation, we find that portfolio 1 stocks with an average mood beta of -0.58 are those sensitive to negative mood, while stocks in portfolio 10 with average mood betas of 0.61 are those sensitive to positive mood. All in all, regardless of whether stocks are sensitive to either positive or negative mood, investors' risk perceptions and trading behaviors are biased by mood via the decision to acquire less information . These stocks are more likely to be affected by the irrational trigger-mood and are more risky than stocks which are less likely to be affected by mood.

First, the value-weighted excess returns of portfolios 1 and 10 are 1.66% and 1.65% per month. The average excess return of mood-insensitive stocks in portfolios 5 and 6 is around 1.1%. A high-low portfolio that takes a long position in both portfolio 1 (negative mood sensitivities) and portfolio 10 (positive mood sensitivities) and short positions in the mood-insensitive portfolios 5 and 6 generates a statistically significant average excess return of 0.52% per month.[11] Analysis based on the Fama-French five-factor model suggests an alpha of 0.48% per month. This rises to 0.50% with the addition of the Carhart momentum factor, and to 0.54% with the further addition of long- and short-term reversal factors. The t-statistics on these alphas range from 3.25 to 4.18.

Second, we construct a mimicking mood factor portfolio to determine whether the risk premium induced by mood can be captured by benchmark pricing factors. This mood factor earns a statistically significant risk premium of 0.56% per month. However,

[9] By studying investors' behavior in Finland which is considered more likely to be affected by people's mood, Kaustia and Rantapuska (2016) argue that positive mood measured by sunshine light length drives investors to buy more than they sell. Goetzmann et al. (2015) find evidence that even institutional or sophisticated investors are subject to cognitive biases such as mood, with optimism increasing buy-sell imbalances.

[10] Shu (2010) develops a model to indicate that the pessimistic mood causes investors' risk aversion and impatience to increase, as a result of which the stochastic discount factor is decreased to price the asset with a higher return.

[11] The long-short strategy takes long positions in positive and negative mood beta stocks (and short positions in zero mood beta stocks). However, the strategy is not a zero-mood-beta smart money scheme, since combining positive and negative mood-sensitive stocks does not remove the exposure to risk from either positive or negative mood changes.

it is positively correlated with the market, size and reversal factors and negatively correlated with profitability and momentum factors. Taking the market factor into account leaves an unexplained mean return of 0.38% per month. As successive extra factors are accounted for, the alpha increases back to 0.56%, equal to the mean return on the mood factor. We construct the orthogonalized mood factor as the component of the mimicking mood factor unexplained by all other factors. Standard factor models fail to explain the returns earned by stocks most affected by mood and the decile of stocks with the largest absolute mood beta earn a statistically significant alpha of between 0.36 and 0.41% per month, while the long-short mood strategy earns an alpha of between 0.47 and 0.55%. Adding the orthogonalized mood factor to the analysis reduces all alphas and removes all statistical significance. As the two portfolios containing stocks with the largest absolute mood betas each load significantly on the orthogonalized mood factor, we infer that it has significant pricing power.

Finally, we construct 25 portfolios based on independent sorts of market capitalization and absolute mood beta. Within each size quintile, the most mood-sensitive stocks earn higher mean returns than the least mood-sensitive stocks. This effect is economically large and statistically significant for all size quintiles. Alphas from the alternative factor models are positive and significant for the most mood-sensitive quintile of stocks in the majority of size quintiles.[12] A high minus low mood sensitivity strategy yields a positive mean alpha in all five size quintiles, largest in magnitude and statistically significant for the smallest quintile. Adding the mood factor to the analysis removes all significant alphas for the most mood-sensitive stocks and for all the high minus low mood sensitivity strategies. Indeed, incorporating the mood factor turns the slope of alphas with respect to mood sensitivity negative for the larger quintiles of stocks.

Additionally, we find that stocks which are sensitive to public mood - positively or negatively - are typically small in size, relatively young, pay lower dividends, have a large R&D ratio, are not profitable, engage in more external financing and have higher levels of idiosyncratic risk. Alternatively, these characteristics share similarities with the link between information asymmetry and stocks that are subject to the effect of mood. The theoretical channel of mood causing less information acquisition in this study, in fact, has a corresponding implication of asymmetric information problem in assets. Meanwhile, our findings are consistent with the study by Bushee and Friedman (2016), who argue that both noise traders and unsophisticated investors are more likely

[12] They are mainly in the smallest (moody stocks in respect of size 1), mid-cap (moody stocks in respect of size 4) and large-cap size quintiles (moody stocks in respect of size 5).

to take feelings or non-fundamental factors as useful information to price asset values and trade stocks when other information is lacking.

Although we emphasize that our study of public mood sensitivity is different from the more general concept of sentiment, the two concepts are clearly related and as such we contribute to the literature on the implications of sentiment and emotions in asset pricing. The exact meaning of sentiment is obscure but, as noted by Baker and Wurgler (2006), one possible definition is the propensity to speculate. Shifts in the propensity to speculate drive shifts in the relative demand for speculative investments and therefore have cross-sectional effects. Much of the literature draws on Baker and Wurgler (2006; 2007), who construct a sentiment index and use this to demonstrate a significant impact of investor sentiment on a cross-section of U.S. stock returns. They find that difficult-to-arbitrage stocks have a negative relation between sentiment and subsequent returns. Baker et al. (2012) confirm the power of investor sentiment in international stock markets. Other studies on the importance of investor sentiment for stock returns include those of Swaminathan (1996), Brown and Cliff (2004), Kumar and Lee (2006), Lemmon and Portniaguina (2006) and Stambaugh et al. (2012).

However, the mood effect we study in this chapter is related to non-fundamental information incorporated into the asset valuation. On the one hand, mood entails to people's feeling states, which need not to be about anything in particular (Wyer Jr and Srull, 2014). On the other hand, sentiment refers to how people or investors in the market feel in combinations with what the feeling is about. For example, investors could have positive or negative sentiment for the market based on their speculation, but they could be in a bad or happy mood that is unrelated to the market. Therefore, the Twitter mood data we use in this study is a non-specific measure for people's general happiness level, which is not as financial market-oriented as the construction of B&W style sentiment index. In fact, in psychological studies, researchers find that mood does not have specific focus as emotions or sentiment (Averill, 1980; Frijda et al., 1986; Clore and Schwartz, 1988). Essentially, both the classical weather proxied mood and innovative measures with textual data are relatively more orthogonalized to stock market information than the investor sentiment which is endogenously implied by the financial market measures.[13]

[13] Unlike our Twitter mood index, the sentiment indices most commonly used in the literature are constructed from proxy measures extracted from financial markets. The Baker and Wurgler (2006) sentiment measure, for example, is based upon closed-end fund discounts, market turnover, IPO numbers and first-day returns, the equity share in new issues and the dividend premium. Huang et al. (2015) use the same six proxies but a different statistical method to construct an alternative sentiment index which they show also supports the pricing power of investor sentiment.

Our study makes a unique contribution to the classical literature that favors the use of measures of mood more exogenous to financial markets. Behavioral finance studies argue that mood can be caused by weather or happy events, therefore channeling the proxied mood to impact investors' pricing or trading behaviors. Saunders (1993) observes that NYSE stocks tend to have positive returns on sunny days and moderate returns on cloudy days. Hirshleifer and Shumway (2003) support the notion that good mood is associated with sunny weather and they find that there is a highly significant correlation between sunshine and stock returns. Kamstra et al. (2003) document the existence of an effect of seasonal affective disorder (SAD) — a psychological condition in which a daylight deficit has a detrimental impact on people's mood — on stock market returns around the world. Kamstra et al. (2000) show that stock market returns on Mondays following daylight saving clock changes are lower than returns on normal Mondays. They propose that market participants' loss of an hour of sleep may result in increased anxiety or risk aversion which adversely affects stock market returns. Edmans et al. (2007) find that sporting events affect investor mood, with soccer defeats in particular being associated with significant market decline in a country.[14] In contrast to proxied mood such as weather, the Twitter message or comparable online data such as the one used in the study by Edmans et al. (2021) directly measure mood. Furthermore, our study concentrates more on cross-sectional analysis instead of the aggregate market. In addition, our key premise is to determine whether the significant effect of mood causing less information acquisition about assets can be seen as a risk factor on cross-sectional asset returns and whether this risk should be compensated for by investors who hold the stocks which are thought of as the "volunteers" of irrational trading or pricing behaviors triggered by public mood.

Furthermore, our study also contributes to the growing body of studies in textual analysis in financial markets,[15] particularly those that use Twitter message data. For example, Bollen et al. (2011) conduct textual analysis from large-scale Twitter feeds based on a computational algorithm to measure mood. They claim, somewhat controversially, that mood calculated from Twitter feeds has predictive power with regard to the DJIA value index. Dey (2014) shows that tweets contain useful information. He studies the polarity value of each tweet by sentiment analysis and finds a significant correlation between changes in stock price and changes in the polarity values of tweets. Both of studies use time series techniques to relate Twitter mood to stock returns. Our

[14] Goetzmann and Zhu (2005) propose that the relationship between mood or weather effects and stock returns may be responsible for the behavior of market makers rather than individual investors.

[15] A comprehensive review and survey can be found in Tetlock (2014a).

analysis focuses on cross-sectional analyses and seeks specifically to quantify the premium associated with mood risk in an asset pricing framework. Additionally, Behrendt and Schmidt (2018) argue that the Twitter-based sentiment marketed by Bloomberg has not contributed valuable information for blue-chip stocks such as the DJIA constituents on future volatility forecasting. However, by using the same database to test on Russell 3000 stocks, Gu and Kurov (2020) find that Twitter-based sentiment predicts stock return without reversals, such that potentially genuine information can be found in stock-level Twitter messages. Although both of them apply Twitter data to cross-sectional studies to discover whether stock-related information can be found in Twitter to incorporate into the asset pricing, we explore the effect of mood measured with non-financially-oriented Twitter messages on investors' choice of fundamental information learning about assets rather than assets' genuine information extracted from Twitter.

The study is structured as follows. Section 1.2 describes the data for mood and the empirical evidence of the mood effect on investors' information acquisition in line with theoretical development in asset pricing. Section 1.3 details mood beta estimation and portfolio. We discuss the key asset pricing tests and results of the mood factor in section 1.4. In section 1.5, we thoroughly discuss theoretical motivations and connections to our empirical findings. Section 1.6 briefly outlines the robustness tests we conducted. Section 1.7 provides our conclusions drawn from the analysis.

1.2 Mood Data and Theoretical Motivation

1.2.1 Mood Measurement from Twitter Message

We use the daily mood score from a Twitter text analysis project supported by Hedonometer.org, based at the University of Vermont Complex Systems Center. In their research, the data generate a "Twitter Happiness Score" which is explored as a function of time, space, demographics and network structure using Twitter feeds as a data source. Hedonometer samples roughly 50 million messages posted on Twitter each day. Words in messages written in English are extracted (resulting in approximately 100 million words each day) and this pool of words is assigned a happiness score based on the average happiness score of the scored words. Hedonometer considers 10,222 scored key words used to calculate mood. These are the most frequently used words in Google Books, New York Times articles, music lyrics and Twitter messages. Each word has been assigned a happiness score (ranging from 1=sad to 9=happy) using Amazon's Mechanical Turk service. Words scoring high on this scale include laughter

(8.5), happiness (8.44), love (8.42), excellent (8.18), joy (8.16), successful (8.16) and win (8.12). Low-scoring words include terrorist (1.30), suicide (1.30), murder (1.48), death (1.54) and cancer (1.54).

The algorithm used to measure the mood score is as follows:

$$h_{avg}(T) = \frac{\sum_{i=1}^{N} h_{avg}(w_i) f_i}{\sum_{i}^{N} f_i} = \sum_{i=1}^{N} h_{avg}(w_i) p_i \tag{1.1}$$

In a given text T, f_i is the frequency of the i^{th} word w_i, and the estimation of happiness for a unique word is $h_{avg}(w_i), p_i = f_i/(\sum_{j=1}^{N} f_j)$. The given text T can be extracted flexibly based on different time intervals. For example, Twitter feeds used to calculate the mood score can be extracted as counting either minutes or days. Hedonometer.org reports that there are about 20 million tweets per day and approximately 14,000 per minute as at August 31 2011 (Dodds et al., 2011).[16] A day is considered happier than usual if happy words are used more frequently than usual, or if sad words are used less frequently than usual.

The daily Twitter score data are available from September 2008; our sample ends in December 2016. The upper panel of Figure 1.1 plots the Twitter mood score at a daily frequency and the lower panel plots the Baker and Wurgler orthogalized sentiment index at a monthly frequency. The time range in the two plots covers the interval between September 2008 and September 2015 for which both data series are available.[17]

1.2.2 Mood Impact on Information Acquisition

We investigate how mood plays a role in affecting investors' economic decisions and behaviors. The key argument we want to address is that investors mistakenly incorporate mood as useful information (Schwarz and Clore, 2007). Consequently, investors rely on less fundamental information to price or trade assets when the average investor becomes moody. Therefore, the hypothesis we are going to test is:

Hypothesis 1: As mood tends to be more volatile, investors tend to learn less information about assets.

[16] The analysis of Twitter mood is from a random sample of 25% of the tweets in the database. There are about 230 million unique words. Due to the computational difficulty, Hedonometer.org analyzes the first 50,000 most frequent words without compromising estimation accuracy.

[17] The sentiment index is downloaded from Wurgler's website where the final observation is September 2015.

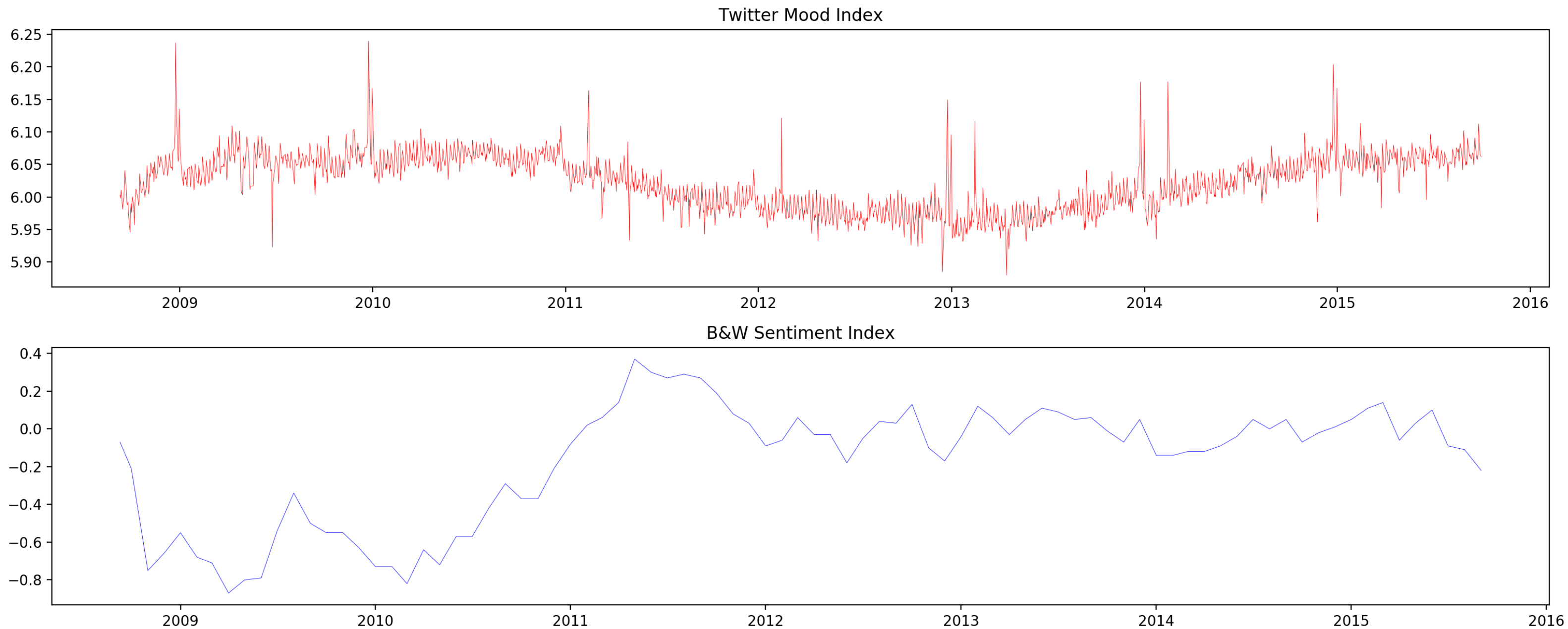


FIGURE 1.1: Twitter Mood vs. Baker & Wurgler Sentiment Index

1.2.2.1 Measure of Information Acquisition

Because investors' information acquisition is not directly observed, we conduct an event study of firm earnings announcements to test the inverse relationship between mood and investors' asset information acquisition.[18] First, we follow a novel study by Weller (2018) to estimate a price jump ratio as the measure of investor information acquisition. The price jump ratio is estimated by taking the post-announcement ACAR divided by the total ACAR that includes pre-announcement periods of around 21 days:

$$Jump_{i,t}^{a,b} = \frac{CAR_{i,t}^{T-1,T+b}}{CAR_{i,t}^{T-a,T+b}} \tag{1.2}$$

where $a = 21$ and $b = 2$ as the pre- and post-announcement window respectively. The CAR is the absolute cumulative abnormal return subject to the study window from the Fama-French five-factor model and also include the momentum factor:

$$CAR_{i,t}^{j_1,j_2} = \sum_{t=j_1}^{j_2} \left(R_{it}^{e} - \alpha_i - \sum_{m=1}^{M} \beta_{i,m} f_{m,t} \right) = \sum_{t=j_1}^{j_2} \epsilon_{i,t} \tag{1.3}$$

where R_{it}^{e} is stock excess return and $f_{m,t}$ is the Fama-French and the momentum factors. We estimate the α_i and $\beta_{i,m}$ based on 252 daily observations and 90 days before the earnings announcement. We require firms that have observations on at least 63 trading days to conduct equation (1.3) as the estimation. To avoid the zero denominator issue in equation (1.2), we follow the instruction from Weller (2018) to set a threshold as $|CAR_{i,t}^{T-21,T+2}| > \sqrt{24}\hat{\sigma}_{i,t}$ where $\hat{\sigma}_{i,t}$ is daily return volatility in 24-day event window.

The rationale for the price jump ratio as a proxy of investors' information acquisition is that, as investors acquire more earnings-related information to learn about the company before the day of the announcement, due to price discovery, the price incorporates more information about earnings in the pre-announcement period. Therefore, we should observe a large denominator in equation (1.2). On the contrary, if investors do not acquire earnings-relevant information to learn about the company before the announcements, the stock price will jump immediately as the earnings are released at the announcement day and afterward.[19] Consequently, we should expect a large numerator relative to the denominator in equation (1.2). In sum, the higher the price

[18] Earnings announcement date and relevant data are extracted from the IBES dataset. Stock return data is from CRSP, and financial fundamentals are from Compustat. See section 1.2.4 and Appendix A.2 for details.

[19] We use 2 days after the announcement as the post-announcement periods to capture the PEAD effect.

jump ratio implies a lower firm-specific information acquisition conducted by investors to learn about the company.

In addition to the price jump ratio measure, we estimate investors' demand for learning information about companies by exploiting the SEC EDGAR logs of access to firm-specific fillings around a quarterly earnings announcement. Specifically, we calculate the average total count of search volume for the files in the most recent month before the announcement. We then take the natural logarithm for the average of total SEC files searching volumes ($LSECV$). To some extent, the count of SEC EDGAR file searching volume is a more straightforward way to understand investors' demand for learning. As the searching volume increases before the announcement, investors are more eager to learn about the company to forecast or estimate its upcoming earnings performance.

1.2.2.2 Measure of Mood Swings

We use two measures to identify mood swings from the Twitter data. First, we take the average absolute change of daily mood in the most recent month before the firm earnings announcement. Therefore, regardless of the positive or negative direction of the mood change, the larger the average value of the absolute difference in the mood, the moodier the investors become. Second, we directly calculate the volatility of the daily mood in the most recent month before earnings announcements as a direct measure of mood swings. Finally, the data in the following test is from September 2008 to December 2016.

1.2.2.3 Hypothesis Testing

We conduct fixed effect regressions to test the *Hypothesis 1* as follows:

$$Dep_{i,t} = \beta_0 + \beta_1 MoodSwing_{t-31,t-1} + X\delta + \epsilon_{i,t} \tag{1.4}$$

$Dep_{i,t}$ is either the price jump ratio from equation (1.2) or the direct measure of investors' information demand $LSECV$ as the proxy of the average investor's learning about firm-specific information. $MoodSwing$ is either the average absolute daily mood change or mood change volatility in the most recent month before announcements. The X is a vector of control variables (see detailed definitions in Appendix A.2) and the δ is a vector of coefficients. Since the investors' information acquisition or demand is significantly related to economic uncertainties (Benamar et al., 2019; Andrei et al., 2020), we add VIX and Economic Policy Uncertainty Index (EPU) by Baker et al.

(2016) as additional control variables to identify the impact of mood more clearly. We are interested in testing whether β_1 has a significant inverse relationship with investors' information acquisition proxies.

TABLE 1.1: Mood Swings Impact on Investors' Information Acquisition

This table presents the results of regressions of the price jump ratio and the counts of EDGAR SEC file downloads as the proxy for investors' firm-specific information acquisition on mood swings measured by either the average absolute change of daily mood or the volatility of daily mood during the firm earnings announcement window. Columns (1)–(4) are based on the fixed-effect regression from equation (1.4): $Dep_{i,t} = \beta_0 + \beta_1 MoodSwing_{t-31,t-1} + X\delta + \epsilon_{i,t}$, where $Dep_{i,t}$ is either the price jump ratio ($jump_{i,t}$ in Panel A) from equation (1.2) or information demand measure that is the natural logarithm for the average of total SEC files searching volumes ($LSECV_{i,t}$ in Panel B) in the most recent month before the announcements. Control variables include: economic uncertainty proxies (VIX or EPU) and the Number of Analyst Forecast is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility, and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Additionally, we control the day-of-the-week effect (DOW) in Panel B when we use $LSECV_{i,t}$ to measure investors' information acquisition. Detailed definition of all variables are available in Appendix A.2. Standard errors are clustered by both firm- and time-fixed effect in column (1)-(4). $***$, $**$, $*$ indicate statistical significance at the two-sided 1%, 5%, 10% levels, respectively.

Panel A: Information Acquisition Measured by Price Jump Ratio

Dependent Variable	(1) $Jump_{i,t}$	(2) $Jump_{i,t}$	(3) $Jump_{i,t}$	(4) $Jump_{i,t}$
$\sigma(Mood_{t-31,t-1})$	0.558***	0.509**		
	(0.216)	(0.216)		
$\lvert\overline{\Delta Mood_{t-31,t-1}}\rvert$			22.955**	25.620**
			(9.961)	(9.967)
$VIX_{t-21,t-1}$	−0.002***		−0.002***	
	(0.000)		(0.000)	
$EPU_{t-21,t-1}$		−0.000		−0.000
		(0.000)		(0.000)
$Size_{i,t-42,t-21}$	0.005	0.010	0.005	0.010
	(0.007)	(0.007)	(0.007)	(0.007)
$Turn_{i,t-42,t-21}$	−0.004	−0.002	−0.004	−0.002
	(0.003)	(0.003)	(0.003)	(0.003)
$Price_{i,t-42,t-21}$	0.006	0.005	0.006	0.005
	(0.007)	(0.007)	(0.007)	(0.007)
$RV_{i,t-42,t-21}$	−0.010*	-0.025***	-0.009*	-0.024***
	(0.005)	(0.005)	(0.005)	(0.005)
$NUMEST_{i,t-21,t-1}$	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$	0.026*	0.023*	0.027**	0.024*
	(0.013)	(0.013)	(0.013)	(0.013)
FE Month	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes
Observations	25,597	25,597	25,597	25,597
R-squared	0.010	0.008	0.010	0.008
Number of Firms	3,442	3,442	3,442	3,442

Clustered Standard Errors in Parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Panel A in Table 1.1 shows both the volatilities of mood change and the absolute change of mood as a measure of investors' mood swings' significant prediction of a positive price jump ratio. This positive predictability implies that when investors become

Panel B: Information Acquisition Measured by $LSECV_{i,t}$

Dependent Variable	(1) $LSECV_{i,t}$	(2) $LSECV_{i,t}$	(3) $LSECV_{i,t}$	(4) $LSECV_{i,t}$
$\sigma(Mood_{t-31,t-1})$	-1.282***	-1.511***		
	(0.321)	(0.326)		
$\lvert\overline{\Delta Mood_{t-31,t-1}}\rvert$			−57.222***	-62.151***
			(14.926)	(15.113)
$LSECV_{i,t-62,t-31}$	0.255***	0.255***	0.255***	0.255***
	(0.008)	(0.008)	(0.008)	(0.008)
$VIX_{t-21,t-1}$	0.003**		0.003**	
	(0.001)		(0.001)	
$EPU_{t-21,t-1}$		0.001***		0.001***
		(0.000)		(0.000)
$Size_{i,t-42,t-21}$	0.024**	0.025**	0.024**	0.025**
	(0.010)	(0.010)	(0.010)	(0.010)
$Turn_{i,t-42,t-21}$	0.037***	0.037***	0.037***	0.037***
	(0.005)	(0.005)	(0.005)	(0.005)
$Price_{i,t-42,t-21}$	−0.020*	-0.021**	-0.021**	-0.022**
	(0.011)	(0.011)	(0.011)	(0.011)
$RV_{i,t-42,t-21}$	0.015**	0.017**	0.014**	0.016**
	(0.007)	(0.007)	(0.007)	(0.007)
$NUMEST_{i,t-21,t-1}$	−0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$	0.032	0.033	0.033	0.034*
	(0.020)	(0.020)	(0.020)	(0.020)
DOW	Yes	Yes	Yes	Yes
FE Year-Quarter	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes
Observations	25,597	25,597	25,597	25,597
R-squared	0.841	0.841	0.841	0.841
Number of Firms	3,442	3,442	3,442	3,442

Clustered Standard Errors in Parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

moody, they acquire less fundamental information to learn about firms' earnings before they are released. Consequently, less earnings-related information is incorporated into the price before the announcements, resulting in a large price jump when the earnings are announced. For example, in column (3) an increase in the absolute change of mood by one standard deviation (2 basis points) is associated with a 1.15% (relative to the median jump ratio of 0.4) decrease in the proportion of earnings-related price impact that arises pre-announcement. Notably, all results across columns are robust after fixed effects and other controls that may explain investors' information acquisition. Panel B in Table 1.1 is the test, using the count of SEC file downloads as the proxy of investors' information acquisition. Not surprisingly, the results are consistent with those of Panel A (with the price jump measure). For instance, in column (1), an increase in one unit of the mean value of mood volatility (0.24) is associated with a 5% (relative to the median $LSECV$ value of 6.24) decrease in investors' downloads of company SEC files. As investors have mood swings, they are less willing to download the company's SEC files, showing a decrease in information demand.

Additionally, we also perform an analysis to disentangle the biasing effect from either negative or positive mood swings. We expect that the absolute change in mood, either upwards or downwards, biases investors' decision to acquire firm-specific information. First, we split the absolute change of mood by positive and negative daily percentage change. Second, we calculate the average absolute change of positive mood ($\overline{|\Delta Mood^{+}|}$) or negative mood ($\overline{|\Delta Mood^{-}|}$) independently in the most recent month extended back to 34 days before the earnings announcement.[20] Finally, we also control the proportion of positive or negative mood change days ($\%Positive, \%Negative$) in the month before the earnings announcement. Unsurprisingly, the results in Table 1.2 are consistent with Table 1.1. By splitting the mood swings into either the upward or downward direction, Panels A and B in Table 1.2 clearly show that the absolute change of positive or negative mood induces less firm-specific information acquisition proxied by pricing jump ratio and SEC file downloads respectively.

Based on the empirical evidence we find in the data, the mood has a significant impact on investors' learning about firm-specific information (here, earnings-related in the test). Compared to studies arguing sentiment or emotions irrationalize how investors respond to the given information (Mian and Sankaraguruswamy, 2012; Karampatsas et al., 2018; Guo et al., 2020), we find evidence in Table 1.1 and 1.2 to explore a different dimension in information incorporation into the valuation of assets. On the one hand, our study takes information that is endogenously acquired or learned by investors whose performance of intentional acquisition on information is subject to be influenced by affective states. On the other hand, information is exogenously endowed to investors whose analysis of given information subjects to behavioral biases in information processing literature. Our study aims to emphasize how affective states, mood swings, bias the agents' performance on acquiring the quantity of information rather than the quality of processing information by behavioral investors. In sum, the key distinction between these two paths in information theory is: investors' choice to acquire

[20]Because either absolute positive or negative change calculates the mood swings in this test, an issue may arise whereby the mood swings data may calculate after the initial days in the proxy of investors' information acquisition. For example, the absolute positive or negative mood changes may be calculated between $t-s$ to $t-1$. In such a case, mood swings may not fully capture the biasing effect in the period before $t-s$ if s is far less than 31. We therefore extend the data back to $t-34$ to mitigate this issue. However, the results are not sensitive to how long we extend the data to calculate the split mood swings.

Table 1.2: Positive or Negative Mood Swings Impact on Investors' Information Acquisition

This table presents the results of regressions of the price jump ratio and the counts of EDGAR SEC file downloads as the proxy for investors' firm-specific information acquisition on mood swings measured by either the average absolute positive change of daily mood or negative change of daily mood during the firm earnings announcement window. Columns (1)–(4) are based on the fixed-effect regression from equation (1.4): $Dep_{i,t} = \beta_0 + \beta_1 MoodSwing_{t-34,t-1} + X\delta + \epsilon_{i,t}$, where $Dep_{i,t}$ is either the price jump ratio ($jump_{i,t}$ in Panel A) from equation (1.2) or information demand measure that is the natural logarithm for the average of total SEC files searching volumes ($LSECV_{i,t}$ in Panel B) in the most recent month before the announcements. Control variables include: economic uncertainty proxies (VIX or EPU) and the Number of Analyst Forecast is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility, and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. We also control the proportion of positive or negative mood change days ($\%Positive, \%Negative$) in the month before the earnings announcement (see detailed definitions in Appendix A.2). Additionally, we control the day-of-the-week effect (DOW) in Panel B when we use $LSECV_{i,t}$ to measure investors' information acquisition. Detailed definition of all variables are available in Appendix A.2. Standard errors are clustered by both firm- and time-fixed effect in column (1)-(4). $***$, $**$, $*$ indicate statistical significance at the two-sided 1%, 5%, 10% levels, respectively.

Panel A: Information Acquisition Measured by Price Jump Ratio

Dependent Variable	(1) $Jump_{i,t}$	(2) $Jump_{i,t}$	(3) $Jump_{i,t}$	(4) $Jump_{i,t}$
$\overline{\lvert \Delta Mood^{+}_{t-34,t-1} \rvert}$	10.763***	8.907***		
	(2.786)	(2.782)		
$\%Positive$	−0.046	-0.016		
	(0.032)	(0.031)		
$\overline{\lvert \Delta Mood^{-}_{t-34,t-1} \rvert}$			7.251***	5.794**
			(2.730)	(2.723)
$\%Negative$			0.128***	0.083**
			(0.037)	(0.036)
$VIX_{t-21,t-1}$	−0.002***		-0.002***	
	(0.000)		(0.000)	
$EPU_{t-21,t-1}$		−0.000		-0.000
		(0.000)		(0.000)
$Size_{i,t-42,t-21}$	0.005	0.011	0.005	0.011
	(0.007)	(0.007)	(0.007)	(0.007)
$Turn_{i,t-42,t-21}$	−0.004	-0.002	-0.004	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
$Price_{i,t-42,t-21}$	0.005	0.004	0.005	0.004
	(0.007)	(0.007)	(0.007)	(0.007)
$RV_{i,t-42,t-21}$	−0.010*	-0.026***	-0.009*	-0.025***
	(0.005)	(0.005)	(0.005)	(0.005)
$NUMEST_{i,t-21,t-1}$	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$	0.026**	0.023*	0.027**	0.024*
	(0.013)	(0.013)	(0.013)	(0.013)
FE Month	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes
Observations	25,592	25,592	25,592	25,592
R-squared	0.011	0.009	0.011	0.008
Number of Firms	3,442	3,442	3,442	3,442

Clustered Standard Errors in Parentheses

$***p < 0.01, **p < 0.05, *p < 0.1$

Panel B: Information Acquisition Measured by $LSECV_{i,t}$

Dependent Variable	(1) $LSECV_{i,t}$	(2) $LSECV_{i,t}$	(3) $LSECV_{i,t}$	(4) $LSECV_{i,t}$
$\overline{\lvert\Delta Mood^{+}_{t-34,t-1}\rvert}$	-9.292^{**}	-10.577**		
	(4.513)	(4.532)		
$\%Positive$	-0.030	-0.033		
	(0.059)	(0.059)		
$\overline{\lvert\Delta Mood^{-}_{t-34,t-1}\rvert}$			-8.577**	-11.348***
			(3.902)	(3.985)
$\%Negative$			-0.061	-0.083
			(0.066)	(0.066)
$LSECV_{i,t-62,t-31}$	0.255***	0.255***	0.255***	0.255***
	(0.008)	(0.008)	(0.008)	(0.008)
$VIX_{t-21,t-1}$	0.003**		0.003**	
	(0.001)		(0.001)	
$EPU_{t-21,t-1}$		0.001***		0.001***
		(0.000)		(0.000)
$Size_{i,t-42,t-21}$	0.024**	0.024**	0.024**	0.024**
	(0.010)	(0.010)	(0.010)	(0.010)
$Turn_{i,t-42,t-21}$	0.037***	0.037***	0.037***	0.037***
	(0.005)	(0.005)	(0.005)	(0.005)
$Price_{i,t-42,t-21}$	-0.020*	-0.021**	-0.020*	-0.021**
	(0.011)	(0.011)	(0.011)	(0.011)
$RV_{i,t-42,t-21}$	0.015**	0.017**	0.015**	0.016**
	(0.007)	(0.007)	(0.007)	(0.007)
$NUMEST_{i,t-21,t-1}$	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$	0.033	0.033	0.033	0.033
	(0.020)	(0.020)	(0.020)	(0.020)
DOW	Yes	Yes	Yes	Yes
FE Year-Quarter	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes
Observations	25,592	25,592	25,592	25,592
R-squared	0.841	0.841	0.841	0.841
Number of Firms	3,442	3,442	3,442	3,442

Clustered Standard Errors in Parentheses

$***p < 0.01, **p < 0.05, *p < 0.1$

information can be thought of as *active learning*; however, investors' performance to analyze given information can be understood as *passive learning*.[21]

Therefore, if investors acquire or learn insufficient fundamental information caused by affective states (mood), economic decisions will always be sub-optimal in equilibrium even the information processing is under rational expectations. All in all, the mood swings add more risks, causing investors to fail to learn enough information about assets. Furthermore, in line with investors' heterogeneous learning across different assets proposed by Van Nieuwerburgh and Veldkamp (2009) and Veldkamp (2011), conditioning on the mood effect implies that the pricing models based on the symmetric information may not be effective as researchers expect in explaining the cross-section asset return predictability. Finally, as mood tends to be either positive or negative, assets learned by investors are sensitive to the volatile mood generate cross-sectional risk premia. Altogether, we conduct a normative analysis through investors' information acquisition channel by arguing mood as a behavioral factor mistakenly incorporated into pricing assets. In sum, using unconditional pricing models in situations in which investors' learning about assets are severely affected by mood contributes to the possibility of the exploration of hundreds of latent risk factors in empirical asset pricing studies.

1.2.3 Theoretical Development

As stated in Veldkamp (2011), the asset's risk (standard deviation of return) will be subject to its payoff information as understood by the average investor. In line with this rationale, other things being equal, an asset that the average investor learns less (more) about is more (less) risky for investors to hold and requires higher (lower) expected returns. Hence assets' risk will be sensitive to investors' level of learning about the assets' payoff. We discuss the theoretical implication of investors' information learning for asset pricing through the general stochastic discount factor (SDF) pricing model, equivalent to factor pricing models. More specifically, we argue in some detail that the behavioral factor-mood can cause the failure of classical (unconditional) beta representation models through the biased effect of insufficient information learning.

A risky asset's return is R_i is priced by SDF pricing model:

$$E(R_i) = \frac{1}{E(m)} - \frac{Cov(m, R_i)}{E(m)} \tag{1.5}$$

[21] See the study by (Veldkamp, 2011) for detailed discussions.

Multiplying and dividing by $Var(m)$,the beta representation is:

$$
\begin{aligned}
E(R_i) &= \alpha + \Big(\frac{Cov(m, R_i)}{Var(m)}\Big)\Big(-\frac{Var(m)}{E(m)}\Big) \\
E(R_i) &= \alpha + \beta_{i,m}\lambda_m. \\
E(R_i) &= \alpha + \rho_{i,m}\frac{\sigma(R_i)}{\sigma(m)}\lambda_m
\end{aligned}
\tag{1.6}
$$

where $\alpha \equiv 1/E(m)$.

Because the SDF m can be an affine function of factors (single-factor such as CAPM or multi-factor such as Fama-French factors etc.), factor models are equivalent to SDFs. The beta representation can be easily transformed into the factor pricing model as the SDF m is a linear combination of factors $f_1, ..., f_k$. For example:

$$
m = a + b_1 f_1 + ... + b_k f_k \tag{1.7}
$$

where a is a constant and $b_1, ..., b_k$ are the factor coefficients. Let $F = (f_1, ..., f_k)'$ and $m = a + b'F$ is an SDF for a constant a and the constant vector b. Therefore equation (1.5) can be written as:

$$
E(R_i) = \frac{1}{E(m)} - \frac{Cov(b'F, R_i)}{E(m)} \tag{1.8}
$$

As we normally work with the excess return $E(R_i^e) = E(R_i) - 1/E(m)$ and $E(mR_i^e) = 0$, equation (1.8) is equivalent to:

$$
\begin{aligned}
E(mR_i^e) &= E(R_i^e) + b'Cov(F, R_i^e) \\
E(R_i^e) &= -b'Cov(F, R_i^e) \\
E(R_i^e) &= -b'Var(F)Var(F)^{-1}Cov(F, R_i^e) = \lambda'\beta
\end{aligned}
\tag{1.9}
$$

Therefore, the β and λ are :

$$
\beta = Var(F)^{-1}Cov(F, R_i^e), \quad \lambda = -Var(F)b
$$

No matter which pricing models we begin from, in a symmetric information world, we customarily assume that the estimated unconditional betas in the right hand side regression carry full pricing information in respect of the risk factors (O'Hara, 2003). However, as stated in Van Nieuwerburgh and Veldkamp (2009), investors choose to learn about assets based on their initial heterogeneous beliefs about the covariance structure of assets' payoff. Because learning will affect the conditional variance of assets' payoff, the betas that are conditional on information that the average investor knows are different from the estimated betas based on unconditional pricing models.

If investors thoroughly learn all available information about assets' payoff, using the unconditional model to price does not generate significantly positive abnormal returns (α). An asset that the average investor learns more about, this implies a lower beta conditional on this information acquisition by investors. As proposed by Veldkamp (2011), learning does not change the correlation structure. However, it reduces the standard deviation of the asset return; other things being equal, $\sigma(R_i)$ is lower; thus, the $\beta_{i,m}$ in equation (1.6) is lower because the asset's information is learned by the average investor.

As we find in section 1.2.2, mood swings in either direction induce investors to acquire less value-relevant information. Hence, assets that the average investor understudies as a result of mood swings are relatively riskier than investors' well-researched assets. If investors' choice to learn information about assets' payoff is not comprehensive, their inadequate learning implies that using the unconditional betas from classical factor models or SDFs to price assets is severely inappropriate. More specifically, when investors do not learn or acquire enough information regarding an asset i's payoff x_i, this insufficient learning increases the asset return risk, such as $\sigma(R_i)$ in equation (1.6). Consequently, the actual $\sigma(R_i)$ becomes higher to investors. The use of unconditional betas that are assumed to capture full pricing information to risk factors will understate the asset's risk.

For instance, the CAPM beta measures the unconditional relationship between an asset return and the market return. The single-factor CAPM model can easily derive that the market beta conditional on investors' insufficient information caused by mood ($\beta^{nl}_{i,CAPM}$, where nl denotes insufficient learning) should be higher than the unconditional beta ($\beta_{i,CAPM}$).[22] Therefore, $\beta^{nl}_{i,CAPM}$ implies higher risk and a higher expected return ($E(R^e_i)$=$\beta^{nl}_{i,CAPM} * \lambda_{MKT}$) than $\beta_{i,CAPM}$ due to investors' insufficient learning indicating a higher $\sigma(R_i)$. By the same token, one can map the logic to multi-factor models to analyze the covariance structure $Cov(F, R^e_i)$ in equation (1.9) for Fama-French three or five factors and q-factor model by Hou et al. (2015). This rationale can be applied to the discussion of hundreds of "innovative" risk factors in the empirical asset pricing studies to explain traditional models' failure in asset pricing. In other words, the beta conditional on insufficient learning should be higher than the unconditional beta. Specifically, we can find that $\beta^{nl}_{i,m} > \beta_{i,m}$ due to the

[22]Note that the $\beta_{i,CAPM} = \frac{Cov(R_i,R_{MKT})}{\sigma^2(R_{MKT})} = \rho_{i,MKT}\frac{\sigma(R_i)}{\sigma(R_{MKT})}$. As investors do not learn enough information about the asset payoff, $\sigma(R_i)$ is higher conditional on the insufficient learning effect. Consequently, the $\beta^{nl}_{i,CAPM}$ is higher than the unconditional CAPM beta as the $\sigma(R_i)$ increases.

$\sigma^{nl}(R_i) > \sigma(R_i)$ in equation (1.6). Working with factor models, we can find that $\sigma(R_i^e)$ in the structure of $Cov(F, R_i^e)$ share similar characteristics based on investors' insufficient learning about the asset i ($Cov^{nl}(F, R^e) > Cov(F, R^e)$). Therefore, when pricing the assets that are marked by a severe lack of investors' learning with the factor model, the betas investors should use are β^{nl} rather than the unconditional betas β in equation (1.9), in fact, $\beta^{nl} > \beta$.

The classical assumption of *Homo economicus* indicates that economic agents do not make sub-optimal economic decisions or behaviors driven by psychological shortcomings. Nevertheless, as we find in the data, the empirical evidence shows when agents have mood swings, investors are less likely to acquire earnings-relevant information to learn about the company's performance than they are in a sober state. Therefore, we argue that mood has a significantly negative impact on investors' fundamental learning about assets. In other words, as investors become moody, they mistakenly rely on their feelings as part of pricing information and acquire less fundamental information to incorporate into the valuation of assets. This mood-driven learning deficit causes the asset risk ($\sigma^{nl}(R_i)$) to be higher than the scenario in which investors do not suffer the mood effect and rationally learn information about the risky asset ($\sigma(R_i)$). As we mentioned above, when using unconditional pricing models without considering this insufficient learning, the betas in classical pricing models are underestimated and do not fully capture the additional risk that is contributed by the mood in our study. Intuitively, we should expect a positive and significant abnormal return (α in equation (1.6)) can be found by using the underestimated beta models (CAPM, Fama-French, etc.). Furthermore, we do not expect all assets to be significantly affected by the mood effect.[23] As a result, only assets that are sensitive to this mood effect, causing investors' insufficient learning, are riskier to hold in an investment portfolio. To hold these mood-sensitive assets, traders require higher expected returns to compensate for the added risk by mood. In sum, the theoretical analysis of the mood effect causing investors' insufficient learning about assets yields the following empirical prediction:

Hypothesis 2: Assets more sensitive to mood swings are riskier due to investors' failure to acquire sufficient information about them; a higher expected return is required for investors to hold these mood-sensitive assets.

[23]It is hard to believe that all assets suffer the mood-biased effect on information acquisition. For example, large firms or companies in an industry that is transparent or easy to analyze for investors. In other words, learning about these firms is, to some extent, costless.

1.2.4 Data for Empirical Tests

We conduct empirical tests in line with the theoretical hypothesis using the Twitter mood index that exhibits three key characteristics. First, there are visible outliers, most of which last a very short time. These outliers fall into two categories. Some outliers repeat and are predictable: Holidays and celebrations such as Christmas, Easter, Thanksgiving and Mother's Day naturally score very highly on the Twitter Mood Index. These are not relevant for our study, as U.S. markets are closed on these days.[24] Other predictable holidays and celebrations - Valentine's Day, for example - do coincide with trading and we deal with this in our analysis. Other outlier days are due to unpredictable events. The saddest day in Figure 1.1 is the Sandy Hook Elementary School shooting on December 14th, 2012. Most, but not all, unpredictable outliers are sad days.

A second pattern not easily visible in Figure 1.1 is the day-of-the-week pattern. Fridays (and Saturdays) are systematically more happy than Mondays.[25] These are the day-of-the-week patterns explored in Hirshleifer et al. (2020). Since they too are predictable and seasonal we remove them from our analysis. Hirshleifer et al. (2020) also use monthly seasonalities, arguing that the months of January and March are positive mood months while Septembers and Octobers are negative mood months. This is less apparent in the Twitter series as a result of its third feature, the slow oscillation about the mean. Twitter mood peaked in early 2010 and late 2015, with troughs in late 2012/early 2013 and mid-2017. Once the effect of regular holidays is excluded, there is very little monthly seasonality in Twitter mood: January 2013 was a lot less happy than September 2015.

More importantly, Figure 1.1 shows that Twitter mood and the Baker-Wurgler (B&W) sentiment series follow very different paths over their common interval. Indeed, key swings in the two series have often been in opposite directions. The Pearson correlation between the orthogonalized B&W sentiment index and our monthly average Twitter mood index is around -0.5.[26] Alternatively, we also investigate the relation between our Twitter mood index and other sentiment or economic expectation measures. For instance, the Pearson correlations between the monthly average Twitter

[24] We conduct a robustness test by adding the lagged Twitter mood into the main regression to estimate the mood beta as a careful consideration for the market closed holiday effect. The results are consistent without adding the lagged Twitter mood.

[25] Abraham and Ikenberry (1994) study the trading behaviors of individual investors resulting in a 'weekend effect' from a relationship between the Friday stock return and the upcoming Monday return. Additionally, Birru (2018) finds that the long-short returns anomaly is related to the mood that has the day of the week patterns.

[26] The result by using the raw B&W sentiment index is similar to the orthogonalized measures.

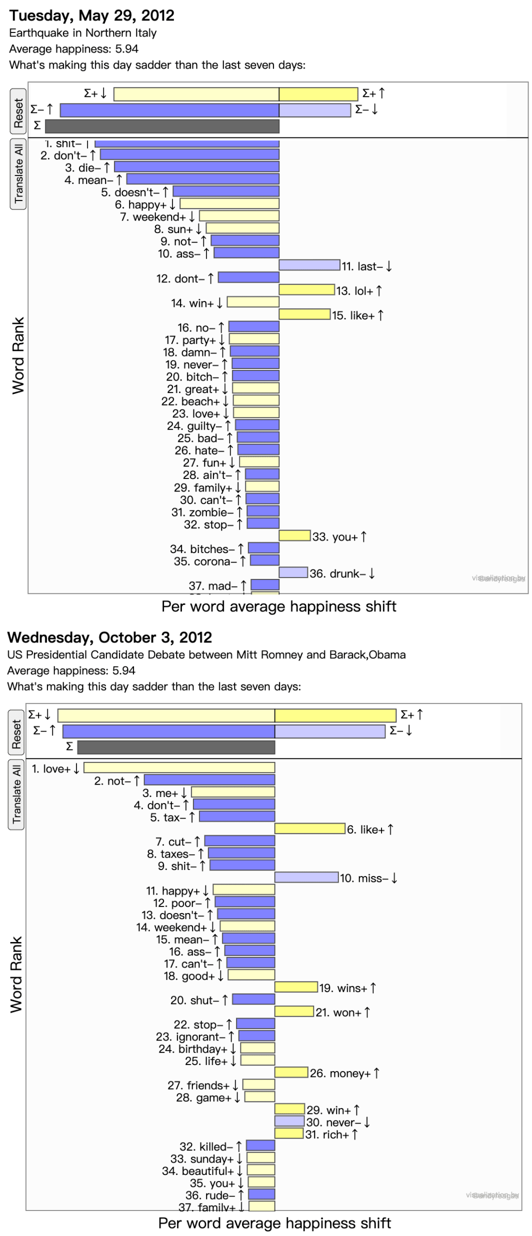


Figure 1.2: Examples for the Use of Word Frequency in Twitter Messages

mood index and Michigan Consumer Sentiment index is about -0.19 and -0.08 with the next year expectation of inflation. The inverse periodic cycle between these measures implies that Twitter mood is not pro-cyclical with respect to investors' sentiment to speculate in the market. More specifically, the moving swings such as the downward move from 2012 to 2013 and upward move from 2014 to 2015 are completely due to the technical measure for the use of word frequency on tweets, especially for exogenous events. For example, Figure 1.2 shows an increasing frequency of using negative words and decrease in the frequency of positive words in tweets on May 29^{th}, 2012 and October 3^{rd}, 2013 – days that coincided with the earthquake in Northern Italy and the U.S. presidential debate respectively.[27] Moreover, the measure for positive events such as holidays or exogenous happy events is related to an increase in the frequency of happy words and decrease in the frequency of negative words in tweets. Essentially, the Twitter mood measure is driven by exogenous events, periodic holidays and classical day-of-the week effect.

In comparison to studies using Twitter-based sentiment (Bollen et al., 2011; Yang et al., 2015), the Twitter mood measure in this study is more subject to the definition of affective state of mood in the psychological sense, free of the endogenous effect from the stock market. In addition, we channel the mood effect of insufficient information learning to the implications of asset pricing on a cross-sectional study rather than the aggregated market indices as addressed by Zhang et al. (2016), Shen et al. (2018) and Zhao (2020) who use the same Twitter data. Therefore, based on both the universe of mood data and the scope of economic theoretical motivation in our study, the sensitivity measure and implications on the variation of the Twitter mood index, mood beta, is different from the measures based on the B&W sentiment index such as sentiment beta reported in studies by Glushkov (2006) and Hirshleifer et al. (2020).

We use the daily change of the Twitter mood index from September 2008 to December 2016 as a proxy for public mood changes. One potential concern arising from the use of the daily change of Twitter mood index is that the index may contain textual information about economic conditions. Most macro conditions are measured at low frequency, such as monthly, quarterly. Therefore, it is a hurdle to orthogonalize the daily change of Twitter mood on potential economic condition information. Figure 1.3 plots the business cycles with the shaded area against the Twitter mood index in the upper panel and its 90 days moving average of daily percentage changes in the lower panel. As the two subplots show, the Twitter mood measure variations are

[27] The two screenshots in Figure 1.2 are taken from the Hedonometer website.

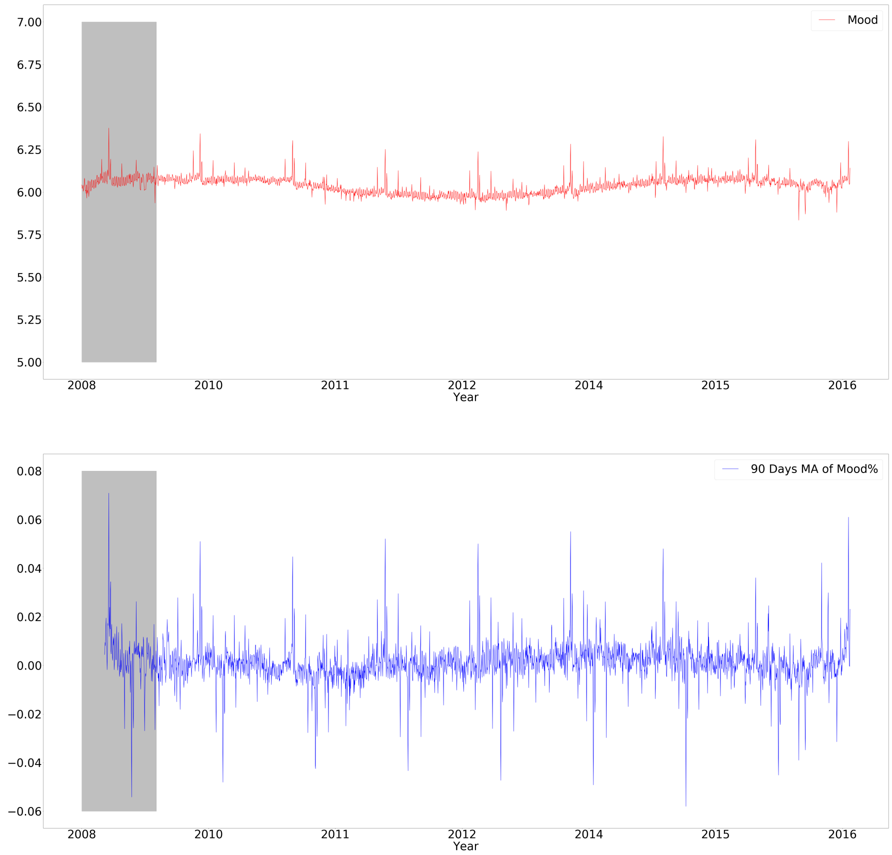


Figure 1.3: Twitter Mood vs. Business Cycles

not affected by the business recession.[28] For example, in the sample's recession time from September 2008 to August 2009, the swings of Twitter mood index movement in this particular time are not related to this business cycle. More importantly, during economic expansion periods (the rest of the years in the sample), long-term troughs and peaks in the Twitter mood index only vary with the characteristics of posted tweets.

Nevertheless, to further clarify the Twitter mood is affected by economic condition information, we look into the daily frequency of economic condition measures. Therefore, we download the daily frequency of economic uncertainty indices by Baker et al.

[28] The business cycle data is downloaded from Federal Reserve Bank of St. Louis website.

(2016). We use two measures from their study: one is the news-based economic uncertainty measure and the other is the Twitter-based economic uncertainty measure.[29] In fact, both of the economic uncertainty measures have negligible Pearson correlations with the percentage of daily Twitter mood index, which are -0.04 and -0.015 for the news-based and Twitter-based respectively. Additionally, the time series regression results do not show significant relations between the economic uncertainty measures and the Twitter mood percentage change. Therefore, the use of the percentage change of the Twitter mood index is not affected by potential economic condition information contained in the index. In the following analysis, we mainly handle the day-of-week effect and holiday effects during the market opens.

The raw mood data are at a daily frequency, and as already noted have a strong day-of-the-week effect. Furthermore, there are two festive days on which the market trades and which are predictably happy; Valentine's Day has an average mood score of 6.15 and Easter Monday has an average mood score 6.08, both of which are higher than the sample average mood score of 6.02. Including these two days in the calculations would lead to predictable outliers in the mood measure. Instead, the mood we wish to analyze is orthogonalized to any foreseeable patterns. Therefore, to obtain a completely exogenous factor for mood, we regress the change of daily mood from Twitter on dummies for days-of-the-week, Valentine's Day and the Easter holiday.

$$\Delta TwitterMoodScore_t = b_1 D_{Monday} + ...b_7 D_{Sunday} + b_8 D_{Valentine} + b_9 D_{Easter} + \epsilon_t \tag{1.10}$$

$$\Delta Mood_t = \epsilon_t$$

Table 1.3 shows that Monday and Sunday have a strong negative impact on the change of mood. This result is consistent with the perception that people are generally less happy on Mondays, with an associated decrease on Sunday since it precedes a Monday.[30] Not surprisingly, Thursday and Friday have strong positive significance. We take the regression residuals as a proxy for mood factor which is orthogonalized to the day-of-the-week effect and predicted holiday effects for the following cross-section stock return study.

Daily stock returns are taken from the Center for Research in Security Prices (CRSP) and financial fundamentals data from the CRSP/Compustat merged database.[31]

[29]The Twitter-based economic uncertainty index is available from 2011 on Economic Policy Uncertainty website.

[30]See studies by Croft and Walker (2001), Areni and Burger (2008), Ryan et al. (2010) and Stone et al. (2012).

[31]We merge CRSP returns data and the CRSP/Compustat merged database by PERMNO and LPERMNO. If there is no match between PERMNO and LPERMNO, we use the tickers and merge

Table 1.3: Day of The Week and Festival Effects

This table is regressions of the daily percentage change of the Twitter mood index on dummy variables for each day of the week, weekends and predictable festivals (Valentine's day and Easter Monday). The sample period of Twitter mood index in our study is from September 2008 to December 2016. $\Delta TwitterMoodScore_t = \frac{TwitterMoodScore_t - TwitterMoodScore_{t-1}}{TwitterMoodScore_{t-1}}$, where $TwitterMoodScore_t$ is the raw measure from Twitter mood index

	Monday	*Tuesday*	*Wednesday*	*Thursday*	*Friday*	*Saturday*	*Sunday*	*Valentines*	*Easter*	R^2
Model 1										
$Coeff$	-0.0026	-0.0002	0.0003	0.0012	0.0031					0.11
t_{Coeff}	-10.70	-1.48	1.80	5.53	13.24					
Model 2										
$Coeff$	-0.0026	-0.0002	0.0003	0.0012	0.0031	0.0003	-0.0020			0.13
t_{Coeff}	-10.70	-1.48	1.80	5.53	13.24	1.25	-10.11			
Model 3										
$Coeff$	-0.0026	-0.0003	0.0003	0.0011	0.0031	0.0002	-0.0021	0.0155		0.16
t_{Coeff}	-10.96	-1.97	1.80	5.48	13.35	0.94	-10.65	12.32		
Model 4										
$Coeff$	-0.0026	-0.0003	0.0003	0.0011	0.0031	0.0002	-0.0022	0.0155	0.0079	0.17
t_{Coeff}	-10.96	-1.97	1.80	5.48	13.35	0.94	-11.64	12.42	8.33	
Average Mood Score	6.015	6.014	6.016	6.023	6.042	6.043	6.031			

We retain all U.S.-based common stocks with share code (SHRCD) value 10 or 11 listed on the NYSE, AMEX, and NASDAQ with exchange code (EXCHCD) 1 or 31, 2 or 32 and 3 or 33 respectively. While we expect that large sensitivities to mood are likely in smaller stocks, we do not want our results to be driven by penny stocks. Therefore we exclude stocks with prices below \$2.50 or a market capital less than the 0.5$^{\text{th}}$ percentile. There are 4962 stocks in our sample analysis.

1.3 Mood Beta Estimation and Portfolios

1.3.1 Estimation of Mood Betas

Because not all assets are subject to the mood effect that causes investors' irrationally insufficient learning, we first identify the sensitivity of the stock returns to mood. We use the Fama-French five-factor model (Fama and French, 2015) augmented with Carhart's (1997) momentum factor as the benchmark model in our analysis. The factors are all taken from Kenneth R. French Data Library. Table 1.4 shows the Pearson correlation among factors and mood variables in our sample period from September 2008 to December 2016. There is no multi-collinearity problem in our time series analysis. The correlation between $\Delta Mood_t$ and the Fama-French factors are around 0.015. The correlation between $\Delta Mood_t$ and the momentum factor is very small and negative.

LPERMNO data based on tickers in the two databases. Detailed information about merging data and the definition of financial fundamentals can be found in Appendix A.1.

Table 1.4: Pearson Correlations among Factors

This table reports Pearson correlations between the change of daily Twitter mood and classical pricing factors that are downloaded from Kenneth R. French Data Library. The sample period in our study is from September 2008 to December 2016. $\Delta Mood_t$ is from equation (1.10), where the residuals are orthogonalized to the day-of-the-week effect and predicted holiday effects.

	MKT	SMB	HML	CMA	RMW	MOM	$\Delta Mood_t$
MKT		0.263	0.419	-0.151	-0.457	-0.395	0.013
SMB			0.122	0.009	-0.342	-0.097	0.018
HML				0.232	-0.462	-0.548	0.024
CMA					0.039	0.127	-0.019
RMW						0.290	-0.011
MOM							-0.0003

Following the similar method applied by Bali et al. (2017), for each stock we perform the following regression:

$$R_{i,t} = \alpha_i + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + b_i \Delta Mood_t + \epsilon_{i,t} \tag{1.11}$$

where $\Delta Mood_t$ is the daily percentage change of the Twitter mood score and $f_{m,t}$ is a vector of pricing factors. The regression coefficient on $\Delta Mood_t$ in equation (1.11) measures the time series sensitivity of each stock to Twitter mood innovations controlling for the other six benchmark factor effects.[32] This coefficient (b_i) is our mood beta. Intuitively, the larger the absolute value of b_i is, the more sensitive the stock return R_i is affected by mood.

We estimate mood betas from rolling regressions. Following Bali et al. (2016) we use windows of 200 daily observations. Since a naïve 200-observation window would ignore some IPO, merger and acquisition activities in a stock, our regression analysis follows the dynamic data-rolling method proposed by Bali et al. (2016) which takes into account all corporate actions.[33] Figure 1.4 plots the data distribution of the regression coefficient on $\Delta Mood_t$, which is the mood beta.

[32] In robustness checks we also include a lagged value of $\Delta Mood$ in the regression since the Twitter mood score has a strong mean-reverting time-series pattern. The Twitter mood score is measured for all Tweets posted within 24 hours of midnight (00:00 until 23:59). It is not hard to identify mood impact on stocks which are sensitive to mood during trading hours (09:30 to 16:00). However, mood out of trading hours may also affect stock returns. Therefore, part of the mood impact on stock return at day t may come from the mood between hours 16:00 to 23:59 at day $t-1$, so we add the lagged change in mood. Our results are unaffected by this adjustment for lagged Twitter mood changes.

[33] In fact, the regression coefficient of mood is largely insensitive to how stocks are incorporated into the sample.

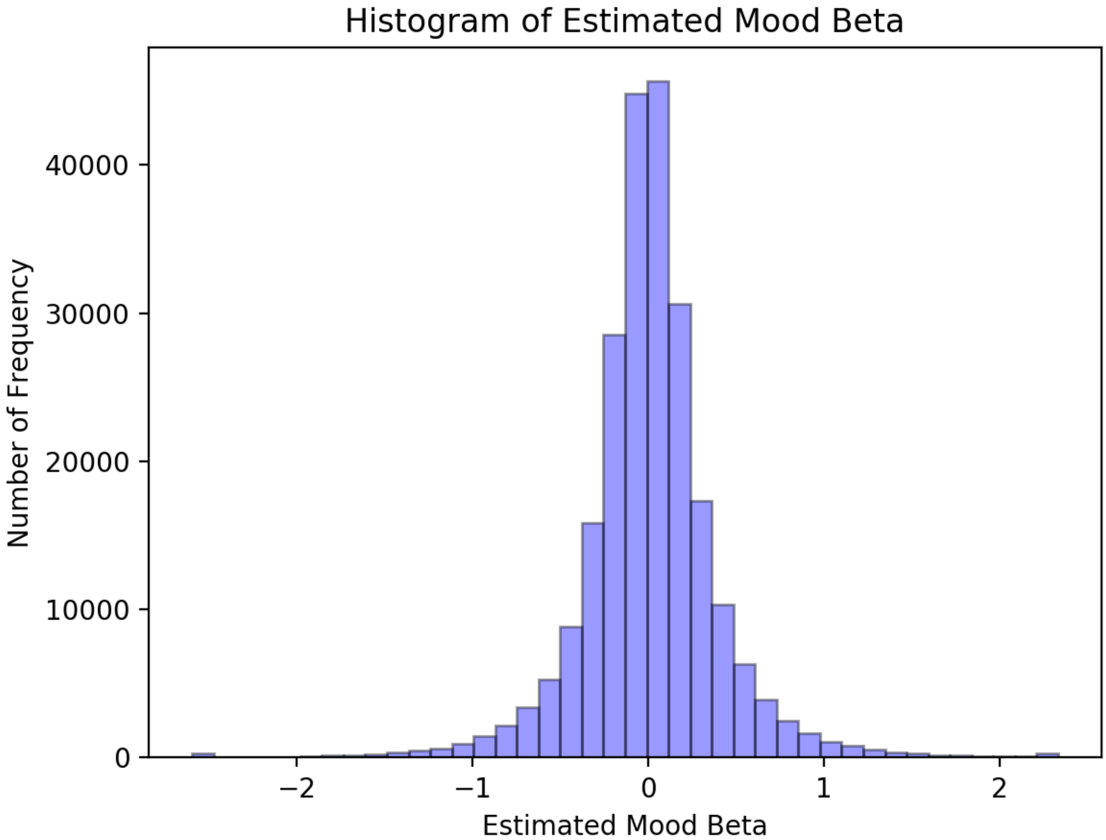


FIGURE 1.4: Estimated Mood Betas from Equation (1.11)

1.3.2 Identification of Mood Impact

We start the analysis by forming portfolios based on mood beta sorts and identifying whether the stocks that are sensitive to mood have high expected returns, as indicated in Hypothesis 2. From the end of December 2009 we sort our stocks based on historical 200-day rolling regression coefficients of $\Delta Mood_t$ into ten portfolios of ascending sensitivity to mood based on breakpoints derived from NYSE stocks, and calculate value-weighted portfolio excess returns for the next month.[34] Portfolio 1 includes stocks with the lowest regression coefficients on $\Delta Mood_t$, that is, stocks that are negatively sensitive to mood. Portfolio 10 includes stocks with the highest regression coefficients on $\Delta Mood_t$. Stocks in portfolio 10 are the most positively sensitive to mood. We then calculate the value-weighted average excess returns for each portfolio over the subsequent month.

[34]Our results are not sensitive to the use of all stocks breakpoints.

The top panel in Table 1.5 reports the average mood beta for the ten mood portfolios together with their monthly value-weighted excess returns. Standard errors are Newey-West-corrected. By construction, the magnitude of mood factor loading increases monotonically from portfolio 1 to portfolio 10. Stocks in portfolio 1 have negative sensitivity to mood change and an average mood beta of -0.58. The mood beta is markedly higher for portfolio 2 at just -0.23. Stocks in portfolio 10 have positive sensitivity to mood change, with a mood beta of 0.61. Again, moving from portfolio 10 to portfolio 9 sees a large decrease in mood beta to 0.26. While mood betas rise as we move from portfolio 2 to portfolio 9, the magnitudes of the changes are much smaller than when considering the extreme portfolios. Extreme sensitivity to mood (either positive or negative) is concentrated in a relatively small number of stocks that lie in portfolios 1 and 10.

The negative and positive mood betas are consistent with the study by Goetzmann et al. (2015), who argue that stock returns have co-movement patterns during optimistic and pessimistic mood days, which are proxied by weather in their study. This understanding can be also interpreted with the argument of Hirshleifer et al. (2020), who propose that the the correlation between stock returns and seasonal patterns is due to the effect of mood. In contrast to their seasonality argument of mood-congruent stock returns, we streamline the congruence of stock returns into two major categories: the tendency of positive and negative mood days measured and the change in the Twitter mood index. More specifically, the extreme sensitivity to public mood, either optimistic or pessimistic, arises from the congruence of investors' trading behaviors which are affected by the unconscious incorporation of mood as information rather than learning sufficient fundamental information about assets into their decision-making. On the one hand, for example, the positive mood sensitivity of stocks is caused by investors with optimistic mood bias tending to overprice stocks and place more irrational buying orders (Kliger and Levy, 2003; Ifcher and Zarghamee, 2011; Goetzmann et al., 2015; Kaustia and Rantapuska, 2016). On the other hand, for instance, negative mood sensitivity is caused by investors with pessimistic mood bias perceiving more risk and becoming more risk-averse. This leads to lower expected firm earnings and gives rise to a requirement for higher returns as compensation for the excess perceived risk biased by the mood (Shu, 2010; Jiang et al., 2019a).

The ten portfolio returns in the top panel of Table 1.5 have an approximate U-shape (see Figure 1.5) as portfolio 1 and portfolio 10 — the most mood-sensitive portfolios — generate higher monthly excess returns (1.66% and 1.65% respectively) than less mood-sensitive portfolios. In fact, from portfolio 5 to portfolio 9, the monthly

Table 1.5: Factor Regression for Monthly Excess Returns of Mood Sorted Portfolios, NYSE Breakpoints,Value-weighted Returns (12/2009-12/2016), 84 Months

For each month from 2010 to 2016, we sort stocks into 10 portfolios based on the factor loading (NYSE breakpoints) of $\Delta Mood_t$ in model (1.11) and calculate the value-weighted monthly excess return for each portfolio. We report the average regression coefficient of $\Delta Mood_t$ on each portfolio and the average portfolio monthly excess returns. The negative slop coefficients indicate stocks which are negatively sensitive to mood change. The positive slop coefficients indicate stocks which are positively sensitive to mood change. Standard errors are subject to Newey-West correction. H/L is the high-low portfolio which is half of a portfolio to long both mood-affected stocks (portfolios 1 and 10) and to short mood-insensitive stocks (portfolios 5 and 6). Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) data are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas across 10 portfolios based on the Fama-French five-factor regressions, ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA}CMA_t + \beta_{RMW}RMW_t + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \epsilon_{p,t}$) and the full behavioral factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{ST,p}ST_t + \beta_{LT,p}LT_t + \epsilon_{p,t}$).

	1	2	3	4	5	6	7	8	9	10	H/L
β_{Mood}	-0.58	-0.23	-0.14	-0.07	-0.02	0.03	0.09	0.16	0.26	0.61	
$Mean$	1.66%	0.97%	0.86%	1.18%	1.13%	1.14%	1.05%	1.07%	1.16%	1.65%	0.52%
t_{Mean}	3.94	2.62	2.19	4.06	3.51	3.17	3.15	2.97	2.76	4.02	3.47
$Fama-French$											
α	0.48	-0.21	-0.33	0.06	0.04	0.11	-0.05	0.01	-0.10	0.47	0.40
t_α	3.25	-1.33	-1.20	0.51	0.31	1.00	-0.41	0.09	-0.54	3.73	2.74
MKT	1.09	1.07	1.04	1.01	1.02	0.96	0.99	0.94	1.12	1.05	0.08
t_{MKT}	22.32	34.93	16.17	38.46	25.76	33.50	34.34	19.35	23.40	29.20	1.68
SMB	0.27	0.10	0.06	0.03	-0.18	-0.10	-0.07	0.02	0.10	0.17	0.36
t_{SMB}	4.15	1.38	0.74	0.60	-3.51	-2.36	-0.91	0.35	1.00	2.20	4.21
HML	0.04	-0.01	0.05	0.01	0.01	-0.03	-0.07	0.04	0.25	0.05	0.06
t_{HML}	0.36	-0.09	0.41	0.11	0.10	-0.48	-0.78	0.38	1.56	0.57	0.59
CMA	-0.19	0.01	0.16	-0.05	-0.02	0.07	0.18	0.14	0.05	0.06	-0.09
t_{CMA}	-1.29	0.04	1.46	-0.59	-0.14	0.60	1.82	1.01	0.27	0.52	-0.63
RMW	-0.14	-0.07	0.15	0.28	0.08	-0.11	0.05	0.02	-0.03	0.05	-0.03
$\bar{R}^2$	0.91	0.90	0.84	0.94	0.91	0.92	0.91	0.88	0.86	0.88	0.32
$CARH$											
α	0.50	-0.21	-0.38	0.06	0.02	0.09	-0.07	0.02	-0.06	0.51	0.45
t_α	3.86	-1.41	-1.40	0.50	0.18	0.84	-0.49	0.17	-0.30	4.05	4.22
MKT	1.08	1.07	1.05	1.01	1.02	0.96	0.99	0.93	1.12	1.04	0.07
t_{MKT}	24.29	35.90	17.70	38.19	26.91	41.19	33.18	18.35	22.67	30.40	1.88
SMB	0.28	0.10	0.04	0.02	-0.19	-0.11	-0.08	0.02	0.11	0.19	0.38
t_{SMB}	4.29	1.32	0.55	0.59	-3.58	-2.68	-1.05	0.43	1.14	2.68	4.88
HML	-0.01	-0.01	0.14	0.01	0.05	0.02	-0.05	0.02	0.17	-0.03	-0.06
t_{HML}	-0.10	-0.12	0.98	0.12	0.76	0.30	-0.43	0.18	1.14	-0.40	-0.55
CMA	-0.15	0.01	0.10	-0.05	-0.04	0.04	0.17	0.15	0.10	0.11	-0.02
t_{CMA}	-1.09	0.05	0.88	-0.57	-0.32	0.33	1.73	1.03	0.55	0.89	-0.13
RMW	-0.13	-0.06	0.12	0.27	0.07	-0.12	0.05	0.02	-0.01	0.07	0.00
t_{RMW}	-1.45	-0.87	0.95	3.57	1.13	-1.52	0.50	0.23	-0.09	0.65	-0.04
MOM	-0.09	0.00	0.15	0.00	0.07	0.08	0.05	-0.03	-0.12	-0.13	-0.18
t_{MOM}	-2.20	-0.03	1.76	0.03	1.78	2.71	1.20	-0.41	-1.50	-1.78	-3.34
$\bar{R}^2$	0.91	0.90	0.85	0.93	0.91	0.92	0.91	0.88	0.87	0.89	0.41
$CARH\&ST\&LT$											
α	0.54	-0.20	-0.39	0.03	0.02	0.08	-0.02	0.05	0.02	0.55	0.50
t_α	3.82	-1.25	-1.52	0.27	0.18	0.62	-0.13	0.35	0.08	4.18	3.69
MKT	1.04	1.07	1.03	1.02	1.01	0.95	0.98	0.93	1.04	1.04	0.06
t_{MKT}	17.39	28.65	26.11	30.86	30.49	37.72	27.46	17.05	17.50	24.21	1.18
SMB	0.27	0.10	0.04	0.03	-0.19	-0.11	-0.08	0.02	0.10	0.19	0.38
t_{SMB}	4.84	1.26	0.54	0.67	-3.51	-2.87	-1.22	0.39	1.12	2.90	5.51
HML	-0.06	-0.01	0.14	0.04	0.05	0.03	-0.10	-0.01	0.07	-0.08	-0.11
t_{HML}	-0.48	-0.17	1.32	0.52	0.62	0.41	-1.22	-0.05	0.61	-0.98	-1.19
CMA	-0.16	0.01	0.15	-0.01	-0.03	0.08	0.08	0.10	0.04	0.03	-0.09
t_{CMA}	-1.36	0.06	1.06	-0.05	-0.19	0.74	0.64	0.58	0.29	0.19	-0.61
RMW	-0.07	-0.06	0.12	0.23	0.07	-0.13	0.11	0.06	0.11	0.12	0.06
t_{RMW}	-0.65	-0.78	0.72	2.43	1.16	-1.42	1.13	0.58	0.72	0.93	0.49
MOM	-0.06	0.00	0.17	0.00	0.07	0.09	0.04	-0.04	-0.09	-0.15	-0.19
t_{MOM}	-1.41	0.01	1.68	-0.01	1.72	3.00	0.96	-0.50	-1.11	-1.98	-3.32
ST	0.11	0.01	0.11	0.01	0.04	0.09	-0.09	-0.06	0.13	-0.10	-0.06
t_{ST}	1.83	0.27	0.95	0.27	0.59	2.40	-1.76	-0.94	2.06	-1.35	-0.94
LT	0.09	0.01	-0.04	-0.09	-0.01	-0.05	0.17	0.09	0.22	0.15	0.15
t_{LT}	0.78	0.07	-0.35	-1.13	-0.10	-0.72	1.54	0.90	1.82	1.07	1.22
$\bar{R}^2$	0.92	0.90	0.84	0.93	0.91	0.92	0.92	0.88	0.87	0.89	0.41

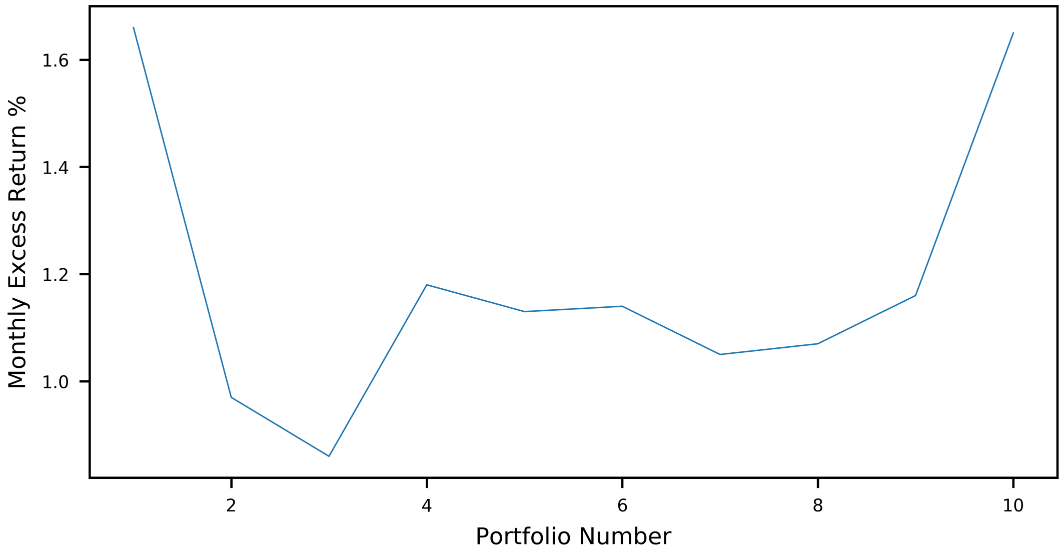


FIGURE 1.5: Portfolio Return Sorted by Mood Betas

excess returns are all around 1.1%. We form a high-low portfolio by taking a long position in both portfolio 1 (the negative-mood portfolio) and portfolio 10 (the positive-mood portfolio) at each investment period, and short positions in the mood-insensitive portfolios 5 and 6. This high-low strategy generates an excess return of 0.52% per month in-sample which is both statistically and economically significant (with a t-stat of 3.47). The mood portfolio results illustrate that stocks that are sensitive to mood in the financial market earn a higher monthly return than stocks which are less sensitive or are insensitive to mood. The mood beta from equation (1.11) already controls for the Fama-French five factors, and the Carhart momentum factor. Therefore, we propose that returns generated by our mood investment strategy can be considered as a possible new factor which cannot be explained by the most commonly used factors. These results are consistent with the study by Hirshleifer et al. (2020), who propose that the bias from mood contributes to the factor that is liable to be affected. Nevertheless, we focus on testing the argument that investors demand a risk premium (in the form of higher expected returns) as compensation for holding mood-sensitive assets about which they acquire less information instead of arguing for a particular misvaluation on factors within a risky asset payoff as in Hirshleifer et al. (2020).

The second panel in Table 1.5 reports the results of a time series regression of the monthly value-weighted returns on each of the ten portfolios sorted by mood beta and of the high-low strategy on Fama-French's five factors. Portfolio 1 generates a positive alpha of 0.48% per month while Portfolio 10 generates a positive alpha of 0.47% per month. Both are more than three standard errors from zero. Conversely, portfolios 2-9 generate smaller and statistically insignificant alphas of between -0.3 - +0.11% per month. Even though the adjusted-R^2 figures are around 90%, the significant positive alphas in the most mood-sensitive portfolios illustrate that the Fama-French five-factor model cannot capture the pricing information based on the mood effect. More importantly, the high-low strategy yields a statistically significant alpha of 0.40% per month.

In the third panel we add the momentum factor. The results are quite similar to the Fama-French five-factor model analysis. The momentum factor is significantly negative for portfolio 1 and for the high-low strategy. The sign of the loading on momentum for the high-low portfolio is unexpected, but the large residual alphas for the most mood-sensitive portfolios and long-short strategy suggests that the momentum factor cannot capture the mood effect on cross-section stock returns.[35]

The mood effect investigated in this study is clearly a behavioral factor, and we believe it is not captured by other documented behavioral patterns such as short- and long-run reversal effects in the stock market. The bottom panel of Table 1.5 adds the short- and long-term reversal factors to the regression. These reversal factors only serve to increase the magnitudes of the significant positive alphas in portfolios 1 and 10, and of the long-short strategy. Our key mood results are robust to controlling for the main behavioral factors in the literature.

More importantly, the empirical results are consistent with our theoretical results. The positive abnormal return in moody stocks (portfolios 1 and 10) is due to the unconditional betas from classical asset pricing understating the risk of stocks sensitive to the mood effect inducing insufficient fundamental information acquisition.

1.3.3 Robustness with Sentiment Beta

Baker and Wurgler (2006) argues that mood can viewed as a special case of sentiment and they propose that hard-to-value and hard-to-arbitrage stocks are sensitive to investors sentiment impact. One concern in this study is that stock sensitivity to the mood effect, the mood beta, may be a proxy of sentiment beta.[36] Therefore, in order to clarify the different implications between the mood effect and the sentiment effect in asset pricing that we explore in this study, we conduct a horse race and follow the

[35]The negative momentum factor exposures for the high-low portfolio can be explained as a consequence of psychological bias leading to irrational decisions of investment strategy and security selection. See the discussions by Shefrin and Statman (1985), Bikhchandani et al. (1992),Barberis et al. (1998), Daniel and Titman (2006) and Frazzini (2006).

[36]The study by Hirshleifer et al. (2020) investigates the same concern by running a horse race with sentiment beta analysis.

Table 1.6: Factor Regression for Monthly Excess Returns of B&W Sentiment Sorted Portfolios, NYSE Breakpoints,Value-weighted Returns (12/2009-12/2016), 84 Months

For each month from 2010 to 2016, we sort stocks into 10 portfolios based on the factor loading (NYSE breakpoints) of B&W sentiment Beta that is estimated from the model: $R_{i,t} = \beta_0 + \beta 1 MKT_t + b_i Sentiment_t + \epsilon_{i,t}$ and calculate the value-weighted monthly excess return for each portfolio. We report the average regression coefficient of the B&W sentiment on each portfolio and the average portfolio monthly excess returns. The negative slop coefficients indicate stocks which are negatively sensitive to investor sentiment. The positive slop coefficients indicate stocks which are positively sensitive to investor sentiment. Standard errors are subject to Newey-West correction. H/L is the high-low portfolio which is half of a portfolio to long both sentiment-sensitive stocks (portfolios 1 and 10) and to short sentiment-insensitive stocks (portfolios 5 and 6). Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) data are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas across 10 portfolios based on the Fama-French five-factor regressions, ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p} MKT_t + \beta_{SMB,p} SMB_t + \beta_{HML,p} HML_t + \beta_{CMA} CMA_t + \beta_{RMW} RMW_t + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p} MKT_t + \beta_{SMB,p} SMB_t + \beta_{HML,p} HML_t + \beta_{CMA,p} CMA_t + \beta_{RMW,p} RMW_t + \beta_{MOM,p} MOM_t + \epsilon_{p,t}$) and the full behavioral factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p} MKT_t + \beta_{SMB,p} SMB_t + \beta_{HML,p} HML_t + \beta_{CMA,p} CMA_t + \beta_{RMW,p} RMW_t + \beta_{MOM,p} MOM_t + \beta_{ST,p} ST_t + \beta_{LT,p} LT_t + \epsilon_{p,t}$).

	1	2	3	4	5	6	7	8	9	10	H/L
$\beta_{Sentiment}$	-0.103	-0.045	-0.028	-0.016	-0.007	0.001	0.01	0.02	0.033	0.077	
$Mean$	0.58%	0.408%	0.851%	0.782%	0.746%	0.763%	0.712%	1.037%	1.037%	0.783%	-0.073%
t_{Mean}	0.984	0.881	1.893	1.491	1.882	1.789	2.019	3.047	1.726	1.755	-0.389
$Fama-French$											
α	-0.807	-0.849	-0.451	-0.483	-0.415	-0.327	-0.224	-0.08	-0.44	-0.314	-0.189
t_α	-3.567	-4.111	-2.357	-2.074	-3.379	-2.273	-1.504	-0.506	-2.511	-1.617	-1.268
$\bar{R}^2$	0.872	0.802	0.854	0.848	0.922	0.871	0.842	0.875	0.845	0.806	0.372
$CARH$											
α	-0.782%	-0.848%	-0.452%	-0.486%	-0.429%	-0.34%	-0.196%	-0.07%	-0.47%	-0.328%	-0.17%
t_α	-3.411	-4.028	-2.305	-2.132	-3.436	-2.631	-1.34	-0.442	-2.541	-1.667	-1.181
$\bar{R}^2$	0.872	0.799	0.852	0.846	0.922	0.87	0.846	0.874	0.848	0.805	0.376
$CARH\&ST\&LT$											
α	-0.605	-0.81%	-0.385%	-0.466%	-0.38%	-0.334%	-0.274%	-0.1%	-0.495%	-0.4%	-0.146 %
t_α	-2.53	-4.022	-2.088	-2.03	-3.123	-2.472	-1.923	-0.629	-2.726	-1.94	-0.979
$\bar{R}^2$	0.896	0.815	0.856 0.847	0.924	0.867	0.854	0.879	0.845	0.809	0.39	

study by Hirshleifer et al. (2020) to estimate the sentiment beta that is stock return sensitivity to B&W sentiment index. We repeat the univariate portfolio analysis based on the estimated sentiment beta.

We regress stock monthly return on the CRSP-monthly value-weighted index return and the B&W sentiment index orthogonalized to macroeconomic variables with the most recent 5 years of monthly data (60 observations). We require that stocks have at least 30 monthly observations for the sentiment beta regressions. At the end of 2009, we sort stocks into 10 portfolios based on the estimated sentiment beta and calculate the next month value-weighted excess return.

Table 1.6 reports the summarized results of the sentiment beta sorting portfolios. By construction, the magnitude of sentiment beta increases from portfolio 1 to 10. Stocks in portfolio 1 and 10 have negative and positive sensitivity to investor sentiment, with average sentiment beta -0.103 and 0.077 respectively. The 10 portfolios' returns do not appear to generate the same U-shape pattern as the Figure 1.5 shown for the mood beta portfolios. More importantly, the 10 sentiment beta portfolios rarely generate significant returns, except the portfolios 7 and 8. The negative sentiment beta portfolio (P1) has a monthly excess return of 0.58% and the positive sentiment beta portfolio (P10) generates a relatively higher excess return of 0.78% per month. However, the excess returns from negative and positive sentiment beta portfolios are not statistically significant. Additionally, the high-low portfolio built by taking a long position in both portfolios 1 and 10, and short positions in the portfolios 5 and 6 generates insignificant excess returns.

The other panels in Table 1.6 reports the results of the alphas of the monthly value-weighted returns on each of the 10 portfolios sorted by sentiment beta and of the high-low portfolios on Fama-French's five factors, momentum factor, short- and long- term reversal factors. By regressing the pricing factors, the 10 portfolios' excess returns are deflated to generate negative significant alphas. Even the negative alphas in portfolios 1 and 10 are similar to the findings of Glushkov (2006), who argues that sentiment-sensitive stocks earning a lower return. The results are statistically exhausted by pricing factors since the raw returns showed in the upper panel are not significant.

We notice that the sample period in our study stretches from 2009 to 2016, which is a shorter interval than in the study by Glushkov (2006). The shorter period could be the potential reason for which the sentiment beta portfolios do not exhibit any clear patterns. However, Gluchkov's method to estimate sentiment beta may be subject to statistical errors such as the recalculation of the change of sentiment index as the new sentiment variable in the regression or the addition of other pricing factors.[37]

More importantly, the argument in Glushkov (2006) is counter-intuitive, claiming that stocks containing sentiment-noise trader risk earn lower expected returns. His findings imply that negative risk premia are required by investors to hold these riskier stocks that are subject to non-fundamental risk. He states that positive sentiment-sensitive stocks are subject to mispricing and that the low subsequent returns are the result of corrections of mispricing. However, there is no specified channel to indicate the implications of sentiment-sensitive stocks earning lower returns. In contrast, our study argues for the existence of a clear risk channel in moody stocks, which is caused by less fundamental information being incorporated into asset valuation. In such cases, stocks subject to this mood effect are more risky and investors demand higher compensation to hold these moody stocks. There is no specified channel to indicate the implications of sentiment-sensitive stocks earning lower returns.

[37] Additionally, Ang et al. (2006) state that adding many pricing factors in the time series regression to estimate the tested sensitivity (beta) of the interest variable may unconsciously add more noise into the model.

1.3.4 Financial Characteristics

The empirical results concerning the excess returns of mood-sensitive portfolios conditional on the Fama-French and Carhart factors demonstrate that mood does indeed have a significant impact on cross-sectional stock returns. Clearly, not all stocks are affected by mood and it is interesting to characterize the kinds of firms that are more likely to be mood-sensitive.

The firm-level data are drawn from the merged CRSP-Compustat database. Size is measured by the market value of equity. The book-to-market ratio is calculated using the method of Fama and French (1992). To identify whether mood-sensitive stocks pay less in dividends we calculate both the dividend yield (D/Y) and the percentage of companies that are paying dividends. For the measurement of profitability, we consider operating cash flow (OCF), earnings per share (EPS), return on assets (ROA), EBITA/Assets and sales revenue. We measure the age of firms by calculating the total number of years for which data are available in the CRSP database back to 1926. To measure financing activity, we consider book leverage and external financing (EF) scaled by asset growth. We also consider the tangible asset ratio (PPE/Asset) and research and development (R&D) expenditure in the analysis. Finally, idiosyncratic risk is measured by taking the RSE of residuals from the Carhart model. Detailed information about the definition of financial fundamentals is available in Appendix A.1 .

Table 1.7 reports the financial characteristics of 10 value-weighted portfolios formed by sorting mood factor loadings. While mood is a specific and distinct type of sentiment, we expect there to be commonality between our findings and those of, for example, Baker and Wurgler (2006) and Glushkov (2006) and to a large extent this is indeed the case. In summary, we find that mood-sensitive stocks, regardless of whether their mood-sensitivity is negative or positive, are small in size, relatively young, pay less in dividends, have more expenditure on R&D, are not profitable, engage in more external financing and have higher levels of idiosyncratic risk. This is in line with previous work identifying more sentiment-sensitive stocks as "hard-to-value and difficult-to-arbitrage." More importantly, the evidence of financial characteristics is consistent with the argument made by Bushee and Friedman (2016), who state that stocks with lack of disclosure are more likely to be invested in by noise traders or unsophisticated investors who are in turn more likely to be affected by non-fundamental factors such as mood. Moreover, mood-sensitive stocks, to some extent, shed light on the study by De Long et al. (1990), who argue that noise traders add risk and contribute to mispricing induced by sentiment. Therefore, mood can be thought of as another trigger that induces noise trading. In contrast to expectations and previous findings, however, there is no clear evidence that mood-sensitive stocks have different book-to-market or tangible asset ratios or different asset growth rates than less mood-sensitive stocks. Of course, given that our results show that mood-sensitive stocks to offer higher average returns than mood-insensitive stocks, while Glushkov (2006) finds the complete opposite for sentiment-sensitive stocks, our approach to the selection of stocks is different to those adopted by previous analyses. We now briefly discuss the defining characteristics of mood-sensitive stocks.

Both negative (portfolio 1) and positive (portfolio 10) mood stocks tend to have lower market values. As the sensitivity to mood decreases, firm size increases. Portfolios 5 and 6 (mood-insensitive firms) have the largest market capital, and are on average about 3.5 times the size of firms in either portfolio 1 or 10. Consistent with these findings, Lee et al. (1991) argue that individual investors who are more responsive to sentiment shifts have significant impact on smaller stocks. The size characteristic in mood-sensitive stocks found in our sample is also consistent with the studies of Baker and Wurgler (2006) and Glushkov (2006).

The B/M ratios are slightly higher for mood-sensitive stocks; however, the differences are not statistically significant. This contrasts with Baker and Wurgler's (2006) study which finds that firms with extreme values for B/M are more subject to the impact of investor sentiment but is in line with our findings in Table 1.5 that the value factor is not relevant to the explanation of mood-beta sorted portfolio returns.

The dividend yield is around 0.2 per share in portfolios 1 and 10, much less than the 0.5 per share found in portfolios 5 and 6. Similarly, only 33% of mood-sensitive firms pay dividends, much fewer than the 57% of mood-insensitive firms paying dividends. Again, this is consistent with the findings of Baker and Wurgler (2006) and Glushkov (2006), who show that non-dividend-paying stocks are more subject to investor-sentiment change.[38]

Operating cash flow, earnings per share, return on assets, company age and sales revenue each present an inverted U-shape, showing that mood-sensitive stocks are less profitable. These differences can be large: mood-insensitive stocks generate more than three-times the operating cash flow of mood-sensitive stocks and return on assets is around 3% across portfolios in low mood sensitivity-portfolios but highly negative in portfolios 1 and 10. While age is related to mood-sensitivity, moody stocks are past their early teenage years. The average moody stock is 21-22 years old, younger than

[38] Chung et al. (2012) find that sentiment impact is more significant on non-dividend-paying stocks during economic expansion states.

Table 1.7: Mean of Statistics on Financial Characteristics of 10 Mood Portfolios

We calculate the average value of financial variables for each portfolio across our sample period. Mkt. Cap is the total market value of equity. B/M is the book value of equity over market value of equity. D/Y is dividend paid per share. Div. is the probability of firms paying a dividend. OCF is operating cash flow. EPS is earnings per share. ROA is return on assets calculated as net income over total assets. EBITDA/Assets is calculated as earnings before interest over total assets. Lever. is total debt over the book value of total assets. PPE/Assets is the value of property, plant and equipment divided by the book value of total assets. R&D is research and development expenditure divided by book value of total assets. Revenue is total sales revenue. Asset growth is the percentage change of total assets between two fiscal years. EF is external financing calculated as difference between asset growth and percentage change of retained earnings. RE is the level of retained earnings between year t to year $t+1$. Age is measured as the date from which data is first available in the database up to December 2016. Risk is idiosyncratic risk measured as the RSE of residuals from the Carhart pricing model for each stock (see detailed information in Appendix A.1).

Port.	Mkt.Cap	B/M	D/Y	Div.	OCF	EPS	ROA	EBITDA/Assets	Lever.	PPE/Assets	R&D	Revenue	Asset growth	EF	RE	Age	Δ RE	Risk
1	2239.18	0.71	0.22	0.33	205.71	0.39	-0.05	0.03	0.17	0.40	0.06	1905.99	0.65	0.74	303.84	21.33	-0.08	0.029
2	3673.01	0.67	0.32	0.46	356.53	1.04	0.01	0.08	0.17	0.40	0.04	3356.67	0.72	0.73	776.54	24.02	-0.06	0.021
3	6816.13	0.64	0.44	0.54	624.73	1.39	0.02	0.09	0.16	0.40	0.03	5231.95	0.77	0.75	2019.37	26.02	-0.07	0.019
4	8280.46	0.63	0.50	0.57	754.20	1.54	0.03	0.10	0.16	0.40	0.03	6122.44	0.61	0.57	2550.52	26.97	-0.01	0.018
5	8546.49	0.63	0.52	0.57	806.64	1.61	0.03	0.10	0.17	0.41	0.03	6305.89	0.69	0.61	2341.86	27.32	-0.02	0.017
6	8203.45	0.64	0.51	0.57	811.12	1.62	0.03	0.10	0.17	0.42	0.03	6191.50	0.65	0.58	2470.35	27.37	0.04	0.017
7	6799.51	0.65	0.48	0.54	700.14	1.52	0.03	0.10	0.16	0.43	0.03	5511.20	0.60	0.53	2054.69	26.81	0.04	0.018
8	5486.10	0.65	0.41	0.50	576.66	1.35	0.02	0.09	0.17	0.43	0.03	4639.71	0.64	0.56	1545.79	26.07	0.06	0.019
9	3520.94	0.69	0.29	0.40	367.00	0.91	0.00	0.08	0.17	0.43	0.04	3048.98	0.75	0.70	849.98	23.55	0.05	0.023
10	2117.83	0.74	0.22	0.33	241.61	0.36	-0.04	0.04	0.17	0.44	0.06	1981.35	0.71	0.73	364.40	21.83	-0.01	0.030

the the mood-insensitive stocks, which are on average 27 years old, but they can hardly be characterized as "new" firms.

Three variables display clear U-shaped relationships with mood betas. Research and development expenditure (R&D), external financing (EF) and idiosyncratic risk are all high for mood-sensitive (moody) stocks, and low for mood-insensitive (sober) stocks.

In general, these U-shaped patterns are consistent with the findings of the sentiment-sensitivity literature. For example, the relation between mood and idiosyncratic risk is noted by Glushkov (2006); however, the intuition behind this specific empirical finding differs. Merton et al. (1987) proposed that when investors do not diversify their portfolio, expected return and idiosyncratic risk have a positive relation. More specifically, De Long et al. (1990) and Lee et al. (1991) state that if noise traders do not trade randomly across assets, the risk created by noise traders cannot be mitigated with diversification. In equilibrium, risk from stochastic investor sentiment will be priced accordingly. The high idiosyncratic risk in our moody stocks illustrates that unsophisticated investors are trading stocks subject to mood swings and that these stocks entail higher firm-specific risk which is not priced in by the traditional factor asset pricing model. While it is outside the scope of this study to consider the causal link between idiosyncratic risk and investor mood, we believe there is the possibility that part of the idiosyncratic risk derives from the trading activities of noise or mood traders (traders who are more likely to be affected by their mood). Hence, a more comprehensive study on this topic is an opportunity for future research.

1.4 Mood Factor Construction and Pricing Power

As stated in section 1.2.3, when mood causes investors to acquire less information about risky assets, the betas should be conditional on this effect. Using unconditional pricing models will lead to positive abnormal returns, which can be considered risk premium to compensate for holding moody stocks. Once the additional risk induced by mood is controlled in the model, we should expect a significant reduction in the abnormal returns as the moody stocks' information about investors' insufficient learning triggered by mood is incorporated into the pricing. Therefore, we follow classical empirical asset pricing studies to conduct risk factor testing to verify our argument about the mood effect.

Before constructing the mood-mimicking portfolio, we re-conduct the factor analysis above based on absolute mood betas; that is, we use NYSE breakpoints of the

Table 1.8: Factor Regression for Monthly Excess Returns of Absolute Mood Beta Sorted Portfolios, NYSE Breakpoints,Value-weighted Returns (12/2009-12/2016), 84 Months

For each month from 2010 to 2016, we sort stocks into 10 portfolios based on the absolute factor loading (NYSE breakpoints) of $\Delta Mood_t$ in model (1.11) and calculate the value-weighted monthly excess return for each portfolio. Standard errors are subject to Newey-West correction. H/L is the high-low portfolio which is to long mood-affected stocks (portfolio 10) and to short mood-insensitive stocks (portfolio 1). Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas across 10 portfolios based on the Fama-French five-factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \epsilon_{p,t}$) and the full behavioral factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{ST,p}ST_t + \beta_{LT,p}LT_t + \epsilon_{p,t}$).

	1	2	3	4	5	6	7	8	9	10	H/L
$\beta_{\lvert Mood \rvert}$	0.01	0.04	0.07	0.10	0.13	0.17	0.21	0.28	0.38	0.68	
$Mean$	0.99%	1.04%	0.96%	1.17%	1.12%	0.82%	1.08%	0.78%	1.19%	1.52%	0.53%
t_{Mean}	2.95	3.22	2.79	3.54	3.06	2.29	2.71	1.65	2.44	4.02	3.16
$Fama-French$											
α	-0.11	-0.06	-0.18	0.10	-0.01	-0.25	-0.13	-0.54	0.03	0.36	0.47
t_α	-0.79	-0.60	-1.16	0.80	-0.08	-1.04	-0.60	-3.28	0.18	3.19	2.49
MKT	1.02	1.00	1.03	0.97	1.01	0.96	1.09	1.18	1.10	1.04	0.02
t_{MKT}	23.57	30.40	23.03	28.97	31.24	14.45	24.71	22.05	26.15	25.74	0.30
SMB	-0.16	-0.07	-0.19	-0.02	0.00	0.01	0.01	0.19	0.31	0.16	0.32
t_{SMB}	-3.09	-0.83	-1.76	-0.27	-0.01	0.14	0.16	2.37	2.43	2.49	3.81
HML	0.04	-0.08	0.02	-0.08	-0.01	0.05	0.17	0.05	0.14	-0.04	-0.08
t_{HML}	0.65	-1.25	0.22	-0.76	-0.10	0.31	1.32	0.42	1.47	-0.49	-0.88
CMA	0.03	0.10	0.13	0.14	0.10	0.15	0.02	0.13	-0.39	0.05	0.02
t_{CMA}	0.27	0.64	0.70	1.02	0.82	1.13	0.13	0.68	-2.55	0.53	0.15
RMW	0.01	0.08	0.09	0.06	0.13	0.01	0.05	-0.20	-0.20	0.03	0.02
t_{RMW}	0.15	0.71	0.76	0.69	1.92	0.04	0.44	-1.28	-1.94	0.33	0.19
$\bar{R}^2$	0.92	0.91	0.87	0.89	0.93	0.82	0.87	0.87	0.88	0.91	0.11
$CARH$											
α	-0.15	-0.07	-0.19	0.11	-0.02	-0.27	-0.12	-0.51	0.08	0.38	0.53
t_α	-1.27	-0.72	-1.17	0.84	-0.18	-1.04	-0.50	-3.14	0.48	3.76	3.69
MKT	1.02	1.00	1.03	0.96	1.01	0.96	1.09	1.17	1.09	1.03	0.01
t_{MKT}	29.54	29.71	22.95	26.64	33.51	14.14	22.61	23.06	23.23	28.40	0.15
SMB	-0.17	-0.07	-0.19	-0.02	0.00	0.00	0.02	0.20	0.34	0.17	0.35
t_{SMB}	-3.77	-0.87	-1.85	-0.21	-0.08	0.05	0.24	2.39	2.99	2.66	4.24
HML	0.12	-0.06	0.05	-0.10	0.01	0.08	0.14	0.00	0.03	-0.08	-0.20
t_{HML}	1.93	-0.84	0.40	-1.03	0.21	0.41	0.96	0.00	0.38	-0.87	-2.12
CMA	-0.02	0.09	0.12	0.16	0.09	0.13	0.04	0.16	-0.32	0.07	0.10
t_{CMA}	-0.21	0.56	0.62	1.14	0.73	0.91	0.25	0.85	-2.20	0.75	0.59
RMW	-0.01	0.07	0.08	0.07	0.12	0.00	0.06	-0.19	-0.18	0.04	0.05
t_{RMW}	-0.14	0.67	0.75	0.76	1.90	-0.01	0.50	-1.20	-1.67	0.44	0.56
MOM	0.13	0.03	0.04	-0.04	0.03	0.05	-0.05	-0.08	-0.18	-0.07	-0.19
t_{MOM}	4.32	0.76	0.92	-0.52	1.56	0.59	-0.76	-1.51	-2.95	-1.49	-3.68
$\bar{R}^2$	0.93	0.91	0.87	0.89	0.93	0.82	0.87	0.87	0.89	0.91	0.20
$CARH\&ST\&LT$											
α	-0.14	-0.14	-0.14	0.10	-0.07	-0.21	-0.07	-0.49	0.13	0.41	0.55
t_α	-1.15	-1.44	-0.75	0.70	-0.63	-0.87	-0.30	-2.96	0.68	3.61	3.21
MKT	0.99	1.03	1.00	0.99	1.04	0.90	1.05	1.12	1.03	1.03	0.04
t_{MKT}	34.41	34.74	16.91	24.58	27.70	19.42	20.64	21.51	20.44	20.51	0.59
SMB	-0.18	-0.07	-0.20	-0.01	0.00	-0.01	0.01	0.19	0.33	0.17	0.35
t_{SMB}	-4.34	-0.83	-2.07	-0.17	0.03	-0.13	0.15	2.32	2.96	2.82	4.90
HML	0.09	0.02	-0.02	-0.08	0.08	0.00	0.08	-0.03	-0.03	-0.12	-0.21
t_{HML}	1.89	0.25	-0.17	-0.84	1.56	-0.01	0.65	-0.24	-0.33	-1.20	-2.03
CMA	0.00	0.20	0.05	0.14	0.16	0.09	-0.01	0.20	-0.36	0.02	0.02
t_{CMA}	0.00	1.20	0.23	1.01	1.60	0.55	-0.05	0.93	-2.46	0.14	0.08
RMW	0.02	-0.03	0.16	0.04	0.04	0.10	0.13	-0.15	-0.10	0.08	0.06
t_{RMW}	0.33	-0.23	1.20	0.42	0.67	0.55	1.13	-0.76	-0.77	0.74	0.49
MOM	0.15	0.04	0.04	-0.05	0.03	0.07	-0.04	-0.05	-0.16	-0.07	-0.22
t_{MOM}	4.01	0.87	1.23	-0.76	1.29	0.74	-0.52	-0.97	-2.74	-1.45	-3.50
ST	0.11	0.08	0.00	-0.09	0.01	0.10	0.04	0.16	0.09	-0.06	-0.18
t_{ST}	1.99	1.51	0.03	-1.26	0.19	0.84	0.68	3.39	1.92	-0.95	-1.94
LT	0.02	-0.24	0.16	-0.02	-0.18	0.18	0.15	0.03	0.14	0.11	0.09
t_{LT}	0.35	-3.57	1.31	-0.19	-3.16	1.36	1.23	0.27	1.01	1.01	0.66
$\bar{R}^2$	0.93	0.91	0.87	0.89	0.93	0.82	0.87	0.88	0.89	0.91	0.23

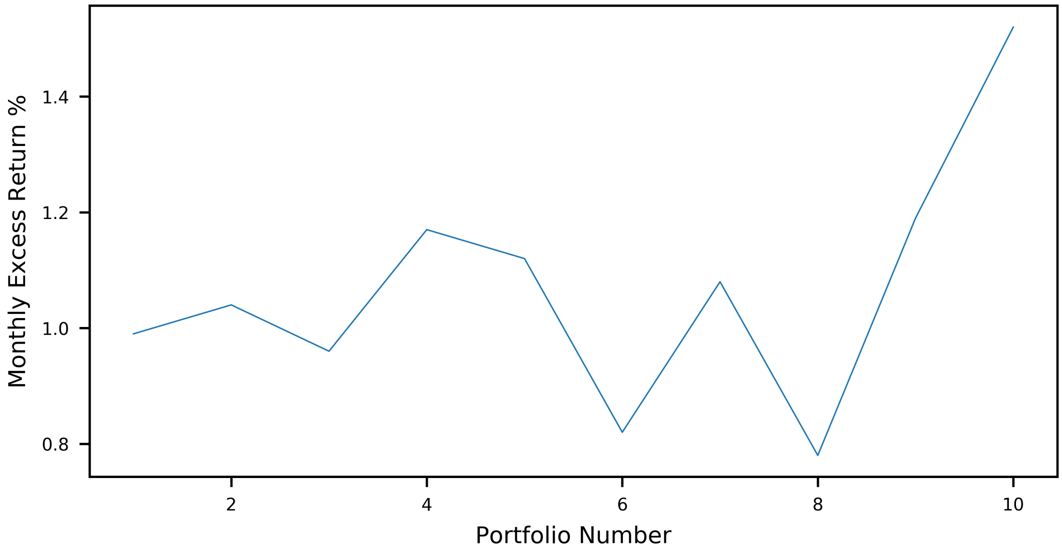


FIGURE 1.6: Portfolio Return Sorted by Absolute Mood Betas

absolute values of mood beta from equation (1.11) to populate 10 mood portfolios. Portfolio 10 contains the stocks most sensitive to public mood, regardless of whether the sensitivity is negative or positive. Portfolio 1 contains the stocks least sensitive to changes in mood. The high-low portfolio returns represent high mood stocks (portfolio 10) minus low mood stocks (portfolio 1). The top panel in Table 1.8 reports value-weighted portfolio excess returns and average regression coefficients of the mood variable for each portfolio. The high mood portfolio earns the highest excess return: about 1.52% per month with high statistical significance. Figure 1.6 shows that there is an increasing pattern of monthly excess returns from the low mood sensitivity portfolio 1 to the high mood sensitivity portfolio 10. The high-low mood portfolio generates significantly positive excess returns of 0.53% per month.

The lower panels in Table 1.8 report the factor analysis of the monthly excess returns of the ten portfolios sorted by absolute mood betas. The results are entirely comparable to those discussed above based on the raw mood betas. Each of the factor models returns large and statistically significant alphas for the most mood-sensitive portfolio and for the long-short strategy. As before, increasing the complexity of the factor model only serves to increase the mood exposure-driven alphas.

1.4.1 Mood Factor Portfolio Return

We construct the mimicking mood factor portfolio following the standard method used in empirical asset pricing studies. At the end of each month, we first use the NYSE breakpoints of market capitalization to split stocks into two size portfolios - small and big. Independently, we use the NYSE breakpoints of the absolute value of mood betas estimated from equation (1.11) to rank stocks into three mood portfolios: low 30%, middle 40% and high 30%. Stocks within the lowest 30^{th} percentile are the most insensitive to mood; stocks within the highest 30^{th} percentile are the most sensitive to mood either negatively or positively; and the stocks within the middle 40% have neutral mood sensitivity. We thus form six interacted value-weighted portfolios in respect of size and mood effect: $L/S, N/S, H/S$, $L/B, N/B, H/B$ sorting by size and the absolute value of mood betas independently. The zero-cost mood factor portfolio is constructed by taking the average of long positions in the two mood-sensitive portfolios high 30% $(H/S, H/B)$ and the average of short positions in the two mood-insensitive portfolios low 30% $(L/S, L/B)$ each month.

Panel A of Table 1.9 gives the Pearson correlations between our mood factor and Fama-French five factors, Carhart's momentum factor, and both short- and long-term reversal factors. The mood factor is positively correlated with the market, size and reversal factors, and negatively correlated with profitability and momentum. These correlations are comparable in magnitude to those between the other previously identified factors. On average, the risk premium of the mood factor is 0.56% per month and is highly statistically significant, with a t-statistic of 4.7. We run time series regressions of the mood factor on subsets of the other factors. The most general regression is of the form:

$$r_{Mood_t} = \alpha + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + \epsilon_t \tag{1.12}$$

where $f_{m,t}$ is the vector of pricing factors. Panel B of Table 1.9 reports the various regression results. Market, size, momentum and short/long-term reversal factors display consistent explanatory power across alternative specifications, but none of the factor models can fully explain the mood factor. Our mood factor consistently earns highly significant positive alphas, and again alphas increase with the complexity of the factor model. With the most basic CAPM regression, the mood factor has an alpha of 0.38% per month. Once orthogonalized to all factors, the mood factor alpha increases to 0.56% per month, exactly equal to its mean return.

In the rest of the study we test the pricing power of the mood factor orthogonalized to all the other factors.[39] We define the mood factor orthogonalized to the other factors

[39] In fact, our results are essentially unchanged if we instead use the original mood factor, but given the weak correlations with other factors we use the orthogonalized version to provide the strictest test of its explanatory power.

Table 1.9: Factor Regression Analysis on Mood Factor

At the end of each month, we use NYSE breakpoints of market capitalization to split stocks into two size portfolios - small and big. Independently, we use NYSE breakpoints of the absolute value of mood betas estimated from (1.11) to rank stocks into three mood portfolios: low 30%, middle 40%, and high 30%. Stocks within the lowest 30$^{\text{th}}$ percentile are the most insensitive to mood; stocks within the highest 30$^{\text{th}}$ percentile are the most sensitive to mood either negatively or positively, and the stocks within the middle 40% are neutral to mood-sensitivity. We thus form six interacted value-weighted portfolios respecting size and mood effect: $L/S, N/S, H/S, L/B, N/B, H/B$ sorting on the size and the absolute value of mood betas independently. The zero-cost mood factor portfolio is constructed by taking the average of long positions in in the two mood-sensitive portfolios-high 30% ($H/S, H/B$) and the average of short positions in the two mood-insensitive portfolios-low 30% ($L/S, L/B$) each month. Panel A reports Pearson correlation between the mood portfolio return factor and other factors. Panel B is the regression analysis of mood factor portfolio returns on Fama-French five factors, momentum and short- and long-term reversal factors. Robust t statistics are in parentheses.

Panel A : Pearson Correlation Matrix

	$Mood$	MKT	SMB	HML	CMA	RMW	MOM	ST	LT
$Mood$	1.00	0.47	0.44	0.01	-0.12	-0.38	-0.30	0.33	0.39
MKT		1.00	0.42	0.18	0.09	-0.38	-0.13	0.47	0.53
SMB			1.00	0.23	0.13	-0.42	-0.05	0.24	0.41
HML				1.00	0.62	-0.16	-0.34	0.18	0.62
CMA					1.00	0.07	-0.09	-0.02	0.46
RMW						1.00	0.15	-0.32	-0.49
MOM							1.00	-0.31	-0.28
ST								1.00	0.39
LT									1.00

Panel B: Factor Regression of Mood factor Portfolio Return

	$Mean$	α	MKT	SMB	HML	CMA	RMW	MOM	ST	LT	$\bar{R}^2$
r_{Mood}	0.56	0.38	0.16								0.21
	(4.70)	(3.26)	(4.67)								
		0.46	0.11	0.16	-0.02	-0.16	-0.11				0.30
		(4.97)	(3.29)	(3.00)	(-0.20)	(-1.34)	(-1.55)				
		0.50	0.10	0.17	-0.09	-0.11	-0.09	-0.13			0.37
		(6.49)	(3.38)	(3.46)	(-1.02)	(-1.00)	(-1.28)	(-2.25)			
		0.50	0.10	0.17	-0.09	-0.11	-0.09	-0.12	0.01		0.37
		(6.52)	(2.83)	(3.48)	(-1.03)	(-0.96)	(-1.25)	(-2.28)	(0.31)		
		0.56	0.07	0.17	-0.16	-0.18	-0.01	-0.12	0.00	0.19	0.40
		(7.19)	(1.60)	(4.12)	(-2.43)	(-1.57)	(-0.11)	(-2.11)	(-0.10)	(1.91)	

as the alpha from equation (1.12) plus the regression residuals:

$$r^{\perp}_{Mood_t} = \hat{\alpha} + \epsilon_t \tag{1.13}$$

We first regress our absolute mood sensitivity-based portfolio excess returns on Fama-French factors and the orthogonalized mood factor:

$$r_{p,t} - r_f = \alpha_p + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + \beta_{Mood^{\perp},p} r^{\perp}_{Mood_t} + \epsilon_{p,t} \tag{1.14}$$

where $f_{m,t}$ includes Fama-French five factors. We subsequently augment the regression with momentum and with reversal factors. The second panel in Table 1.10 reports the regression results for equation (1.14). With the orthogonalized mood factor included in the model, portfolio 10, which contains the most moody stocks, has a loading on the mood factor of 0.79, which is highly significant ($t = 8.12$), and a statistically

Table 1.10: Orthogonalized Mood Factor Pricing of Portfolios Sorted by Absolute Value of Mood Betas

For each month from 2010 to 2016, we sort stocks into 10 portfolios based on the absolute factor loading(NYSE breakpoints) of $\Delta Mood_t$ in model (1.11) and calculate the value-weighted monthly excess return for each portfolio. Standard errors are subject to Newey-West correction. H/L is the high-low portfolio which to long both mood-affected stocks (portfolios 10) and to short mood-insensitive stocks (portfolio 1). Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas across 10 portfolios based on the Fama-French five-factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{Mood^{\perp},p}r^{\perp}_{Mood,t} + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{Mood^{\perp},p}r^{\perp}_{Mood,t} + \epsilon_{p,t}$) and the full behavorial factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{ST,p}ST_t + \beta_{LT,p}LT_t + \beta_{Mood^{\perp},p}r^{\perp}_{Mood,t} + \epsilon_{p,t}$).

	1	2	3	4	5	6	7	8	9	10	H/L
$\beta_{\|Mood\|}$	0.01	0.04	0.07	0.10	0.13	0.17	0.21	0.28	0.38	0.68	
$Mean$	0.99%	1.04%	0.96%	1.17%	1.12%	0.82%	1.08%	0.78%	1.19%	1.52%	0.53%
t_{Mean}	2.95	3.22	2.79	3.54	3.06	2.29	2.71	1.65	2.44	4.02	3.16
$Fama-French$											
α	0.02	0.11	0.08	0.12	0.09	-0.14	-0.03	-0.74	-0.25	-0.08	-0.10
t_α	0.14	0.88	0.54	0.90	0.74	-0.64	-0.13	-5.39	-1.34	-0.73	-0.46
MKT	1.02	1.00	1.03	0.97	1.01	0.96	1.09	1.18	1.10	1.04	0.02
t_{MKT}	24.71	30.49	25.15	28.64	31.64	14.32	24.74	22.25	27.78	34.80	0.35
SMB	-0.16	-0.07	-0.19	-0.02	0.00	0.01	0.01	0.19	0.31	0.16	0.32
t_{SMB}	-3.00	-0.88	-1.77	-0.27	-0.01	0.14	0.16	2.49	2.38	2.85	3.99
HML	0.04	-0.08	0.02	-0.08	-0.01	0.05	0.17	0.05	0.14	-0.04	-0.08
t_{HML}	0.62	-1.38	0.21	-0.77	-0.10	0.29	1.26	0.41	1.57	-0.58	-0.98
CMA	0.03	0.10	0.13	0.14	0.10	0.15	0.02	0.13	-0.39	0.05	0.02
t_{CMA}	0.30	0.65	0.71	1.03	0.86	1.11	0.12	0.65	-2.65	0.55	0.19
RMW	0.01	0.08	0.09	0.06	0.13	0.01	0.05	-0.20	-0.20	0.03	0.02
t_{RMW}	0.14	0.78	0.75	0.68	2.06	0.04	0.43	-1.27	-1.83	0.42	0.18
$r_{Mood^\perp}$	-0.24	-0.30	-0.46	-0.05	-0.17	-0.21	-0.18	0.36	0.51	0.79	1.02
$t_{Mood^\perp}$	-2.28	-1.88	-3.33	-0.42	-1.38	-1.66	-1.25	1.26	2.66	8.12	6.69
R^2	0.92	0.91	0.88	0.89	0.93	0.82	0.87	0.88	0.89	0.94	0.42
$CARH$											
α	-0.02	0.10	0.06	0.13	0.08	-0.15	-0.02	-0.71	-0.20	-0.06	-0.04
t_α	-0.13	0.78	0.44	0.91	0.67	-0.67	-0.06	-5.71	-1.06	-0.63	-0.28
MKT	1.02	1.00	1.03	0.96	1.01	0.96	1.09	1.17	1.09	1.03	0.01
t_{MKT}	31.19	30.23	25.40	26.31	33.69	13.96	22.44	23.44	24.57	39.83	0.19
SMB	-0.17	-0.07	-0.19	-0.02	0.00	0.00	0.02	0.20	0.34	0.17	0.35
t_{SMB}	-3.61	-0.93	-1.91	-0.20	-0.08	0.04	0.24	2.49	2.91	2.91	4.27
HML	0.12	-0.06	0.05	-0.10	0.01	0.08	0.14	0.00	0.03	-0.08	-0.20
t_{HML}	1.87	-0.93	0.36	-1.04	0.22	0.40	0.92	0.00	0.44	-1.05	-2.60
CMA	-0.02	0.09	0.12	0.16	0.09	0.13	0.04	0.16	-0.32	0.07	0.10
t_{CMA}	-0.23	0.56	0.61	1.14	0.75	0.90	0.24	0.82	-2.24	0.75	0.68
RMW	-0.01	0.07	0.08	0.07	0.12	0.00	0.06	-0.19	-0.18	0.04	0.05
t_{RMW}	-0.13	0.73	0.76	0.75	2.06	-0.01	0.49	-1.20	-1.63	0.55	0.62
MOM	0.13	0.03	0.04	-0.04	0.03	0.05	-0.05	-0.08	-0.18	-0.07	-0.19
t_{MOM}	4.28	0.88	0.78	-0.53	1.40	0.57	-0.77	-1.87	-3.37	-2.57	-5.70
$r_{Mood^\perp}$	-0.24	-0.30	-0.46	-0.05	-0.17	-0.21	-0.18	0.36	0.51	0.79	1.02
$t_{Mood^\perp}$	-2.72	-1.86	-3.49	-0.39	-1.38	-1.57	-1.16	1.32	2.96	9.36	9.44
R^2	0.93	0.91	0.88	0.89	0.93	0.82	0.87	0.88	0.90	0.94	0.51
$CARH\&ST\&LT$											
α	-0.01	0.03	0.12	0.12	0.02	-0.09	0.03	-0.70	-0.15	-0.02	-0.02
t_α	-0.04	0.22	0.66	0.78	0.18	-0.43	0.13	-4.90	-0.75	-0.25	-0.12
MKT	0.99	1.03	1.00	0.99	1.04	0.90	1.05	1.12	1.03	1.03	0.04
t_{MKT}	38.51	34.02	19.66	24.78	29.48	19.21	20.49	23.60	23.49	30.40	0.86
SMB	-0.18	-0.07	-0.20	-0.01	0.00	-0.01	0.01	0.19	0.33	0.17	0.35
t_{SMB}	-4.30	-0.88	-2.19	-0.17	0.03	-0.13	0.16	2.37	2.85	3.05	5.18
HML	0.09	0.02	-0.02	-0.08	0.08	0.00	0.08	-0.03	-0.03	-0.12	-0.21
t_{HML}	1.80	0.23	-0.17	-0.85	1.58	-0.01	0.62	-0.23	-0.34	-1.60	-2.63
CMA	0.00	0.20	0.05	0.14	0.16	0.09	-0.01	0.20	-0.36	0.02	0.02
t_{CMA}	0.00	1.25	0.22	1.02	1.64	0.56	-0.05	0.88	-2.56	0.16	0.11
RMW	0.02	-0.03	0.16	0.04	0.04	0.10	0.13	-0.15	-0.10	0.08	0.06
t_{RMW}	0.31	-0.26	1.17	0.42	0.73	0.54	1.09	-0.76	-0.84	0.90	0.53
MOM	0.15	0.04	0.04	-0.05	0.03	0.07	-0.04	-0.05	-0.16	-0.07	-0.22
t_{MOM}	4.15	1.08	1.08	-0.76	1.30	0.73	-0.53	-1.14	-2.92	-2.71	-6.15
ST	0.11	0.08	0.00	-0.09	0.01	0.10	0.04	0.16	0.09	-0.06	-0.18
t_{ST}	2.16	1.54	0.03	-1.26	0.19	0.84	0.70	3.28	1.75	-1.13	-2.65
LT	0.02	-0.24	0.16	-0.02	-0.18	0.18	0.15	0.03	0.14	0.11	0.09
t_{LT}	0.40	-3.97	1.25	-0.19	-3.37	1.37	1.25	0.26	1.10	1.46	1.19
$r_{Mood^\perp}$	-0.24	-0.30	-0.46	-0.05	-0.17	-0.21	-0.18	0.36	0.51	0.79	1.02
$t_{Mood^\perp}$	-2.89	-2.05	-3.76	-0.39	-1.53	-1.61	-1.17	1.33	3.04	9.61	9.58
R^2	0.93	0.92	0.88	0.89	0.93	0.83	0.87	0.88	0.90	0.94	0.55

insignificant alpha of -0.08% per month, down from the 0.36% per month reported in Table 1.8. Similarly, the high-low portfolio alpha decreases from 0.47% to a statistically insignificant -0.10% per month. The adjusted-R^2 figure for portfolio 10 is slightly improved by adding the mood factor to the model, increasing from 0.91 to 0.96, but the improvement is most noticeable for the long-short strategy, increasing from 0.11 to 0.42. Portfolio 9 loads positively and significantly on the mood factor, while portfolio 1 loads negatively. The adjusted-R^2 for portfolio 1 is barely affected, and the mood factor, while statistically significant, does not appear to add much explanatory power but pushes the alpha back closer to zero.

The other panels in Table 1.10 report the regression results from more complete factor models, including momentum and reversal factors. Results differ only slightly from those in the second panel. In short, the mood factor is highly significant for portfolio 10 and the high-low portfolio, which earn insignificant alphas once the mood factor is included in the regression. The significant beta on the mood factor captures the additional risk from the mood effect on returns.

1.4.2 Pricing Analysis for 25 Size-Mood Portfolios

Beginning with December 2009, we use the NYSE breakpoints to split stocks into quintiles of market capitalization at the end of each month. Independently, we use the NYSE breakpoints to split stocks into quintiles based on absolute mood beta estimated from equation (1.11). We then form 25 size and mood portfolios by taking intersections.[40] Monthly value-weighted portfolio returns are calculated from the end of month t to the end of month $t+1$, and portfolios are re-balanced at the end of each month $t+1$.

Table 1.11 reports average excess returns and alphas from the Fama-French five-factor, Carhart six-factor, and full eight-factor models, including reversals. There is clear evidence that moody stocks earn high excess returns per month. From the upper left panel of Table 1.11, we see that the smallest stocks with high sensitivity to mood earn 1.71% per month. Average returns for moody stocks are reasonably constant for successively larger size quintiles. Even the largest moody stocks earn returns of 1.52% per month. In fact, within all five mood quintiles, size does not appear to be an important differentiating factor in terms of average returns. However average returns

[40]Here we perform independent sorts on size and mood beta. As noted, stocks that are sensitive to mood tend to be smaller and so the portfolios are unbalanced in terms of numbers of stocks. Each month, the portfolio of large, moody stocks contains much fewer stocks than the portfolio of small moody stocks. Nevertheless, there are sufficient stocks in even the smallest portfolio to perform the analysis. In the robustness tests we perform conditional sorts, first on size then on mood betas, in order to balance the portfolios better. Our findings are robust to this approach.

Table 1.11: 25 Size and Mood Value-Weighted Portfolio Analysis

Starting from December 2009, we use the NYSE breakpoints to split the NYSE, Amex and NASDAQ stocks into quintiles on the market equity at the end of each month. Independently, we use the NYSE breakpoints to split the NYSE, Amex and NASDAQ stocks into quintiles on absolute value of mood beta calculated from model (1.11). We form 25 size and mood portfolios by taking intersections. Monthly value-weighted portfolio returns are calculated from the end of month t to the end of month $t+1$, and portfolios are rebalanced at the end of each month $t+1$. H/L is the high-low portfolio to long mood-affected stocks (portfolio 5 ranked on mood) and to short mood-insensitive stocks (portfolio 1 ranked on mood) respecting each size portfolio. Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas across 25 portfolios based on Fama-French five-factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \epsilon_{p,t}$) and the full behavioral factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{ST,p}ST_t + \beta_{LT,p}LT_t + \epsilon_{p,t}$).

	Sober	2	3	4	Moody	High-Low
			Mean			
Small	0.77	1.46	0.87	0.94	1.71	0.93
2	0.77	1.34	1.06	1.12	1.28	0.51
3	0.63	0.91	1.10	1.25	1.37	0.75
4	0.82	1.13	1.15	0.79	1.79	0.96
Large	0.74	0.99	0.67	0.85	1.52	0.78

	Sober	2	3	4	Moody	High-Low
			t_{Mean}			
Small	1.30	2.66	1.70	1.68	2.87	4.34
2	1.74	2.64	2.22	2.14	2.28	2.30
3	1.36	1.93	2.07	2.53	2.11	2.54
4	2.07	2.41	2.66	1.75	3.30	3.94
Large	2.23	3.09	1.83	2.27	3.42	3.15

	Sober	2	3	4	Moody	High-Low
			α^{FF}			
Small	-0.41	0.20	-0.37	-0.36	0.43	0.84
2	-0.43	0.25	-0.19	-0.19	-0.16	0.27
3	-0.54	-0.32	-0.20	0.01	-0.08	0.46
4	-0.30	-0.02	0.02	-0.42	0.40	0.70
Large	-0.38	-0.13	-0.45	-0.42	0.39	0.77

	Sober	2	3	4	Moody	High-Low
			$t_{\alpha^{FF}}$			
Small	-1.77	0.82	-2.21	-2.42	2.53	3.06
2	-4.57	1.35	-1.54	-1.70	-0.95	1.48
3	-2.93	-2.12	-1.15	0.08	-0.30	1.71
4	-1.95	-0.13	0.14	-1.69	2.08	3.29
Large	-2.82	-0.98	-2.08	-2.72	2.52	3.46

	Sober	2	3	4	Moody	High-Low
			α^{CARH}			
Small	-0.40	0.27	-0.35	-0.30	0.54	0.93
2	-0.39	0.30	-0.14	-0.13	-0.06	0.33
3	-0.54	-0.30	-0.20	0.05	-0.02	0.52
4	-0.29	0.00	0.03	-0.37	0.47	0.76
Large	-0.39	-0.13	-0.47	-0.39	0.45	0.84

	Sober	2	3	4	Moody	High-Low
			$t_{\alpha^{CARH}}$			
Small	-1.75	1.22	-2.08	-2.19	2.93	3.26
2	-3.94	1.72	-1.17	-1.23	-0.37	1.88
3	-2.90	-2.03	-1.15	0.26	-0.07	2.16
4	-1.94	0.00	0.23	-1.70	2.66	3.62
Large	-3.30	-0.91	-2.02	-2.21	2.89	4.64

	Sober	2	3	4	Moody	High-Low
			$\alpha^{CARH\&ST\&LT}$			
Small	-0.37	0.31	-0.29	-0.23	0.58	0.95
2	-0.41	0.22	-0.14	-0.09	0.04	0.45
3	-0.57	-0.30	-0.22	0.07	0.11	0.67
4	-0.31	0.02	0.04	-0.34	0.61	0.92
Large	-0.41	-0.08	-0.48	-0.36	0.47	0.89

	Sober	2	3	4	Moody	High-Low
			$t_{\alpha^{CARH\&ST\&LT}}$			
Small	-1.61	1.26	-1.85	-1.68	3.20	3.40
2	-3.57	1.11	-1.16	-0.85	0.22	2.22
3	-3.12	-1.77	-1.12	0.35	0.44	2.77
4	-2.06	0.14	0.25	-1.47	4.00	5.08
Large	-3.31	-0.52	-2.10	-1.98	2.66	3.82

clearly drop as we examine successively less moody stocks within each size category. The most sober (least moody) stocks typically earn no more than half the average return of the most moody stocks. The high-low portfolios defined as taking a long position in moody stocks and a short position in sober stocks within each size quintile earn economically and statistically significant positive returns.

The lower panels in Table 1.11 report alphas from progressively more complex factor models. Considering first the estimates based on the Fama-French five-factor model, alphas are typically positive and significant for the most moody stocks, and negative and significant for less moody stocks (including, in several places, portfolios

Table 1.12: 25 Size and Mood Value-Weighted Portfolio Analysis with Mood Factor

Starting from December 2009, we use the NYSE breakpoints to split the NYSE, Amex and NASDAQ stocks into quintiles on the market equity at the end of each month. Independently, we use the NYSE breakpoints to split the NYSE, Amex and NASDAQ stocks into quintiles on absolute value of mood beta calculated from model (1.11). We form 25 size and mood portfolios by taking intersections. Monthly value-weighted portfolio returns are calculated from the end of month t to the end of month $t+1$, and portfolios are rebalanced at the end of each month $t+1$. H/L is the high-low portfolio to long mood-affected stocks (portfolio 5 ranked on mood) and to short mood-insensitive stocks (portfolio 1 ranked on mood) respecting each size portfolio. Market factor (MKT), size factor (SMB), value factor (HML), investment factor (CMA), profitability factor (RMW), momentum factor (MOM), short-term reversal factor (ST) and long-term reversal factor (LT) are downloaded from the Kenneth R. French Data Library. We report the mean monthly alphas of the orthogonalized mood factor pricing across 25 portfolios based on Fama-French five-factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{Mood^\perp,p}r^\perp_{Mood,t} + \epsilon_{p,t}$), the Carhart momentum factor regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{Mood^\perp,p}r^\perp_{Mood,t} + \epsilon_{p,t}$) and the full behavorial factors regressions ($r_{p,t} - r_f = \alpha_p + \beta_{MKT,p}MKT_t + \beta_{SMB,p}SMB_t + \beta_{HML,p}HML_t + \beta_{CMA,p}CMA_t + \beta_{RMW,p}RMW_t + \beta_{MOM,p}MOM_t + \beta_{ST,p}ST_t + \beta_{LT,p}LT_t + \beta_{Mood^\perp,p}r^\perp_{Mood,t} + \epsilon_{p,t}$).

FF & Mood

	Sober	2	3	4	Moody	High-Low
			$\alpha^{Mood^\perp}$			
Small	-0.39	0.32	-0.38	-0.23	0.23	0.62
2	-0.13	0.38	-0.05	-0.03	-0.09	0.05
3	-0.28	0.04	-0.06	0.04	-0.25	0.03
4	-0.13	0.05	0.03	-0.25	0.15	0.28
Large	-0.15	0.02	-0.31	-0.55	-0.24	-0.09

	Sober	2	3	4	Moody	High-Low
			$t^{\alpha^{Mood^\perp}}$			
Small	-1.89	1.19	-1.80	-1.30	1.40	2.62
2	-1.20	2.03	-0.28	-0.23	-0.43	0.21
3	-1.46	0.20	-0.32	0.22	-0.86	0.10
4	-0.86	0.23	0.19	-1.03	0.56	1.20
Large	-1.09	0.13	-1.68	-3.15	-1.10	-0.31

	Sober	2	3	4	Moody	High-Low
			$\beta_{Mood^\perp}$			
Small	-0.04	-0.22	0.02	-0.22	0.36	0.40
2	-0.54	-0.25	-0.26	-0.29	-0.12	0.41
3	-0.46	-0.65	-0.24	-0.04	0.31	0.78
4	-0.32	-0.12	-0.02	-0.31	0.44	0.76
Large	-0.42	-0.27	-0.26	0.23	1.14	1.55

	Sober	2	3	4	Moody	High-Low
			$t_{\beta_{Mood^\perp}}$			
Small	-0.23	-0.93	0.11	-1.06	2.04	1.71
2	-3.43	-1.63	-1.47	-1.50	-0.81	2.10
3	-3.32	-3.23	-1.61	-0.25	1.39	3.14
4	-1.73	-0.62	-0.12	-1.31	1.88	3.23
Large	-2.46	-2.45	-1.60	1.40	4.49	5.62

CARH & Mood

	Sober	2	3	4	Moody	High-Low
			$\alpha^{Mood^\perp}$			
Small	-0.38	0.39	-0.36	-0.18	0.33	0.71
2	-0.09	0.44	0.00	0.03	0.01	0.10
3	-0.28	0.06	-0.06	0.07	-0.19	0.09
4	-0.11	0.07	0.05	-0.20	0.23	0.34
Large	-0.16	0.02	-0.32	-0.52	-0.18	-0.02

	Sober	2	3	4	Moody	High-Low
			$t_{\alpha^{Mood^\perp}}$			
Small	-1.83	1.72	-1.72	-1.24	1.73	2.73
2	-0.92	2.47	0.03	0.27	0.07	0.48
3	-1.45	0.31	-0.32	0.37	-0.78	0.36
4	-0.83	0.37	0.25	-0.93	0.99	1.64
Large	-1.34	0.14	-1.63	-2.60	-1.02	-0.09

	Sober	2	3	4	Moody	High-Low
			$\beta_{Mood^\perp}$			
Small	-0.04	-0.22	0.02	-0.22	0.36	0.40
2.00	-0.54	-0.25	-0.26	-0.29	-0.12	0.41
3.00	-0.46	-0.65	-0.24	-0.04	0.31	0.78
4.00	-0.32	-0.12	-0.02	-0.31	0.44	0.76
Large	-0.42	-0.27	-0.26	0.23	1.14	1.55

	Sober	2	3	4	Moody	High-Low
			$t_{\beta_{Mood^\perp}}$			
Small	-0.22	-1.03	0.11	-1.26	2.66	2.57
2.00	-4.04	-1.39	-2.13	-1.97	-0.83	1.89
3.00	-3.32	-3.25	-1.60	-0.26	1.27	3.05
4.00	-1.72	-0.63	-0.11	-1.08	2.32	3.92
Large	-2.46	-2.40	-1.54	1.49	5.88	7.50

CARH&ST&LT&Mood

	Sober	2	3	4	Moody	High-Low
			$\alpha^{Mood^\perp}$			
Small	-0.34	0.44	-0.30	-0.11	0.38	0.72
2.00	-0.11	0.35	0.01	0.07	0.10	0.22
3.00	-0.31	0.06	-0.08	0.09	-0.07	0.24
4.00	-0.14	0.09	0.05	-0.17	0.36	0.50
Large	-0.18	0.07	-0.34	-0.49	-0.16	0.03

	Sober	2	3	4	Moody	High-Low
			$t_{\alpha^{Mood^\perp}}$			
Small	-1.69	1.70	-1.45	-0.73	2.00	2.82
2.00	-1.15	1.68	0.04	0.59	0.55	1.01
3.00	-1.59	0.31	-0.39	0.45	-0.31	1.09
4.00	-0.92	0.51	0.26	-0.73	1.68	2.78
Large	-1.47	0.45	-1.72	-2.37	-0.81	0.11

	Sober	2	3	4	Moody	High-Low
			$\beta_{Mood^\perp}$			
Small	-0.04	-0.22	0.02	-0.22	0.36	0.40
2.00	-0.54	-0.25	-0.26	-0.29	-0.12	0.41
3.00	-0.46	-0.65	-0.24	-0.04	0.31	0.78
4.00	-0.32	-0.12	-0.02	-0.31	0.44	0.76
Large	-0.42	-0.27	-0.26	0.23	1.14	1.55

	Sober	2	3	4	Moody	High-Low
			$t_{\beta_{Mood^\perp}}$			
Small	-0.23	-1.03	0.10	-1.24	2.71	2.46
2.00	-4.22	-1.35	-2.22	-1.96	-0.87	2.12
3.00	-3.19	-3.30	-1.57	-0.27	1.54	3.60
4.00	-1.69	-0.64	-0.11	-1.11	2.25	3.93
Large	-2.54	-2.76	-1.58	1.42	6.03	7.46

comprising stocks with intermediate levels of mood sensitivity). The high-low strategy alphas are always positive, often significantly so. Again, consistent with earlier results, both the magnitude and statistical significance of alphas increase as the factor models become more complex. In the bottom panel reporting the results of the Fama-French five-factor model augmented with momentum and short- and long-term reversal factors, high-low strategy alphas are very similar to the average excess returns (even though individual portfolio alphas are very different). It is evident that conventional factor models cannot adequately price size-mood beta-sorted portfolios.

We now test the pricing power of our mood factor constructed in the section above on the same 25 portfolios. The top panel of Table 1.12 reports the key results from pricing these portfolios using the Fama-French five-factor model augmented with the orthogonalized mood factor. The first block reports alphas. In most cases, the alphas are much smaller than reported in Table 1.11, and statistical significance is lost. The main exceptions are for the smallest quintile of stocks where the most sober stocks report a marginally significant and negative alpha which, when paired with the positive but insignificant alpha from the most moody stocks gives, a positive long-short alpha. This is one-third smaller than the equivalent alpha reported in Table 1.11, but it remains statistically significant. The second block reports the loadings on the mood factor in the regression. The pattern of these loadings is as expected: positive and significant for the most moody stocks and for the long-short strategy, negative and significant for the less moody portfolios. Subsequent panels report results for more complex factor models but the inferences are quite similar to those from the Fama-French model. Loadings on the mood factor are often very significant and follow the expected patterns in both sign and magnitudes, while the magnitudes of alphas for the 25 portfolios are typically so reduced as to lose statistical significance. High-low portfolios, however, still offer positive alphas that are sometimes statistically significant.

We conclude that the orthogonalized mood factor has important additional pricing power beyond that offered by the benchmark Fama-French five-factor model, even when augmented with previously identified behavioral factors. The mood factor does a reasonable but not perfect job of explaining the "mispricing" caused by incorporating mood as pricing information rather than acquiring fundamental information to learn about assets across these 25 portfolios. In subsequent drafts, we will test its ability to explain returns on portfolios sorted on the basis of other known anomalies.

1.4.3 Zero Mood Beta Test

The high excess monthly returns on mood-sensitive portfolios (portfolio 1 and portfolio 10) indicate that moody stocks earn higher expected returns and are consistent with the theoretical result in section 1.2. In fact, the excess returns of moody stocks are due to the irrational information acquisition decision induced by mood contributes an additional risk that can not be captured by traditional pricing factors or the unconditional betas. Naively, looking at the sign of the sensitivity of mood risk, the mood beta in equation (1.11), it could be questioned whether the mood risk can be canceled out by holding both negative and positive sensitive mood stocks. Therefore, in this section we test whether the exposure to mood risk is hedged by taking a long position in both negative mood portfolio (P1) and positive mood portfolio (P10).

To test whether this mood risk hedging strategy is feasible, we separate the mood beta for each stock into upward mood betas and downward mood betas. The upward mood beta measures the exposure of stocks to increases in the public mood, while the downward mood beta measures the exposure to mood when public mood decreases. For each stock we conduct the following regression:

$$r_{i,t} = \alpha_i + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + \beta_{Mood^+,i} D * \Delta Mood_t^+ + \beta_{Mood^-,i}(1-D) * \Delta Mood_t^- + \epsilon_{i,t} \quad (1.15)$$

where $f_{m,t}$ includes Fama-French five factors and the momentum factor. β_{Mood^+} captures the mood exposure for each stock using days when the public mood improves and β_{Mood^-} captures the mood exposure days of worsening mood. D is a dummy variable to identify whether a day has an upward mood change. Panel A in Table 1.13 gives the time series average of cross-sectional upward and downward mood betas. Portfolio 10 has an upward mood beta of 0.65 and a downward mood beta of 0.59. Returns on this portfolio increase more on days when mood improves than they fall on days when mood worsens by an equivalent amount. Conversely, portfolio 1 has a -0.52 beta to upward mood and a -0.64 beta to downward mood. Its returns on bad mood days are larger than its losses on equivalent good mood days. There is clear (and statistically significant) asymmetry between upward and downward mood betas in the most moody stock portfolios. It is noticeable that such asymmetry is only found for the two extreme portfolios (P1 and P10). Less moody portfolios have symmetric mood betas.

To assess the performance of the hedging strategy further, we analyze the upward and downward mood betas of the portfolio formed by combining portfolios 1 and 10. Panel B in Table 1.13 gives the time series averages of each beta. The downward mood beta is -0.035, more than 3 standard errors below zero. The average upward

Table 1.13: Zero Mood Beta Test

Panel A: The upward mood beta and downward mood beta are calculated based on the regression model : $r_{i,t} = \alpha_i + \beta_{MKT,i}MKT_t + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{CMA,i}CMA_t + \beta_{RMW,i}RMW_t\beta_{MOM,i}MOM_t + \beta_{Mood^+,i}D * \Delta Mood_t^+ + \beta_{Mood^-,i}(1-D) * \Delta Mood_t^- + \epsilon_{i,t}$. D is a dummy variable to identify if a day has an upward mood change. We conduct time series regression for each stock to isolate risk exposure to upward and downward mood change. The upward mood beta for negative mood sensitive stocks is the sensitivity of portfolio 1 return decreased on the days with upward mood change. The downward mood beta for positive mood sensitive stocks is the sensitivity of portfolio 10 return decreased on the days with downward mood change. Panel B: For each regression period, we hold stocks from portfolio 1 and 10 and calculate the cross-sectional mean of upward and downward mood betas. The upward mood hedging is to add upward mood beta from portfolio 1 and portfolio 10 at each period. The downward mood hedging is to add downward mood beta from portfolio 1 and 10.

Panel A : Upside and Downside Mood Beta

Portfolio	β^{Mood^+}	β^{Mood^-}
1	-0.52	-0.64
2	-0.23	0.24
3	-0.14	-0.14
4	-0.08	-0.07
5	-0.03	-0.02
6	0.02	0.04
7	0.08	0.10
8	0.16	0.16
9	0.26	0.26
10	0.65	0.59

Panel B: Mood Beta Hedging Test

Mood Hedging	Upside Mood	Downside Mood
$Coef$	0.061	-0.035
$t_{statistics}$	6.42	-3.01

mood beta is 0.061, with an even larger t-statistic. These findings imply that taking long positions in both positive and negative beta stocks does not eliminate exposure to mood. On good mood days, the gains from the positive mood beta stocks outweigh the losses on negative mood beta stocks - with the opposite holding true on bad mood days - because mood betas are asymmetric. Positive (negative) mood beta stocks gain more on good (bad) mood days than they lose on bad (good) mood days.

In summary, combining positive and negative mood beta stocks does not hedge mood risk exposure. Taking a long position on portfolio 1 and portfolio 10 will always involve an exposure to mood risk on either upward mood days or downward mood days. As argued in the literature, mood can irrationalize investors with biased decision making and asset valuation through different channels such as deflecting risk version, tolerance, biasing common pricing factors etc. We contribute to the literature by exploring the mood effect through the biasing channel on investors' information acquisition in respect of assets. More importantly, the additional risk originating from the biased information acquisition triggered by mood implies a higher return as required by investors to hold the mood-sensitive assets. In the next section, we conduct a detailed discussion about the theoretical motivation in our study through our interpretation of how mood gives rise to bias and becomes an indispensable risk factor that should be recognized and compensated for by investors.

1.5 Discussion

The empirical results support our key argument that stocks that are sensitive to mood as a bias factor in investors' decision to acquire information about assets' payoff and valuation earn higher expected returns. These can be thought of as risk premia to compensate investors who want to hold these moody stocks. In fact, our study is closely connected to studies that discuss theoretical perspectives of how mood creates bias in financial markets.

On the one hand, some studies argue that there is a negative relationship between people's risk aversion and mood (Kamstra et al., 2003; Kramer and Weber, 2012; Bassi et al., 2013; Kaplanski et al., 2015). More specifically, investors in a positive mood tend to be less risk-averse or more risk-tolerant and vice versa. In fact, the effect of risk aversion deflected by mood causes mispricing into two ways: either through trading behavior from misvaluation of asset payoff or through incorrect perception of the stochastic discount factor.

First, positive mood causes investors to feel less risk-averse and more likely to perceive lower risk in stocks, or to believe that stocks are more likely to be underpriced (Goetzmann et al., 2015). Therefore, investors choose either to invest in more risky assets or to conduct more buying rather than selling as they are not consciously aware of their positive mood bias. Negative mood makes investors feel more risk-averse and more likely to perceive stocks as overpriced. As stated by Nagel (2005), the short sale constraint acts as an indispensable condition in the market, especially for investors who are more likely to be biased by mood factors (Goetzmann et al., 2015). As a consequence, investors in a pessimistic mood choose stocks more carefully, with high expected returns as rewards for taking risk (Raghunathan and Pham, 1999). Eventually, as found by Goetzmann et al. (2015), stock returns that are liable to be affected by mood generate comovement during positive and negative mood days respectively. This is consistent with the evidence found in our mood beta estimation, in which mood-sensitive stocks have positive and negative mood betas. In other words, positive moody stock returns comove with positive mood days; by the same token, negative moody stock returns move together with negative mood days.

Second, the lower risk aversion from the positive mood effect could also bias investors to increase the stochastic discount factor (SDF) (Shu, 2010). As investors use this subjective discount factor to price assets, they anticipate a lower expected return as lower risk aversion implies higher risk tolerance. Therefore, they overprice the stocks and induce more buying instead of selling. Of course, higher risk aversion from a negative mood effect decreases the SDF. According to the short sale constraint, moody investors seek to invest in stocks with higher expected returns as risk compensation, as investors over-perceive the risk on negative mood days.

Overall, stocks invested in by moody investors are risky on average, as shown in our findings on financial characteristics (small in size, high idiosyncratic risk etc.) for these stocks subject to mood risk. More importantly, mood as "feeling" information is added to investors' valuation of risky assets, which contributes to risks in addition to the risk factors contained in fundamental information. Therefore, in our study, we seek to answer the question of how much the risk induced by the mood through investors' insufficient information acquisition in respect of risky assets should be compensated for with high expected returns as a risk premium by introducing the existence of mood risk in a proportion of stocks.

On the the other hand, the recent seminal study by Hirshleifer et al. (2020) proposes a multi-factor asset pricing model. They argue that the hard-to-value factor is liable to be biased by investors' mood in subsequent trading periods. In the mean time, the aggregate market factor is biased by the public mood as well. They too use the term 'mood beta', although in their application this is constructed by regressing a stock's returns on equal-weighted market returns during periods conjectured to be associated with investor mood swings. Stocks with a higher mood beta earn higher (lower) returns during future positive (negative) mood seasons. These results are consistent with several recent papers identifying seasonality in a cross-section of security returns (Heston and Sadka, 2008; Keloharju et al., 2016; Birru, 2018). The focus of Hirshleifer, Jiang, and DiGiovanni's (2020) paper is on seasonality and as such they naturally use seasonal patterns in mood to identify mood sensitivities.

In contrast, Bali et al. (2017) measure stocks' sensitivity to economic uncertainty by running time-series regressions of a multi-factor pricing model including an uncertainty index as another proposed factor. In line with this, we directly estimate our mood betas as sensitivities to an exogenous measure of mood in the factor pricing model. For this reason, our mood betas are probably very different to those of Hirshleifer et al. (2020), who argue that a particular factor in asset payoff is liable to be biased by mood. In particular, while their analysis draws on the conjecture that investors are in a happier mood in January and March compared with the supposedly sadder months of September and October, we note that our Twitter mood index of happiness follows a low-frequency cycle. This cycle is not the business cycle (Twitter mood peaked in early 2010 and late 2015, and hit troughs in late 2012 and mid-2017). It also means that once the effect of regular holidays is excluded there is very little monthly seasonality in Twitter mood.

The key argument in this paper is inspired by studies about information demand or learning (Grossman and Stiglitz, 1980; Van Nieuwerburgh and Veldkamp, 2009; Veldkamp, 2011). We incorporate human affective states as a behavioral factor serving to challenge the classical assumption of *Homo economicus* in neoclasscial economics. One may find that connections or overlapping theories between information processing and demand are irrationalized by incorporating psychological factors. In fact, the biased information demand caused by mood can be viewed as a different dimension to explain the argument of sub-optimal decision-making in economics. First, as moody investors acquire less fundamental information, their investment decisions are more likely to be subject to heuristic thinking. Therefore, investors could be either irrationally processing information as argued by Daniel et al. (1998) and Holden and Subrahmanyam (2002), or adding the behavioral factor as useful information to bias the valuation of assets (Hirshleifer and Shumway, 2003; Hirshleifer et al., 2020). Second, as investors irrationally process information, they unconsciously incorporate insufficient fundamental information, as the rational theory assumed. Eventually, their sub-optimal decisions caused by irrational information processing imply that they may not acquire or learn enough fundamental information as they should under rational expectations. In sum, the insufficient information learning caused by the psychological bias (mood), to some extent, interacts with the irrational information processing addressed in the existing literature. Essentially, we propose a new understanding of economic agents' irrational behaviors or decisions by incorporating another dimension of information economics.

Additionally, affective states, especially for negative mood, are multi-dimensional with different impacts on decision-making (Wyer Jr and Srull, 2014). For example, anxiety is one of the dimensions in negative or sad mood. People with anxious feelings may perceive more uncertainty and pay more attention to available information (Smith and Ellsworth, 1985; Ortony et al., 1990). In line with information acquisition theory, there is a positive relation between uncertainty and the decision to acquire information (Veldkamp, 2006; Benamar et al., 2019; Andrei et al., 2020). The link between negative mood and insufficient information learning in this study might appear obscure. However, based on the universe of data we use,[41] the negative mood measured in Twitter is more subject to depression[42] that contributes to the major issue of attenuate focus or attention; as a result, using less of the available information in psychological studies (Dobson and Dobson, 1981; Silberman et al., 1983; Conway and Giannopoulos, 1993). Indeed, disentangling the different dimensions of negative affective states and investigating the non-monotonic effect in decision-making will further enrich classical economic analysis. We leave this opportunity for future research.

It might be tempting to think that the factor biased by investor mood in the study by Hirshleifer et al. (2020) takes a view similar to our proposal regarding mood sensitivity. However, we conduct our study in a more parsimonious or generalized setting and do not specify how the bias contributes specific parameters (risk aversion, time preference or factors etc.) in asset pricing models. The bias might be affecting any of the key factors in a pricing model, or the mood itself might serve as a factor mistakenly incorporated by investors into pricing models. In actual fact, those particular biasing channels to investors' mispricing can be summarized by the the problem of information incorporation in asset valuation. If investors' decisions on the acquisition of fundamental information about assets are not as rational expected under *Homo economicus*, investors eventually incorporate insufficient fundamental information when they price assets. Therefore, we take the behavioral finance perspective to argue that the mood swings can result in insufficient acquisition of information that should be incorporated in valuation. Finally, this comes back to our key premise, which is to find which stocks are the "volunteers" to the mood risk regardless of how stocks or investors are affected by mood as its existence of bias on particular factors in decision-making and asset pricing has been broadly addressed. More importantly, our interest is to elucidate whether the mood is priced as a risk premium which investors need to consider in holding the stocks which are liable to be biased by the mood effect.

In sum, we propose that not all stocks in financial markets are liable to be biased by investor mood; that there is a subset of stocks which are traded by investors who are more likely to be affected by mood with consequent insufficient information acquisition; that regardless of the specific learning factors through which mood bias causes mispricing, the mood effect acts as a risk factor in addition to fundamental factors in

[41] An article posted in *Times* discusses researchers find tweets or messages posted by Twitter users can serve to predict depression. See the article *How Twitter Knows When You're Depressed* for more details.

[42] Quercia et al. (2011) and Park et al. (2012) find evidence that words of language used in Twitter messages are efficient to detect users' negative emotions or sentiment, especially for depressive symptoms.

asset pricing models; and that stocks which are subject to mood risk earn a higher expected return (risk premium) as a compensation for investors to hold them.

1.6 Robustness Tests

Our analysis entailed several conscious choices. In this section we test whether our results depend on any of them. Specifically, we replicate our results after making the following changes:

1. Break points:

In previous sections, we allocated stocks to portfolios according to break points determined by NYSE stocks. We repeat our analysis using break points determined by poling stocks from all three main venues-NYSE, Nasdaq and Amex. Our results are not affected by this change in any material way.

2. Portfolio weights:

We used value weighs to construct portfolios. We repeat our analysis using equal weights. Naturally, this does change the magnitude of some of our estimates, but less than expected. We suspect that this is because sorting by mood sensitivities is positively correlated with sorting by size. The decile of most positively mood-sensitive stocks are all relatively small, as are the stocks in the decile of most negatively mood-sensitive stocks. The large stocks typically - but not always - fall into the mood-insensitive deciles. Equal weighting within each decile is then not very different from value weighting. Since value weighting is the norm in this field, we report results based on this approach in the study.

3. Factor models:

Earlier drafts of this study used the Fama-French three-factor model as the starting point of all analyses. Mood betas were calculated after conditioning on these factors in equation (1.11), and subsequent analysis began with the Fama-French three-factor model before augmenting this with momentum and reversal factors. Over the last two decades, there have been studies that clearly point out the inability of the Fama-French three-factor model to explain a range of anomalies in cross-sectional stock returns.[43] For this reason, we adopt the Fama and French (2015) five-factor asset pricing model, which adds investment and profitability factors to the original three

[43]The studies of, for instance, Ikenberry et al. (1995), Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), Chan et al. (1996), Dichev (1998), Griffin and Lemmon (2002), Ang et al. (2006), Daniel and Titman (2006), Campbell et al. (2008) and Hou et al. (2015) all question the performance of the Fama-French model.

Fama-French factors. We could equally have taken the q-factor model of Hou et al. (2015) as our starting point.

Overall, our findings are all robust to the use of the Fama and French three-factor (1992), the five-factor (2015) or the q-factor models as a starting point. We obtain very similar results if we compute mood betas conditioned on a simple CAPM or Fama-French three-factor model (or, indeed, allow for no conditioning in equation (1.11)), even if we apply more complex models in subsequent stages of the analysis.

The traditional fundamental financial factors—market, size, value, profitability and investment (however defined)—cannot capture the effects of sensitivity to mood, even when augmented by behavioral factors such as Carhart's (1997) momentum and reversal factors.

4. Mood factor construction:

We orthogonalized the mood factor to our most general factor model by taking only the unexplained components of equation (1.12). This assigns all explanatory power common to mood and another factor to that other factor, thereby giving the mood factor the least possible credit for any explanatory power it might have. Nevertheless, we find that it has considerable explanatory power over and above the previously identified factors. Repeating our analysis based on the raw mood factor changes our results in the expected way, increasing its power slightly, but it does not affect inference.

In summary, we have tried to follow the standard empirical path in testing for a new factor. Deviations from this path would not have materially affected our conclusions, and our results are robust.

1.7 Conclusions

The neoclassical finance paradigm answers questions about financial markets by applying models in which the economic agents are rational. However, it is becoming increasingly apparent that this framework struggles to elucidate the essential facts with respect to the aggregate stock market, the cross-section of average returns and individuals' trading behavior (Barberis and Thaler, 2003). The effects of mood on the stock market have been addressed in the literature by attempting to connect either investors' shifting risk tolerance or directly biased pricing factors. As a consequence, mood irrationalizes investors' decision-making and trading behaviors. In line with existing studies on the mood effect in behavioral finance, we explore an innovative argument that mood can bias investors' decision to acquire information pertaining to an asset. As investors' mood swings, they tend to acquire less earnings-related information to learn about companies' performance.

Based on the empirical evidence in the data showing that mood significantly affects investors' information acquisition, we study this effect by understanding the asset risk and discussing the failures of classical pricing models with investors' inadequate learning about assets induced by mood. We test theoretical predictions implied by the mood effect on a cross-section of U.S. equity returns using a high-frequency Twitter-based mood index and applying a traditional asset-pricing approach. First, we sort stocks into portfolios on the basis of their sensitivity to changes in Twitter mood. Stocks that are either highly negatively or positively sensitive to changes in Twitter mood earn higher monthly excess returns than stocks with low or no mood sensitivity. Our empirical results offer strong evidence that risk created by agents who are more likely to be affected by psychological feelings such as mood, particularly on information acquisition decisions, earns greater expected returns, and that mood risk is priced. Second, we examine the financial characteristics of mood-sensitive stocks. In particular, stocks that are small in size, relatively young, pay less in dividends, are non-profitable, engage more in external financing, and have higher levels of idiosyncratic risk are more sensitive to changes in public mood.

To identify the quantity of risk affected by public mood, we construct mimicking portfolios by taking long positions in mood-sensitive stocks and short positions in mood-insensitive stocks. The mood risk factor earns an average return of 0.56% per month, which is not captured by traditional asset-pricing models such as the Fama-French five-factor model even if augmented with more behavioral factors such as Carhart's momentum factor. When we include our mood factor in the pricing regressions, the alphas of mood sensitivity-sorted portfolios are significantly reduced, usually to levels of insignificance. As we document in this study, in addition to fundamental risks, the mood effect adds more risk to the valuation of assets through its inducement of investors' insufficient information acquisition. Investors require risk premia as compensations to hold stocks which are more likely to be affected by mood through bias channels.

Chapter 2

Biased News and Irrational Investors: Evidence from Biased Beliefs about Uncertainty and Information Acquisition

2.1 Introduction

A theory of what drives investors' decision-making on acquiring information is explored in models of how rational investors perceive the uncertainty surrounding risky assets (Grossman and Stiglitz, 1980; Veldkamp, 2006; Andrei et al., 2019; Benamar et al., 2019). By contrast, in studies adopting the behavioral perspective, researchers customarily assume that investors suffer from psychological bias such as overconfidence, which causes the equilibria achieved in the information acquisition model to diverge from rational expectations equilibria (Odean, 1998; García et al., 2007; Ko and Huang, 2007). One argument in the study by Tirole (2002) is that rationalists have legitimate concerns about the shortcomings of the *Homo economicus* paradigm, and that the field of neoclassical economics study can be enriched by contributions from behavioral studies without losing the rigor of quantitative economics analysis.

Adopting insights from behavioral studies, an interesting question arises in situations in which the perception of uncertainty in the risky asset's payoff is not rationally formed. A question of similar interest relates to the cause of the irrationality that drives investors' biased decision-making on information acquisition. Therefore, in line with the inspiration of Tirole's (2002) study, I seek to answer the question of how an irrational decision concerning the acquisition of further information can be made by investors by drawing on behavioral views to investigate the drivers of irrationality.

The traditional view of investors' irrationality originating from psychological bias fails to adequately address how biased information transmission contributes to irrational decision-making by investors. Specifically, linguistic or rhetorical tone measured by sentiment, as a partial order on reporting strategy in publicly available news stories through newswires or online media, may bias investors toward irrational decisions concerning whether or not to acquire private information in investment. This study addresses this gap by examining how, by using biased public information about the market or companies as measured by sentiment from news stories, investors' acquisition of private, firm-specific information deviates from the rational expectations equilibrium.

Building on the model by Grossman and Stiglitz (1980), I develop a three-period model by extension from the seminal study by Andrei et al. (2019), who argue that investors' rational perception of economic uncertainty affects their attention to firm-specific information. I introduce an additional medium to relax the assumption of rationality in the model, namely, the consideration of biased public information from news to which investors are exposed exogenously before they begin to trade. Although rational agents are found to be subject to biased information in the media for decision-making (Baron, 2006; Kamenica and Gentzkow, 2011), to simplify the analysis, I adhere to Hirshleifer's (2020) study and add a parsimonious friction-naiveté assumption in the model. As stated in Hirshleifer (2020), information receivers' naiveté about bias in the messaging is due to people's general tendency to take the information at face value, rather than adjusting for the features of the data-generating process. Therefore, investors are naïve about bias in the news when considering their investment choices; as a result, their acquisition of firm-specific information will deviate from the equilibrium in rational expectations.

The key difference in the model I develop in this study compared to existing studies on biased information acquisition is that irrationality arises from the bias in the news information, rather than from investors' behavioral irrationality as the sole cause. The investors' biased acquisition decision about firm-specific information is made through the channel of their beliefs about the uncertainties in the risky asset payoff, which are biased by the public information from news articles that tend to be either optimistically or pessimistically reported.[1] When there is a positive (upward biased) tone in

[1] In section 2.4.1, I first verify this channel of irrationality as motivation from empirical evidence, arguing that the tone in the news biases the variance of distribution rather than the mean in the risky asset payoff components.

the news that investors read, they feel more optimistic or less uncertain about economic conditions or a firm's individual performance surrounding the risky investment. Accordingly, investors are biased towards an under-perception of the systematic uncertainty or idiosyncratic uncertainty in the payoff of a risky asset, which causes investors either to overstate the informativeness of price or understate the value of firm-specific information respectively. In a biased belief equilibrium, investors eventually acquire less firm-specific information than they would if the decision were made under rational expectations. By the same token, when the news is marked by a negative tone (downward biased), it leads investors to acquire more firm-specific information, due to them feeling more uncertain about the economy or the firm itself. This more uncertain perception leads investors to understate the informativeness of price or overstate the value of firm-specific information.

The model yields three testable predictions. First, since investors' perception of uncertainties in risky assets is inversely related to the tone in the news media, news sentiment, as a proxy for biased public information in the model, negatively predicts investors' acquisition of firm-specific information. Second, the deviation of firm-specific information acquisition, especially from firm-specific news sentiment, indicates a different degree of price informativeness and hence a deviation of risky assets' information asymmetry risk from the rational expectations equilibrium. As proposed by O'Hara (2003), investors require a risk premium to hold the risky assets which are subject to high information risk; thus, the compensation of the information risk in this model varies with the biased decision to acquire firm-specific information. This bias is caused by sentiment in firm-specific news. Third, firm-specific news sentiment predicts positive cross-sectional variation of stock returns in the form of variation in information risk, led by a shift from the rational expectations equilibrium of the proportion of informed investors.

To test these predictions, I use a novel dataset from Thomson Reuters MarketPsych (TRMI). To collect this dataset, Thomson Reuters develops an algorithm to conduct textual analysis of worldwide news and online media sources to provide a sentiment index. This takes the form of linguistic tone measured by counting the usage of positive and negative words in the news stories about the aggregate market or individual firms. Therefore, I use TRMI news sentiment indices as measures of biased tone in the news to test its impact on information acquisition behavior. I find strong evidence of an inverse relationship between news sentiment and uncertainties. On the one hand, it is clear that stock market news sentiment is significantly and negatively correlated with customary measures of systematic uncertainties, such as the stock market expectation of volatility on S&P500 index options (VIX) or the Economic Policy Uncertainty (EPU) indices (Baker et al., 2016). On the other hand, by using a bundle of proxies for firm-specific uncertainties – such as the variance of regression residuals from an AR(1) process of firm earnings per share (Griffin, 1977), the absolute value of unexpected earnings (Hirshleifer et al., 2008) and idiosyncratic volatility shock (Bali et al., 2018) – firm-specific news sentiment is found to be consistent in negatively predicting all proxies of firm-specific uncertainties.

Next, I examine how the news sentiment indices affect investors' decision to acquire firm-specific information. First, a proxy for firm-specific information acquisition, in line with the study by Weller (2018), is measured by earnings-related information incorporated into price before announcements. Second, I show the empirical evidence to confirm the model's theoretical implication of an inverse relationship between news sentiment and investors' acquisition of firm-specific information. In fact, when a more optimistic tone is found in the news about either the stock market or a particular firm, investors tend to acquire less earnings-relevant information before it is released and vice versa. These results hold after controlling for fixed effects, firms' fundamental variables and benchmark uncertainty measures, namely, the VIX and EPU. Overall, these findings confirm my theoretical results that the biased public information contained in news shifts investors' acquisition of firm-specific information away from the rational expectations equilibrium. I also show the predictability of the effect of sentiment in firm-specific news on cross-sectional variation of stock returns by proposing an argument that information risk in risky assets varies with firm-specific news sentiment. Specifically, I conduct daily cross-sectional Fama-Macbeth (1973) regressions to show that firm-specific news sentiment positively predicts future stock returns without reversal. These empirical results hold after including firm-fundamental control variables, volume–return predictors, and other influential effects from news variables such as value-relevant information (Tetlock et al., 2008; Chen et al., 2014) and reduction of information asymmetry (Tetlock, 2010). These findings are consistent with theoretical results. Sentiment in the firm-specific news drives a biased belief equilibrium in investors' firm-specific information acquisition which deviates from rational expectations; the information risk in the risky assets eventually becomes relatively higher or lower to traders through a price discovery process.

As an additional test of the risk premium argument, I conduct a factor pricing test by constructing a zero-cost portfolio sorted by daily cross-sectional firm-specific news sentiment. On average, the news sentiment factor earns around a 6.6-basis point return per day, which is equal to annualized return of about 16.63%. In addition, controlling for classical asset pricing factors such as the Fama–French five factors (Fama and French, 2015), the momentum factor (Carhart, 1997), the Pastor and Stambaugh liquidity factor (Pástor and Stambaugh, 2003), and short- and long-term reversal factors does not accommodate for abnormal returns as fully as the news-sentiment portfolio does. In sum, the factor pricing results support this study's theoretical proposition that sentiment, particularly from firm-specific news, affects information risk in risky assets, in that the proportion of informed investors, in a biased belief equilibrium, departs from rational expectations.

My study makes a unique contribution to the literature on information acquisition by investors. Through both neoclassical and behavioral economics perspectives, prior studies have addressed how investors' perceptions of uncertainty or the value of signals create demand for information about assets' fundamental payoff (Grossman and Stiglitz, 1980; Veldkamp, 2006; Odean, 1998; García et al., 2007; Ko and Huang, 2007; Andrei et al., 2019; Benamar et al., 2019). In line with the behavioral school's tendency to relax strict rationality in economic studies, this research is enriched by the introduction of a new biased channel that is motivated by Hirshleifer's (2020) seminal study, which argues that biased information or signals stemming from information transmission significantly affect investors' decision-making and may cause asset mispricing. Therefore, in contrast to the majority of extant behavioral studies in finance and economics that examine the behaviors of irrational agents, this study focuses on biased information percolation as argued for by Hirshleifer (2020) and proposes that investors should not necessary be presumed to be irrational agents. Investors can, in fact, be 'forced' into behaving sub-optimally when they receive and apply biased public information from news in their decision-making on acquisition of firm-specific information.

My study also contributes to the growing body of research that makes use of textual data in finance and economics. This literature includes studies by Tetlock (2007), Akhtar et al. (2011) and Garcia (2013) on negative news sentiment regarding aggregate markets predicting market returns; studies by Tetlock et al. (2008), Tetlock (2010), Chen et al. (2014) and Ke et al. (2019) on firm-specific news or online media sentiment containing valuable information for predicting positive future returns; and studies on the effect of media on stock markets by Bhattacharya et al. (2009), Engelberg and Parsons (2011), Peress (2014), Hillert et al. (2014), and Bonsall IV et al. (2020). However, news sentiment plays a key role in my study in demonstrating investors' biased decision-making on firm-specific information acquisition, which has not been addressed in the literature. Additionally, contrary to the argument that value-relevant information may be found in the news, the empirical result that sentiment in firm-specific news predicts positive future stock returns supports the theoretical prediction that information risk varies with firm-specific news sentiment.

Furthermore, my study sheds light on other studies that address how information purveyors such as journalists or media companies reflect different tones in news or media which bias or slant audiences' economic or political opinions (Mullainathan and Shleifer, 2005; Baron, 2006; Gentzkow et al., 2015). More importantly, media bias can be persistent as information in the news is suppressed or withheld by news organizations, in that the bias cannot be undone by rational or sophisticated agents since they do not know how much information the news supplier has and when information is being withheld (Bernhardt et al., 2008; Anderson and McLaren, 2012). In financial markets, preference for or disagreement with a journalist's report or media channels' views can affect stock market behaviors and financial valuation (Dougal et al., 2012; Gurun and Butler, 2012; Hillert et al., 2018). In line with these studies on media bias, I provide additional evidence that tone in the news, measured by sentiment, leads investors to form a biased perception of uncertainties in risky assets, and thus make a biased decision to acquire firm-specific information in equilibrium. To the best of my knowledge, this analysis is the first study to bridge this gap on the effect of biased public information in the news on investors' acquisition of firm-specific information.

Finally, the theoretical result regarding investors' biased information acquisition decision in this study is also in line with studies on information rigidity (Sims, 2003; Coibion and Gorodnichenko, 2012, 2015; Bouchaud et al., 2019) and extrapolation (Alti and Tetlock, 2014; Greenwood and Shleifer, 2014; Hirshleifer et al., 2015; Choi and Mertens, 2019). On the one hand, an investors' reluctance to take on board new information, as expounded in information rigidity studies, is similar to the implications of the model developed in this study. Sticky information acquisition, whereby investors are less willing to acquire firm-specific information in a biased belief equilibrium, is caused by positive sentiment in the news. On the other hand, the overweighted amount of recent information used by investors in information extrapolation research is similar to the present study's understanding of negative sentiment in the news. Investors acquire too much firm-specific information compared to what they would acquire in a rational expectations scenario. Although the biased incorporation of information for the purposes of making an investment decision in the model presented in this study shares similar psychological behaviors to those described in the information rigidity and extrapolation studies, the channel for bias in this study's model is different, as bias mainly originates from the news media itself, rather than from investors.

The study is organized as follows. Section 2.2 introduces a theoretical model of biased information acquisition and develops testable predictions. Section 2.3 describes the dataset used for the empirical studies and provides data summary statistics. Section 2.4 details the empirical results of the tests, which show that with news sentiment held as a proxy for biased public information, investors' biased perception of uncertainties gives rise to biased information acquisition. Section 2.5 entails a test conducted on the pricing power of firm-specific news sentiment on cross-sectional stock returns. Section 2.6 offers the study's conclusions. Robustness tests are in the Appendix B.

2.2 Information Acquisition Model with Biased Beliefs

This study reports the development of a model for how investors become informed as a way of reducing the uncertainty of risky asset investments. I assume that the acquisition of firm-specific information is costly. This cost can be understood as, among other things, hiring financial advisers, analyzing financial reports, gathering information about consumers' preferences, buying financial data or outsourcing financial data analysis. Therefore, only a fraction of investors will choose to pay for such costly information. This study demonstrates how the tone of exogenous costless public information from news media may give investors a biased rather than rational perception of the uncertainty surrounding risky assets. As a consequence, firm-specific information acquisition deviates from the rational expectations equilibrium.

2.2.1 Model Setup

The principles of this static model for information acquisition are based on Grossman and Stiglitz (1980), and those of biased public information are based on the proposition of biased information transmission by Hirshleifer (2020). The economy of the current model is similar to that of Kacperczyk et al. (2016) and Andrei et al. (2019). The biased belief draws on work by Odean (1998), García et al. (2007), Ko and Huang (2007) and Heller and Winter (2020) in allowing irrationality in the economy. However, the key argument of biased belief in this model is the result of biased public information such as news sentiment and not investors' psychological bias, which has been broadly addressed in the behavioral literature.

In a hypothetical economy populated by a continuum of investors indexed by $i \in [0, 1]$, there are three periods $t \in \{0, 1, 2\}$. At $t = 0$ investors read costless news about the market or particular firms they are considering an investment in and make a decision on whether or not to acquire more private information about firm-specific conditions to inform their investment decision. Investors trade competitively at $t = 1$ in the financial market. At $t = 2$ the payoff of financial assets will be realized and investors will consume their terminal wealth.

Investors trade a risk-free asset and a risky asset. The risk-free asset pays a gross interest rate of r_f and the supply is infinitely elastic. The risky asset (stock) has an equilibrium price P_1 at $t = 1$ and pays a risky dividend at $t = 2$:

$$D_2 = \overline{D} + m_2 + e_1 \tag{2.1}$$

The risky dividend payoff has three components: a mean payoff $\overline{D} > 0$, a market component $m_2 \sim N(0, \sigma_m^2)$ and a firm-specific component $e_1 \sim N(0, \sigma_e^2)$. The firm-specific component will be available at $t = 1$ to investors who choose to become informed. Therefore, informed investors will perfectly observe e_1. Additionally, m_2 and e_1 are independent.

The mean payoff $\overline{D}$ is common knowledge for all investors at $t = 0$. Investors with rational expectations know the variance (uncertainty) of the market component σ_m^2, and the variance (uncertainty) of the firm-specific component σ_e^2 at $t = 0$. However, investors' knowledge about σ_m^2 and σ_e^2 are biased by reading news with non-neutral tones about the market or a firm at $t = 0$.[2]

This understanding of biased information in the news sheds light on one of the major propositions stated by Hirshleifer (2020), namely, that information transmission bias results from misreporting, in that a signal received by investors is subject to an upward or downward bias in the signal itself. In addition, information receivers interpret the biased information from news naively and without adjusting for the bias in the news. In fact, investors' unawareness or naivety about the bias in the news can be easily relaxed, because Bernhardt et al. (2008) and Anderson and McLaren (2012) developed models to confirm that rational agents cannot undo this bias caused by the suppression or withholding of information by suppliers.

The assumption of rational or sophisticated investors may make the model in the current study even more parsimonious or generalized, but without including a verification of the biased effect from public information in the news, I retain the customary assumption of naivety in the model proposed by Hirshleifer (2020). Hence, following the Hirshleifer (2020) research, this study defines the tone from news – which is measured by sentiment in the way news providers describe the stock market or particular

[2] I outline a simple model to describe why news or media always has bias $E[b] \neq 0$ in the Appendix B.7.

firms – as tending to be either more optimistic or pessimistic. This is the bias (b) in costless information reporting to investors. Investors' prior beliefs of both market or firm-specific components' uncertainty is subject to bias through the tone of the market- or firm-specific news respectively, which they receive at $t = 0$. Furthermore, all investors are homogeneously biased by the tone of news.[3]

For simplicity, I assume that the biased effect of the news sentiment about the whole market (S_m) is independent of the firm-specific news sentiment (S_e).[4] Therefore, the uncertainty of the market component σ_m^2 is only biased by the market news sentiment, and σ_e^2 is only biased by the sentiment in firm-specific news. Finally, as investors are naive about the validity of news tone, they make trading or investment decisions based on their unconscious, biased beliefs.

As argued by Odean (1998), Ko and Huang (2007) and Heller and Winter (2020), I assume that all investors' subjective beliefs follow a bias function $\beta(S_j, \sigma_j^2)$, where σ_j^2 is a constant of correct beliefs, and $j \in (m, e)$. This posits that the biased prior belief of both market and firm-specific components' uncertainty is parameterized by the bias function:

$$\beta_{j \in (m,e)}(S_j, \sigma_j^2) = \sigma_{b,j}^2 \begin{cases} S_j \uparrow & \sigma_{b,m}^2 < \sigma_m^2, \;\; \sigma_{b,e}^2 < \sigma_e^2 \\ S_j = 0 & \sigma_{b,m}^2 = \sigma_m^2, \;\; \sigma_{b,e}^2 = \sigma_e^2 \\ S_j \downarrow & \sigma_{b,m}^2 > \sigma_m^2, \;\; \sigma_{b,e}^2 > \sigma_e^2 \end{cases} \tag{2.2}$$

where b denotes the investors' subjective biased belief throughout the study. Notably, bias in the news is not intended to advance a false perception or convince investors to alter their own perceptions. In fact, the effect of bias can be understood as presented in the study of Gentzkow et al. (2015), who defined the bias as a partial order on reporting strategies that shift agents' beliefs about a firm strategy to either the right or the left. In my study, the bias shifts investors' beliefs towards either more optimistic or more pessimistic perceptions of the uncertainties. Therefore, the biased information from news media slants investors' perception, causing them either to overestimate or underestimate σ_m^2 and σ_e^2, and does not mislead investors into changing the mean of the distribution about m_2 and e_1.

The rationale for biased beliefs in the model is as follows: as the tone in news about the market or a particular firm grows more positive or optimistic, investors' certainty

[3]Since news is costless and available to all investors at $t = 0$, I assume all investors have the same biased beliefs about the uncertainties for tractability.

[4]Even though I make this assumption of independence in the theoretical model, I control the market news sentiment in all the empirical testing for robustness.

regarding the market or the firm's future performance will also grow, and vice versa.[5] If the tone in the news is neutralized ($S = 0$), meaning that the information from news is genuinely objective and devoid of bias, investors have a rational prior belief about the uncertainties. Since all investors are naive about the validity of biased information from news, they are behaving optimally by believing that their biased understanding of those uncertainties is indeed correct, even though, in fact, it is not.

At $t = 1$, there is a public signal about the market in the economy and the signal is available for all investors:

$$M_1 = m_2 + \eta_1 \tag{2.3}$$

where $\eta_1 \sim N(0, \sigma_\eta^2)$ and is independent from m_2 and e_1.

Following the Grossman and Stiglitz (1980) information acquisition model, at $t = 0$, all investors decide if they want to acquire the private information about e_1, which will be perfectly observed at $t = 1$. I denote the decision of investor i with variable L_0^i , where $L_0^i = 1$ denotes when investor i chooses to become informed and $L_0^i = 0$ indicates that she wishes to stay uninformed.

I assume that investors have CARA utility function with zero initial wealth[6] and maximize their expected utility with biased beliefs :

$$\mathbf{U}_b^i = \mathbb{E}_{b,0}^i \Big[- e^{-\alpha(W_2^i - cL_0^i)} \Big] \tag{2.4}$$

where α is the risk aversion coefficient and c is a positive information cost for those who choose to become informed about e_1 at $t = 1$. W_2^i is investor i's terminal wealth at $t = 2$.

Investors choosing to become informed by perfectly observing e_1 at $t = 1$ are denoted by I. Investors who choose to remain uninformed are denoted by U. Following the noisy rational expectations model proposed by Grossman and Stiglitz (1980), the uninformed investors are still able to learn e_1 partially through the perceived equilibrium price. This is described below.

At $t = 1$ investors choose their optimal portfolios:

$$q_1^i = \frac{\mathbb{E}_{b,1}^i[D_2] - r_f P_1}{\alpha Var_{b,1}^i[D_2]}, \qquad \text{for } i \in \{I, U\} \tag{2.5}$$

[5]In other words, the biased uncertainty is a monotonically decreasing function of news sentiment. I do not assume a particular form of the function between biased uncertainty and news sentiment. However, without loss of generality, one can simply assume a linear form $\sigma_{b,j}^2 = (1 - S_j)\sigma_j^2$.

[6]Without loss of generality, I suppress $W_0 = 0$ because the CARA utility maximization problem is independent of initial wealth.

where $\mathbb{E}^i_{b,1}$ and $Var^i_{b,1}$ are subject to investor i's biased beliefs. Following O'Hara (2003), I assume that the risky asset random supply x_1 is independent of m_2, e_1, η_1, and that x_1 is normally distributed with mean $\overline{x}$ and variance σ^2_x, or $N(\overline{x}, \sigma^2_x)$. With the exception of the case in which the random supply prevents a perfect revelation of e_1 through the price, the positive expected supply of the risky asset implies a risk premium in the model as traders demand compensation to hold the risky assets in equilibrium. Finally, with λ_1 denoting the proportion of informed investors, the equilibrium price of the risky asset is determined by the market clearing condition:

$$\lambda_1 q^I_1 + (1 - \lambda_1) q^U_1 = x_1 \tag{2.6}$$

Because investors are naive about the validity of the news tone, investors with biased perceptions of uncertainties believe they are acting optimally and the equilibrium is determined by investors' biased beliefs. Similar rationales can be found in Heller and Winter (2020). In two-player games, the authors argue that players are blind to their biased beliefs regarding the opponent's strategy and choose the best response strategy to their biased beliefs. The equilibrium yielded by the model of Heller and Winter's (2020) study is subject to the players' biased belief. Therefore, the equilibrium achieved in the model I discuss in this study falls within the ambit of the *biased belief equilibrium* proposed by Heller and Winter (2020).

2.2.2 Equilibrium

By virtue of investors' naivety about biased tones in the news information they consume, the biased belief equilibrium (BBE) in my study is obtained in the same manner as in a noisy rational expectations equilibrium model (REE). I posit that the investors' perceived pricing function is:

$$P_1 = A\overline{D} + BM_1 + Ge_1 - Kx_1 + H\overline{x} \tag{2.7}$$

As uninformed investors are able to partially learn the e_1 for free from the price, the informative signal from price revealing is defined as:

$$\hat{p}_1 = \frac{P_1 - A\overline{D} - BM_1 + (K - H)\overline{x}}{G} = e_1 - \frac{K}{G}(x_1 - \overline{x}) \tag{2.8}$$

The information set for informed and uninformed investors is different. For informed investors, the information set is $\mathcal{F}_I = \{\overline{D}, M_1, e_1, \hat{p}_1\}$. For uninformed investors, the information set is $\mathcal{F}_U = \{\overline{D}, M_1, \hat{p}_1\}$. Therefore, the following equations define optimal portfolio choice from (2.5) for informed and uninformed investors (see Appendix B.2 for the derivation):

$$q_1^I = \frac{\overline{D} + \frac{\sigma_{b,m}^2}{\sigma_{b,m}^2+\sigma_\eta^2} M_1 + e_1 - r_f P_1}{\alpha Var_{b,1}^I[D_2]} \tag{2.9}$$

$$q_1^U = \frac{\overline{D} + \frac{\sigma_{b,m}^2}{\sigma_{b,m}^2+\sigma_\eta^2} M_1 + \frac{\sigma_{b,e}^2}{\sigma_{b,e}^2+\frac{K^2}{G^2}\sigma_x^2}\hat{p}_1 - r_f P_1}{\alpha Var_{b,1}^U[D_2]} \tag{2.10}$$

The optimal portfolio from equations (2.9) and (2.10) clearly indicates that, on average, informed investors hold more of the risky assets ($q_1^I > q_1^U$) when the expected return is positive. This is because they perfectly observe e_1 at $t = 1$, thus reflecting a lower risk ($Var_{b,1}^I[D_2] < Var_{b,1}^U[D_2]$) which is bestowed on them by their superior information (O'Hara, 2003).

As noted above, investors are naive about their biased beliefs and use $\sigma_{b,m}^2$ and $\sigma_{b,e}^2$ instead of rational perceptions (σ_m^2, σ_e^2) to make their optimal investment decision. Therefore, the model is solved by the standard procedure introduced by Grossman and Stiglitz (1980) which uses the market clearing condition (2.6) to find the equilibrium price with investors' biased beliefs. The proof is provided in the Appendix B.2.

Proposition 1. *In equilibrium, the coefficients on the fundamental, public signal, private signal and supply noise in the investors' perceived pricing function are given by:*

$$A = \frac{1}{r_f}, \quad B = \frac{\sigma_{b,m}^2}{(\sigma_{b,m}^2 + \sigma_\eta^2) r_f}, \quad G = \frac{\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U\Phi}{r_f Z}, \quad K = \frac{(1-\lambda_1)\gamma\phi_U\Phi\frac{K}{G} + 1}{r_f Z},$$
$$H = \frac{(1-\lambda_1)\gamma\phi_U\Phi\frac{K}{G}}{r_f Z}, \quad \Phi = \frac{\sigma_{b,e}^2}{\sigma_{b,e}^2 + \frac{K^2}{G^2}\sigma_x^2}, \quad \frac{K}{G} = \frac{\alpha Var_{b,1}^I[D_2]}{\lambda_1}, \quad \phi_I = \frac{1}{Var_{b,1}^I[D_2]},$$
$$\phi_U = \frac{1}{Var_{b,1}^U[D_2]}, \quad Z = (\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U) r_f, \quad \gamma = \frac{1}{\alpha} \tag{2.11}$$

2.2.3 Information Acquisition in Investors' Biased Belief Equilibrium

As stated in Grossman and Stiglitz (1980), in equilibrium, investors must be indifferent when choosing whether to become informed or uninformed. The indifference condition yields the following equation (see the proof in Appendix B.3):

$$\frac{\mathbf{U}_b^I}{\mathbf{U}_b^U} = e^{\alpha c}\sqrt{\frac{Var_{b,1}^I[D_2]}{Var_{b,1}^U[D_2]}} = 1 \tag{2.12}$$

Proposition 2. *In investors' biased belief equilibrium, the proportion of investors who become informed* λ_1 *can be solved by the benefit and cost function:*

$$\Pi(*) = \frac{\lambda_1^2\sigma_{b,e}^2\delta + \alpha^2 Var_{b,1}^I[D_2]^2\sigma_x^2\delta - \alpha^2 Var_{b,1}^I[D_2]\sigma_x^2\sigma_{b,e}^2}{\alpha^2 Var_{b,1}^I[D_2]\sigma_x^2\sigma_{b,e}^2\delta} = 0, \quad \text{where} \quad \delta = e^{2\alpha c} - 1 \tag{2.13}$$

The implicit function (2.13) is jointly determined by λ_1 and the uncertainties ($Var_{b,1}^I[D_2]$, $\sigma_{b,e}^2$). The model yields investors' biased belief equilibrium λ_1 which depends on how investors perceive the uncertainties of market and firm-specific components. Therefore, the proportion of investors who are willing to observe e_1 in this model deviates from the rational expectations equilibrium which is customarily implied by Grossman and Stiglitz (1980).

On the one hand, if investors hold correct beliefs about σ_m^2 and σ_e^2, in which $S_j = 0$, the model yields the same results as would be found under rational expectations. This is mainly addressed by Andrei et al. (2019), who argue that investors' information demand depends on systematic (market) uncertainty. In fact, their study rests on the assumption that investors do not suffer information transmission bias, which is represented as $S_j = 0$ in the current study.

On the other hand, this study will relax the assumption of investors being devoid of biased beliefs. The model developed in this study comprehensively analyzes comparative statics concerning how investors' information acquisition about e_1 deviates from rational expectations. This is explained by information transmission bias derived from news sentiment. Correspondingly, the positive expected supply of the risky asset ($E[x_1] = \overline{x}$) in this study's model contributes an additional implication for how firm-specific news sentiment has return predictability as an information risk premium on the risky asset.

Corollary 1. *In equilibrium, from equation (2.13) under a necessary condition* $\Pi'(\lambda_1) \geqslant 0$*, since* $\frac{\partial \lambda_1}{\partial \sigma_{b,j}} > 0$ *and from equation (2.2),* $\sigma_{b,j}^2$ *monotonically decreases with* S_j*, the model predicts* $\frac{\partial \lambda_1}{\partial S_j} < 0$*, where* $j \in \{m, e\}$. (The Proof is available in Appendix B.4)

2.2.4 Information Acquisition with Biased Beliefs of Systematic Uncertainty

On the basis of Proposition 1, the price informativeness is defined as (see Appendix B.2 for the proof):

$$n_b = \frac{\rho^2}{1-\rho^2} = \frac{\lambda_1^2\sigma_{b,e}^2}{\alpha^2 Var_{b,1}^I[D_2]^2\sigma_x^2} \tag{2.14}$$

where ρ is the correlation between e_1 and $\hat{p}_1$. Holding $\sigma^2_{b,e}$ constant, price informativeness increases as more investors become informed ($\lambda_1 \uparrow$), are less risk-averse ($\alpha \downarrow$), have less systematic uncertainty ($Var^I_{b,1}[D_2] \downarrow$) or the random supply is less volatile ($\sigma^2_x \downarrow$).

Figure 2.1 depicts the relationship between λ_1 which is the investors' information demand and biased belief of systematic uncertainty $Var^I_{b,1}[D_2]$. It should be noted that, since investors' belief about the uncertainty of m_2 is biased by sentiment from the consumption of news on the condition of the market, as a consequence, the $Var^I_{b,1}[D_2]$ is directly biased by linear projection of $\sigma^2_{b,m}$ and σ^2_η (see Appendix B.2 for the proof). First, if we keep σ^2_e unbiased, the blue line in Figure 2.1 shows zero bias ($S_m = 0 \rightarrow b = 0$) in the news consumed by investors about the market. Therefore, the model is reconciled with the rational expectations as argued by Andrei et al. (2019). The theoretical maximum of information demand is reached when the systematic uncertainty is at:

$$Var^I_{b=0,1}[D_2]^* = Var^I_1[D_2]^* = \frac{\sigma^2_e}{2(e^{2\alpha c} - 1)} \tag{2.15}$$

and the informed investors' information quality under rational expectations is defined as:

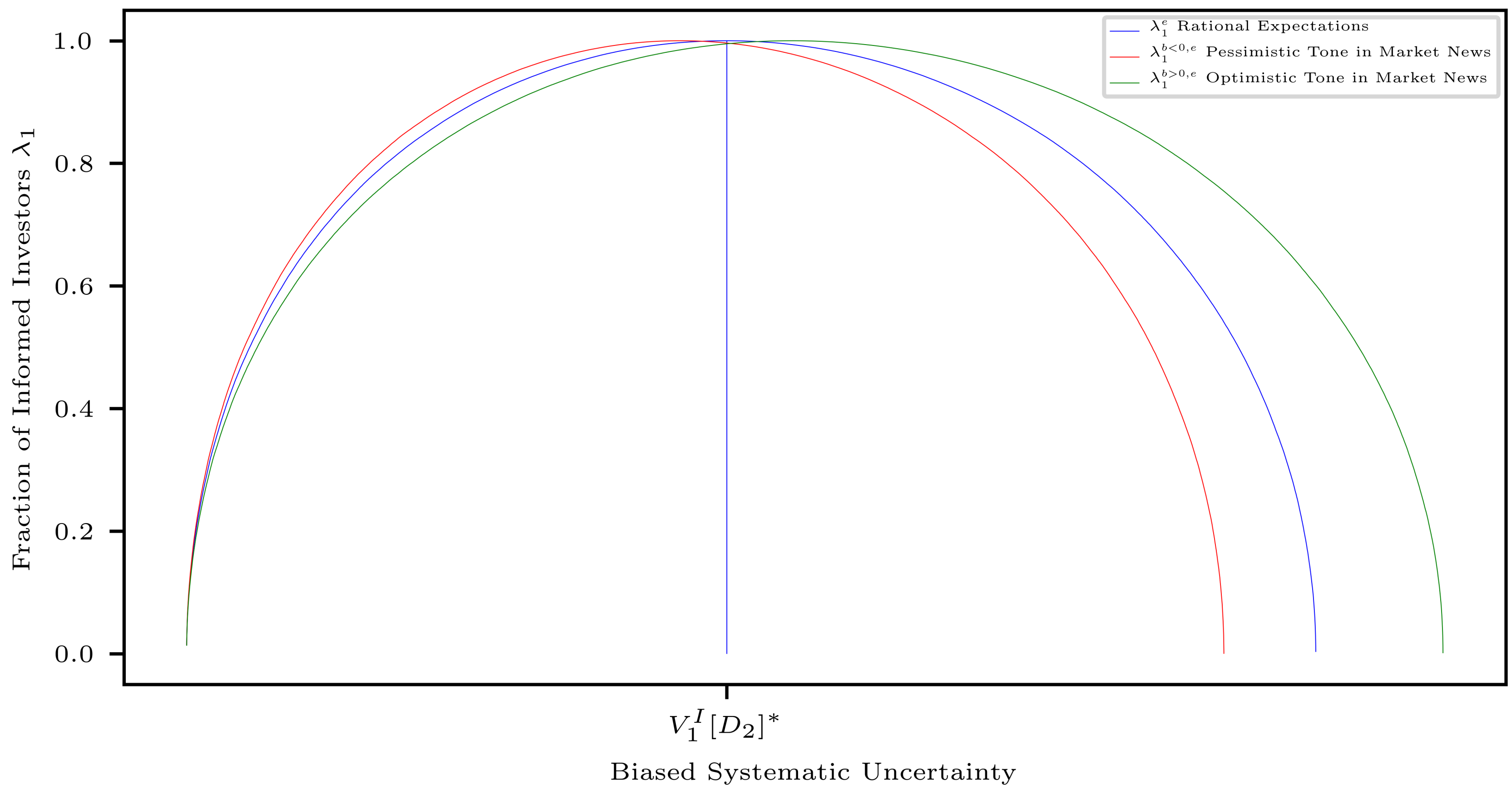
$$v = \frac{\sigma^2_{b=0,e}}{Var^I_{b=0,1}[D_2]}. \tag{2.16}$$

The hump shape is due to the trade-off between price informativeness (n) and informed investors' quality of information (v). Before the systematic uncertainty reached at $Var^I_1[D_2]^*$, as the market becomes more uncertain, higher systematic uncertainty, which reduces price informativeness, motivates investors' desire to acquire private information about e_1. Nevertheless, if the market becomes too uncertain (above the level $Var^I_1[D_2]^*$), it is worthless for investors to acquire information about e_1, because the significantly decreased quality of informed investors' information makes them reluctant to pay anything at all to observe e_1. This link between investors' information demand and economic uncertainty is mainly addressed by Andrei et al. (2019).

The novel study of Dougal et al. (2012) finds evidence that journalists are significant predictors of the positive–negative words balance of writing in the "Abreast of the Market" column in *The Wall Street Journal*. Their persistent bullishness and bearishness has a significant impact on the financial market. As a consequence, investors consume news about the market or economic conditions before they make investment or trading decisions, and as long as the sentiment from market news is not neutral ($S_m \neq 0$), their beliefs are biased by the market news sentiment, either overstating $\sigma^2_{b,m} > \sigma^2_m$ as $S_m \downarrow$ or understating $\sigma^2_{b,m} < \sigma^2_m$ as $S_m \uparrow$.

FIGURE 2.1: Stock Market News Sentiment Impact on Information Acquisition as a function of $Var^I_{b,1}[D_2]$

This figure plots the equilibrium proportion of investors who want to acquire firm-specific information e_1 as a function of perceived systematic uncertainty $Var^I_{b,1}[D_2]$ by investors. λ^e_1 and $\lambda^{b,e}_1$ denote the equilibrium level of the proportion of informed investors under the rational expectations and biased beliefs, respectively. I calibrate cost of information about e_1, $c = 0.002$; the variance of supply $\sigma^2_x = 0.2$; the variance of the public signal noise σ^2_η is randomly drawn from a uniform distribution; and investors' coefficient of risk aversion $\alpha = 3$. I set the rational perception of firm-specific uncertainty $\sigma^2_e = 0.0623$ as the median value of the sum of squared residuals from AR(1) of firms' $EPS_{i,t}$ in the sample period. The rational perception of systematic component uncertainty σ^2_m is randomly drawn from the distribution of S&P 500 realized variance in the sample period. Without loss of generality, I assume the bias function $\beta(S_m, \sigma^2_m)$ in equation (2.2) is linear as $\sigma^2_{b,m} = (1 - S_m)\sigma^2_m$ to generate the biased perception of systematic component uncertainties. The blue curve is the equilibrium fraction of informed investors under the rational expectations without the biased impact from market news sentiment ($S_m = 0$) on σ^2_m. The red curve is the equilibrium fraction of informed investors, as $\sigma^2_{b,m} \uparrow$ is upward (positively) biased when the tone in the news about the stock market is decreased, or made more pessimistic, by one standard deviation $S_m \downarrow$ from $S_m = 0$ under the rational expectations. The green curve is the equilibrium fraction of informed investors, as $\sigma^2_{b,m} \downarrow$ is downward (negatively) biased when the tone in the news about the stock market is increased, or made more optimistic, by one standard deviation $S_m \uparrow$ from $S_m = 0$ under the rational expectations. The standard deviation of stock market news sentiment σ_{S_m} is from the TRMI stock market news sentiment index in the U.S. and the value is 0.183.



Tesser and Rosen (1975) state that people's reluctance to report bad news is a means of shielding discomforting feelings from public display. This drives more positive reporting by information disseminators, as acknowledged by Hirshleifer (2020). The green line in Figure 2.1 shows that as market news sentiment S_m grows to be more optimistic, the fraction of investors who want to become informed about e_1 in the biased belief equilibrium is always less than that seen in the rational expectations equilibrium at every level of rationally perceived systematic uncertainty ($Var^I_{b=0,1}[D_2]$) before it reaches $Var^I_1[D_2]^*$. This is because, at each level of $Var^I_1[D_2]$, investors' belief about $\sigma^2_{b,m} \downarrow$ is negatively biased. Similarly, $Var^I_{b,1}[D_2] \downarrow$, from the rational perception σ^2_m is due to investors consuming news containing an optimistic tone or sentiment about the market. Investors irrationally place more aggressive orders with the negatively biased systematic uncertainty; thus, investors with this biased belief of $Var^I_{b,1}[D_2] \downarrow$ perceive the price as more informative than the price informativeness in rational expectations. Because of the systematic uncertainty's inverse relationship with price informativeness and its dominant effect on investors' information demand to observe e_1, investors are less willing to acquire information about e_1 in the biased belief equilibrium due to the positively biased price informativeness ($n_b \uparrow$) differing from the negatively biased $Var^I_{b,1}[D_2] \downarrow$.

Negativity bias has been broadly addressed in the psychological literature. Rozin and Royzman (2001) and Baumeister et al. (2001) argue that people have a tendency to pay more attention to negative information and to interpret information negatively. Hence, journalists use negative tones in their work to attract investors' attention to consume news and improve the profit of selling news (Arango-Kure et al., 2014). The red line in Figure 2.1 shows that as the market news sentiment becomes more pessimistic, the proportion of informed investors in the biased belief equilibrium is greater than the proportion of investors who want to become informed in the equilibrium under rational expectations at every level of $Var^I_{b=0,1}[D_2]$ before it reaches $Var^I_1[D_2]^*$. When investors consume market news with a negative tone, this engenders greater perception of uncertainty about economic conditions and investors tend to perceive a higher, $\sigma^2_{b,m} \uparrow$, then $Var^I_{b,1}[D_2] \uparrow$. Thus, the positive biased $Var^I_{b,1}[D_2] \uparrow$ drives investors irrationally to trade less aggressively. As a consequence, investors with the biased belief of $Var^I_{b,1}[D_2] \uparrow$ perceive that price is not as informative ($n_b \downarrow$) as in the rational expectations model. This negatively biased price informativeness ($n_b \downarrow$) motivates investors to pay costs for observing e_1. In equilibrium, the positively biased perception of market uncertainty from negative news sentiment leads to more information acquisition regarding e_1 than observed in the rational expectations scenario.

2.2.5 Information Acquisition with Biased Beliefs of Firm-Specific Uncertainty

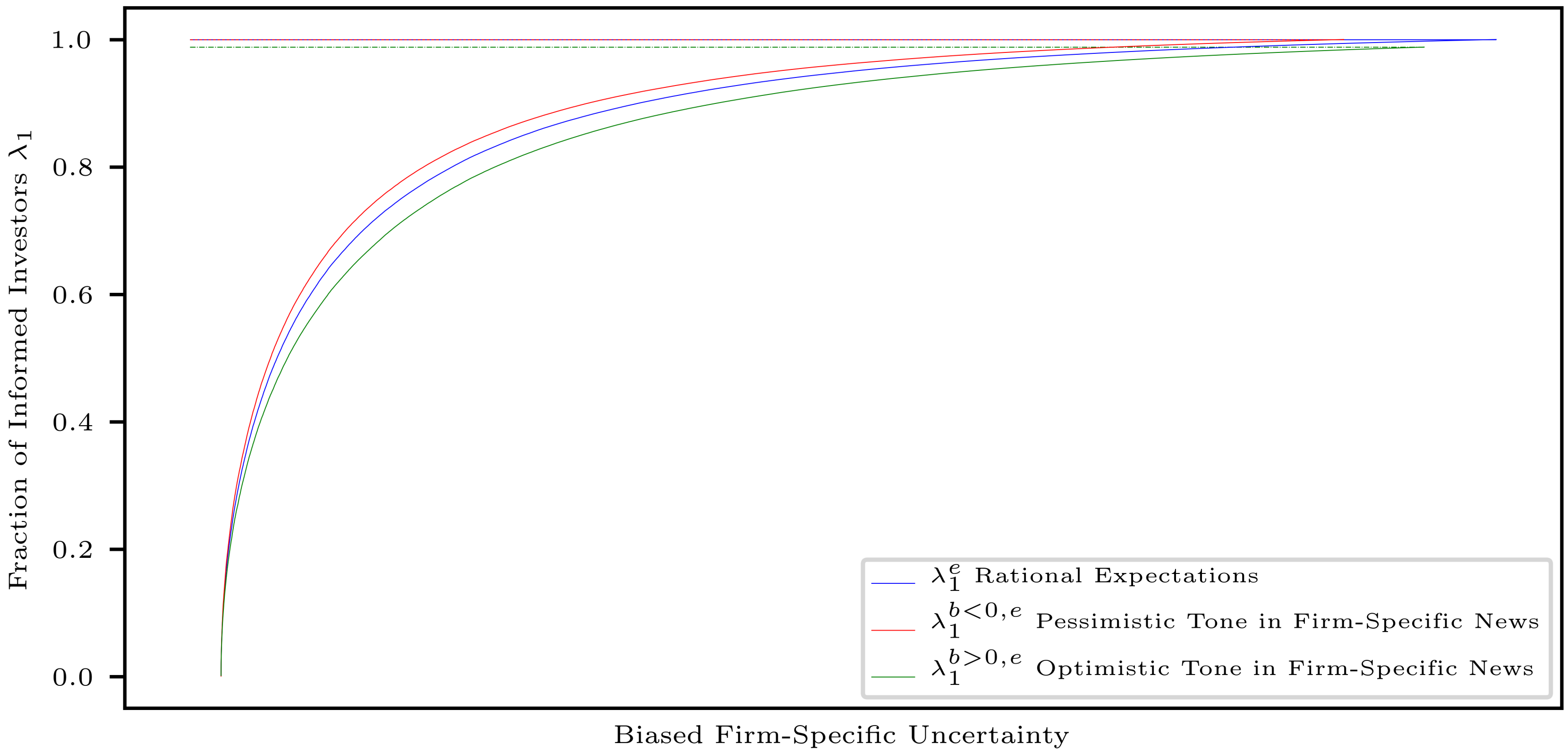
To study the comparative statics of the impact of the biased perception of $\sigma^2_{b,e}$ on investors' information acquisition, I first reconcile the model with rational expectations ($S_e = 0 \to b = 0$) regarding the relationship between λ_1 and $\sigma^2_{b=0,e}$. Equation (2.13) implies that λ_1 is a non-decreasing function of $\sigma^2_{b=0,e}$ in the range of $\Pi^{'}(\lambda_1) \geq 0$ and it yields $\frac{\partial \lambda_1}{\partial \sigma_{b=0,e}} > 0$ (see Appendix B.4 for the proof). Increasing $\sigma^2_{b=0,e}$ for a given λ_1 and $Var^I_{b=0,1}[D_2]$ indicates that the variance of e_1 ($Var^U_{b=0,1}[e_1|\hat{p}_1]$) perceived by the uninformed investors must be increased and that the indifference condition function shifts downward from the equilibrium level. As a result, and to maintain the indifference condition at the equilibrium level, there must be more investors becoming informed, and thus a higher λ_1 in equilibrium (Grossman and Stiglitz, 1980). This intuition is also consistent with the findings presented in Veldkamp (2006), whereby the uncertainty of the given price of asset payoff is largely relative to the uncertainty of given information (here e_1) on the payoff. Therefore, when $\sigma^2_{b=0,e}$ is high, information that reveals e_1 is more valuable because the degree of reduction of $Var^U_{b=0,1}[D_2|\hat{p}_1]$ is considerable. Thus, risk-averse investors are more willing to become informed to remove the firm-specific uncertainty $\sigma^2_{b=0,e}$ when it is higher, more specifically, at every level of market uncertainty.

If we assume that the market news is not biased by any tone ($S_m = 0$), the blue line in Figure 2.2 is the λ^e_1 denoted as the equilibrium level under the rational expectations ($S_e = 0$) as λ_1 increases with σ^2_e. Despite investors' optimal behavior in the market, their perception of $\sigma^2_{b,e}$ may be biased by the tone (sentiment) in the firm-specific news. Investors are unconscious of their being biased by the news sentiment; consequently, the λ^e_1 deviates to $\lambda^{b,e}_1$ and the b denotes biased belief equilibrium.

As discussed by Berger and Milkman (2012) and Berger (2014), people are more likely to share and discuss positive content in the news or media rather than negative content, in order to maintain a reputation for providing useful information. For example, when choosing a wide range of products, advising on what to buy is more helpful than advising on what not to buy, as discussed in the marketing study of Hirshleifer (2020). Gurun and Butler (2012) find the evidence that local media tend to provide a positive slant when reporting on local firms, typically to encourage advertising expenditure from local firms. Additionally, as argued in the accounting literature, managers tend to release good news vs. bad news strategically for their own benefit - a manifestation of the agency problem in corporations (Kothari et al., 2009; Bao et al., 2019; Ahn et al., 2019).

FIGURE 2.2: Firm-Specific News Sentiment Impact on Information Acquisition as a function of $\sigma^2_{b,e}$

This figure plots the equilibrium fraction of investors who want to acquire firm-specific information e_1 as a function of perceived firm-specific uncertainty $\sigma^2_{b,e}$ by investors. λ^e_1 and $\lambda^{b,e}_1$ denote the equilibrium level of the proportion of informed investors under the rational expectations and biased beliefs, respectively. I calibrate cost of information about e_1, $c = 0.002$; the variance of supply $\sigma^2_x = 0.2$; the variance of the public signal noise σ^2_η is randomly drawn from a uniform distribution; and investors' coefficient of risk aversion $\alpha = 3$. The rational perception of firm-specific uncertainty σ^2_e is drawn randomly from the distribution of the sum of squared residuals from AR(1) of firms' $EPS_{i,t}$. I set rational perception of systematic component uncertainty $\sigma^2_m = 0.031$ as the mean S&P 500 realized variance in the sample period. Without loss of generality, I assume the bias function $\beta(S_e, \sigma^2_e)$ in equation (2.2) is linear as $\sigma^2_{b,e} = (1 - S_e)\sigma^2_e$ to generate the biased perception of firm-specific uncertainties. The blue curve is the equilibrium fraction of informed investors under the rational expectations without the biased impact of firm-specific news sentiment ($S_e = 0$) on σ^2_e. The red curve is the equilibrium fraction of informed investors, as $\sigma^2_{b,e} \uparrow$ is upward (positively) biased when the tone in the news about particular firms' performance is decreased, or made more pessimistic, by one standard deviation $S_e \downarrow$ from $S_e = 0$ under the rational expectations. The green curve is the equilibrium fraction of informed investors, as $\sigma^2_{b,e} \downarrow$ is downward (negatively) biased when the tone in the news about firm-specific information is increased, or made more optimistic, by one standard deviation $S_{b,e} \uparrow$ from $S_e = 0$ under the rational expectations. The standard deviation of firm-specific news sentiment σ_{S_e} is from the TRMI news sentiment indices of U.S.-listed firms and the value is 0.394.



The green line in Figure 2.2 shows that the curve of biased belief equilibrium $\lambda_1^{b,e}$ is shifted downward and ends earlier in comparison to λ_1^e in rational expectations at every level of σ_e^2. The decrease in information acquisition from investors is due to an increment in the positive tone of firm-specific news ($S_e \uparrow$) which leads to a negatively biased perception of firm-specific uncertainty $\sigma_{b,e}^2 \downarrow$. Since investors are biased to believe that a firm's future performance is more certain, *ceteris paribus*, the benefit derived from a reduction in the uncertainty about the payoff $Var_{b,1}^U[D_2]$ by knowing e_1 is underestimated by investors. Additionally, the quality of information v_b is under-perceived, because the negatively biased uncertainty about $\sigma_{b,e}^2$ makes investors feel less inclined to shed risk while keeping the systematic uncertainty unchanged. Overall, in the equilibrium with a biased belief that is more optimistic about firm-specific uncertainty $\sigma_{b,e}^2 \downarrow$, investors are less willing to pay the extra cost of acquiring the private information about e_1 and $\lambda_1^{b,e} < \lambda_1^e$ as $S_e \uparrow$.

As argued in the financial textual analysis literature, researchers find evidence that the frequency of negative words found in firm-specific news or online media dictates the overall tone of the report (Tetlock et al., 2008; Chen et al., 2014). However, the impact of negative tone in firm-specific news on investors' information acquisition decisions is unexplored. As shown in Figure 2.2, the red curve is investors' positively biased information demand from rational expectations. This is due to investors' positive bias about the firm-specific uncertainty $\sigma_{b,e}^2 \uparrow$ giving rise to an increment in the negativity or more pessimism in the tone of the firm-specific news. Intuitively, by reading firm-specific news with a more pessimistic tone, investors tend to predict that the firm's performance will be more uncertain in the future. As a consequence, investors over-perceive the value of information e_1 and the benefit of the reduction in $Var_{b,1}^U[D_2]$ by acquiring the information about e_1. Additionally, the quality of information is also overstated by a positively biased $\sigma_{b,e}^2$ while holding the systematic uncertainty constant. In sum, investors are willing to become informed as more negative sentiment ($S_e \downarrow$) exists in the firm-specific news; thus, there is an excess information acquisition in equilibrium.

In the Appendix B.10, I plot another figure (Figure B.1) as a different view to show the fraction of informed investors as a function of rational perception of market uncertainty respecting biased beliefs of firm-specific uncertainty. Overall, the tone in either market news or firm-specific news raises a deviation of investors' information acquisition in equilibrium. As long as there is a non-neutral tone ($S_j \neq 0$) in the news, investors are either "sticky" or "extrapolated" to acquire private information about the firm-specific component.

2.2.6 Deviation of Information Risk from Rational Expectations

As argued in previous sections, news sentiment deflects investors' information acquisition about e_1 away from rational expectations due to biased beliefs about uncertainties arising from the biased tone in the news. Consequently, the monotonically decreasing relationship between the proportion of informed investors (λ_1) and news sentiment, especially for firm-specific news (S_e), results in a deviation of information risk in the risky asset and, as a consequence, in the predictability of expected returns.

Proposition 3. *Expected risky asset return is* $E[R_2] = \frac{\alpha\overline{x}}{\lambda_1\phi_I+(1-\lambda_1)\phi_U}$ *and* $\frac{\partial E[R_2]}{\partial\lambda_1} < 0$. *From Corollary 1,* $\frac{\partial\lambda_1}{\partial S_e} < 0$*, therefore, sentiment in the firm-specific news has a positive predictability on the risky asset expected return* $\frac{\partial E[R_2]}{\partial S_e} > 0$. (see Appendix B.5 for the proof.)

When news is not biased in its tone, the positive expected supply $\overline{x}$ implies a risk premium ($E[R_2]$) by holding the risky asset, as proposed by O'Hara (2003), due to the information risk between informed and uninformed investors in forming their investment portfolios. However, in my study, firm-specific news sentiment generates deviations in information risk because of the deviation in the firm-specific information acquisition by investors. As a consequence, there is an implied return predictability by firm-specific news sentiment. The theoretical foundation of sentiment predictability on stock returns from firm-specific news is under-explored and quite different from studies in the extant literature.[7] Therefore, this study discusses the theoretical implications of Proposition 3 through deviations in information risk resulting from firm-specific news sentiment, to argue why firm-specific news sentiment can predict expected returns.

First, as long as price-revealing does not perfectly uncover the private information acquired by informed investors (here, e_1), this causes a non-diversified information risk to arise in the risky asset (O'Hara, 2003).[8] Additionally, as implied in a partially revealing rational expectations model, it is not possible for all investors to acquire the private information for all assets. This is because investors will value the benefit and cost in line with the indifference condition in equilibrium in order to make information acquisition decisions (Grossman and Stiglitz, 1980). Therefore, the extent to which information is private differs across assets based on the different degree of information

[7] Additionally, there is a paucity of studies on positive firm-specific news sentiment in the existing literature. See related studies of firm-specific news sentiment predictability by Busse and Green (2002); Antweiler and Frank (2004); Tetlock et al. (2008);Chen et al. (2014); and Ke et al. (2019).

[8] As stressed by O'Hara (2003), even where investors hold portfolios with the same assets, they will have different beliefs about the expected payoff of each asset due to different information advantages between informed investors and uninformed investors. As a consequence, uninformed and informed investors hold different relative weights of risky assets in their portfolios.

risk in the assets. Consequently, traders demand extra compensation or expected returns to hold the assets when the information risk is large. Intuitively, the more investors choose to acquire private information ($\lambda_1 \uparrow$), the more the price will become informative in reflecting private information e_1. This will also serve to reduce the rate of privateness of the information, since the price discovery becomes more effective in revealing the private information (O'Hara, 2003). This intuition yields $\frac{E[R_2]}{\partial \lambda_1} < 0$, and the expected return in the model can be seen as a risk premium to compensate the information risk of the risky asset in a rational expectations equilibrium.

Second, as discussed above, the firm-specific news can be biased with either a more positive ($S_e \uparrow$) or more negative tone ($S_e \downarrow$). Investor's information acquisition about e_1 deviates through the channel of a biased belief about $\sigma^2_{b,e}$ determined by S_e. In the biased belief equilibrium, the risky asset has a proportion of informed investors that is greater or lower than that which deviates from the number of investors who become informed about e_1 under the rational expectations. This $\lambda_1^{b,e}$ deviation causes the information risk-compensating expected return of the risky asset to be higher or lower than the expected return at λ_1^e.

Corollary 2. *If the tone (sentiment) in the firm-specific news tends to be more positive ($S_e \uparrow$), in a biased belief equilibrium, this positive tone predicts relatively higher expected returns than the rational expectations equilibrium $E_b[R_2] > E_r[R_2]$, where b and r denote the biased belief and rational expectations equilibrium respectively.* (See the proof in Appendix B.5.)

The more positive sentiment in the firm-specific news results in investors feeling less uncertain about the firm-specific component e_1 and perceiving a negatively biased $\sigma^2_{b,e}$. In the biased belief equilibrium, there are fewer informed investors than the situation brought about by rational expectations ($\lambda_1^{b,e} < \lambda_1^e$). When less informed investors trade in the market, their trading incorporates little new information into the price through the price discovery process. Correspondingly, uninformed investors face more information risk, because they cannot learn much from the equilibrium price about the private information obtained by informed investors. Compared to the rational expectations equilibrium, the risky asset in this biased belief equilibrium is in fact riskier because the price discovery process is not as informative as it should be. Intuitively, traders require greater compensation to hold this risky asset since its information risk is increased by the incremental "privateness" of information. This incremental information risk comes from investors with the biased belief of $\sigma^2_{b,e} \downarrow$ as $S_e \uparrow$ being reluctant in their acquisition of private information. In sum, the more positive sentiment bias in the firm-specific news generates more information risk which is compensated for by a higher expected return of the risky asset. Finally, Corollary 2 yields an empirical prediction:

Hypothesis 1: As sentiment increases or becomes more positive or optimistic in firm-specific news, the expected return of the risky asset increases.

A more negative sentiment in the firm-specific news yields the opposite effect. In fact, the theoretical implication of negative tone in firm-level news implies less information risk in the equilibrium with biased beliefs.

Corollary 3. *If the tone (sentiment) in the firm-specific news tends to be more negative ($S_e \downarrow$), in a biased belief equilibrium, this negative tone predicts relatively lower expected returns than in the rational expectations equilibrium $E_b[R_2] < E_r[R_2]$, where b and r denote the biased belief and rational expectations equilibrium respectively.* (See the proof in Appendix B.5.)

The more pessimistic or negative tone in the firm-specific news causes investors to feel more uncertain about the firm's future performance, resulting in a positively biased perception of $\sigma^2_{b,e}$. Because investors are risk-averse and may place more value on information about e_1 to reduce the uncertainty, in the biased belief equilibrium, more investors are inclined to acquire the information about e_1 than in the case of rational expectations ($\lambda_1^{b,e} > \lambda_1^e$). Since there are more informed investors trading in this biased belief equilibrium, the price discovery process sees additional new information incorporated into the price. Uninformed investors can learn more about the firm-specific component e_1 from the equilibrium price through the trading process. Compared to the rational expectations equilibrium, the asset traded in the market is less risky due to an excess of investors becoming interested in being informed, causing the price discovery process to be more informative than it should be in respect of the asset. Hence, uninformed investors face relatively less information risk than they face in the rational expectations model.[9] Traders require less compensation or a lower expected return to hold the asset in equilibrium, as there is less information risk than when there is more negative sentiment in the firm-specific news. Finally, Corollary 3 yields the following empirical prediction:

[9] One could think of the extreme case as $\lambda_1^{b,e} = 1$, where, if the tone in the news about a company is strongly pessimistically biased, all risk-averse investors will panic and seek to acquire the information about e_1 to reduce their positively biased uncertainty. Intuitively, the asset is no longer risky as a consideration of information asymmetry, because the effect of excess information demand minimizes the information risk in the asset.

Hypothesis 2: As sentiment decreases and tends to be more negative or pessimistic in the firm-specific news, the expected return of the risky asset decreases.

2.2.7 Discussion

The theoretical model in my study shows that the effect of biased tone or sentiment found in the news deflects investors' acquisition of firm-specific information regarding the asset's fundamental payoff, in contrast to rational expectations. Essentially, investors' eagerness or reluctance to acquire private information in this model shares similar characteristics with studies concerned with information rigidity and extrapolation.[10] Although the model discussed in this chapter shares a key premise with these studies - namely, that investors' biased belief formation drives different information acquisition behaviors concerning their forecast or investment decision - the rationale for the deviation from the null to full information in equilibrium is quite different.

Most studies in the literature address investors' psychological irrationality including overconfidence, representative bias, etc. as proposed by Tversky and Kahneman (1974). However, in the present study, the main driver of biased decisions made by investors is the consumption of biased public information in the news. It may be objected that the naivety assumption still contributes to the factor of agents' psychological bias as a trigger of irrational decision-making based on the concept of *Homo economicus*. As a matter of fact, the naivety assumption can be thought of as a concession to the main argument that biased information in the news as another channel results in investors' irrational decision-making in addition to behavioral irrationality. In fact, media bias is persistent, and even rational or sophisticated consumers can not perfectly adjust for it. Information suppliers can manipulate the bias by suppressing or withholding information, motivated by either profit-seeking or political preference (Bernhardt et al., 2008; Anderson and McLaren, 2012). By the same token, the broadly addressed issue of information withholding in financial markets[11] contributes this particular type of bias to the process of information supply, and as a result, the Bayesian investors cannot perfectly adjust for the bias in the financial news they receive.

[10] First, the information rigidity model indicates that investors tend to undervalue new information and overvalue old information. Thus, predictability comes from the slow update of new information. Second, the information extrapolation model argues that investors overweight recent information and incorporate too much of it into forecasting. As a long-run correction, there is a reversal effect. See related studies by Coibion and Gorodnichenko (2012; 2015); Bouchaud et al. (2019); Alti and Tetlock (2014); and Bordalo et al. (2019).

[11] Studies in the accounting literature have comprehensively addressed managers strategically disclosing both negative and positive news to investors (Sletten, 2012; Amir et al., 2018; Baginski et al., 2018; An et al., 2020).

Moreover, in a seminal psychological study, Le Mens and Denrell (2011) propose that even when the naivety assumption is relaxed, systematic judgment errors are still made by rational agents. This is due to the possibility that they may be subject to asymmetry of information access or their information search is interested, rather than disinterested.[12] Le Mens and Denrell stress that even when rational agents without cognitive limitations apply legitimate corrections to the bias in the sample, the corrected bias might be skewed. Thus, using skewed estimators for judgment or decision-making causes either overestimation or underestimation by the population of interest in a study.

Altogether, naivety is not necessarily a key assumption in the model in order to cause systematically biased decision-making and can be easily relaxed.[13] Therefore, agents can be rational and behave optimally as they are under rational expectations, but to some extent they are affected by the biased news. Alternatively stated, the generation of a biased belief equilibrium by biased decision-making need not necessarily be the product of an investor's psychological irrationality.

In fact, if news sentiment can be seen as the impact of investor sentiment generating incorrect beliefs about firms' fundamentals, it should also have a short-term momentum followed by a long-run reversal correction (Tetlock, 2014b). However, instead of arguing for the biased belief in the value of fundamental payoff, which is broadly addressed in the literature,[14] this study argues that sentiment from news is the cause of investors' biased beliefs about fundamental uncertainty; and that this results in biased decisions on information acquisition. Finally, in equilibrium, the private information reflected in the price through the price discovery process is subject to these biased beliefs. Therefore, the "mispricing" in the presented theoretical model is not the result of the deviation in assets' fundamental value, but deviation in information acquisition. As a consequence, the theory suggests an empirical and testable prediction that firm-specific news sentiment has predictability on cross-sectional stock returns. Furthermore, the informativeness of the price is synchronized with investors'

[12] For example, when rational investors receive news, they may have their own preference on searching or analyzing information from the news based on their rational choice for constructing portfolios to maximize the payoff.

[13] The assumption of naivety only serves to simplify the study without solving a sub-game between investors and information suppliers such as news companies or journalists. In fact, the model can be extended to the solution of a sub-game, first between rational investors and news suppliers as studied in Kamenica and Gentzkow (2011) and Baron (2006) who show that even rational investors are subject to bias in the news. The rest of the analysis is followed by section 2.2.

[14] See related studies by De Long et al. (1990); Barberis et al. (1998); Baker and Wurgler (2006); and Huang et al. (2015).

information acquisition in the biased belief equilibrium. Hence, firm-specific news sentiment is expected to have persistent predictability on cumulative stock returns, up to a certain length of trading periods without reversal correction.

In addition to the return predictability of firm-specific news sentiment as discussed in section 2.2.6, one might question whether or not the sentiment from market- or economy-wide news is comparatively predictive of stock returns. As mentioned in section 2.2.4, Figure 2.1 shows a non-monotonically increasing relationship between fractions of informed investors and biased perception of systematic uncertainty. Therefore, under normal economic conditions, the market news sentiment yields positive predictability, much like the firm-specific news sentiment. Under very uncertain economic conditions - for example, an economic bubble or recession - the market news sentiment has a reverse effect in biasing investors' information acquisition. For instance, optimistic market news sentiment makes investors under-perceive genuine market uncertainty; when the market is very uncertain above the $Var^I_{b,1}[D_2]^*$, investors acquire more private information than they should according to rational expectations. As a result, the positive market news sentiment negatively predicts stock returns under highly uncertain economic conditions and vice versa. These reversal effects of the predictability of market news sentiment are consistent with studies by Tetlock (2007) and Garcia (2013). Although the compelling non-monotonic predictability from market news sentiment, subject to different economic conditions, yields interesting theoretical and empirical predictions, a more comprehensive study on this topic is an opportunity for future research.

2.3 Data

The daily stock-level news data used in the empirical study are collected from Thomson Reuter MarketPsych (TRMI). TRMI derives newsfeeds of newly published content from approximately 40,000 internet news sites. More specifically, the news or social media content of information is assembled via TRMI crawls through hundreds of financial news sites, including, for example, *The New York Times, The Wall Street Journal, The Financial Times, Seeking Alpha* and many other sources that are widely read by financial professionals. In contrast to the traditional method of lexical analysis used in textual study, the technology used to create TRMI overcomes several shortcomings of the conventional approach broadly used in extant finance and economics studies (detailed information can be found in Peterson (2016)).

Figure 2.3: Stock Market News Sentiment vs. Firm News Sentiment I

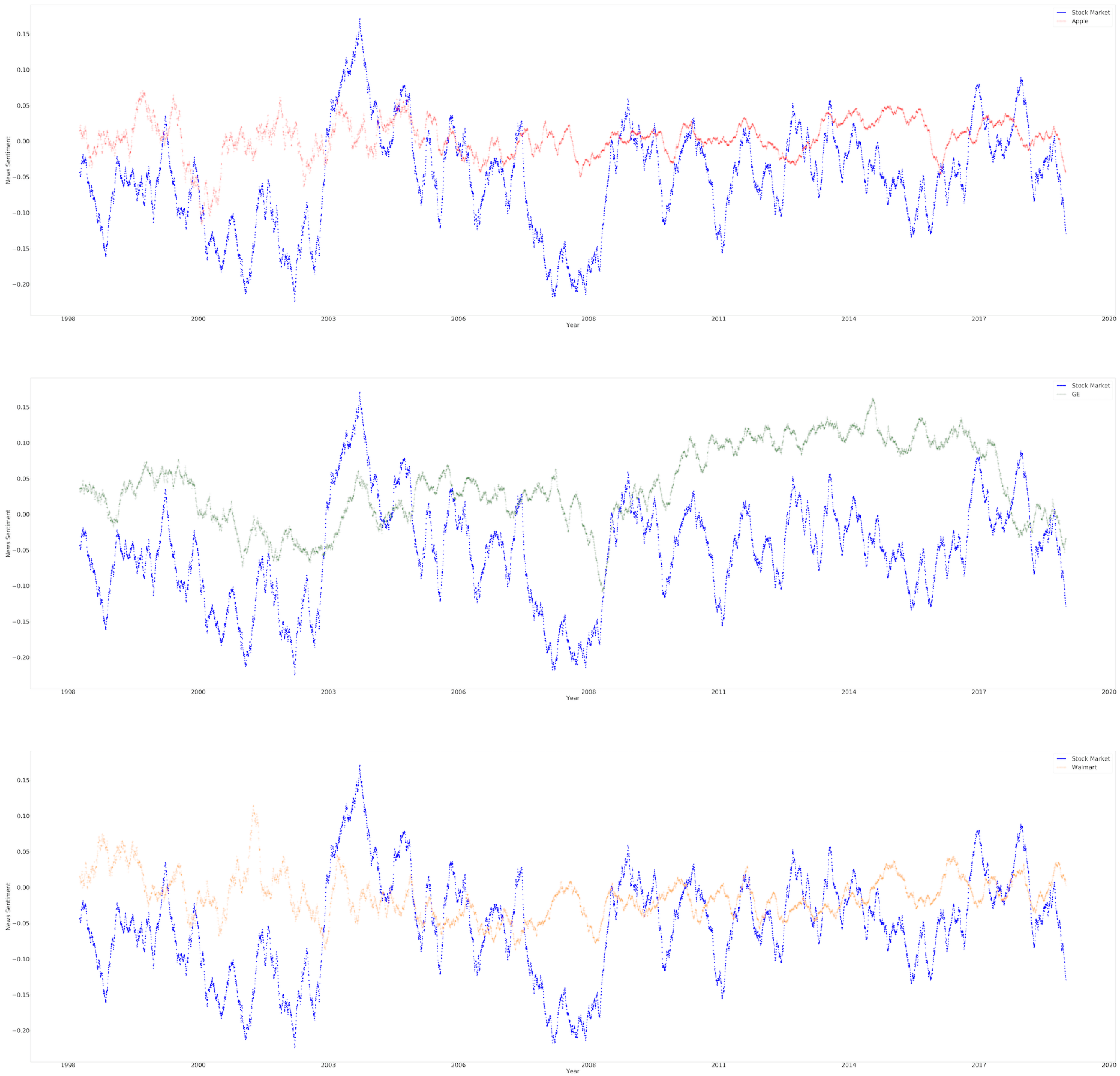
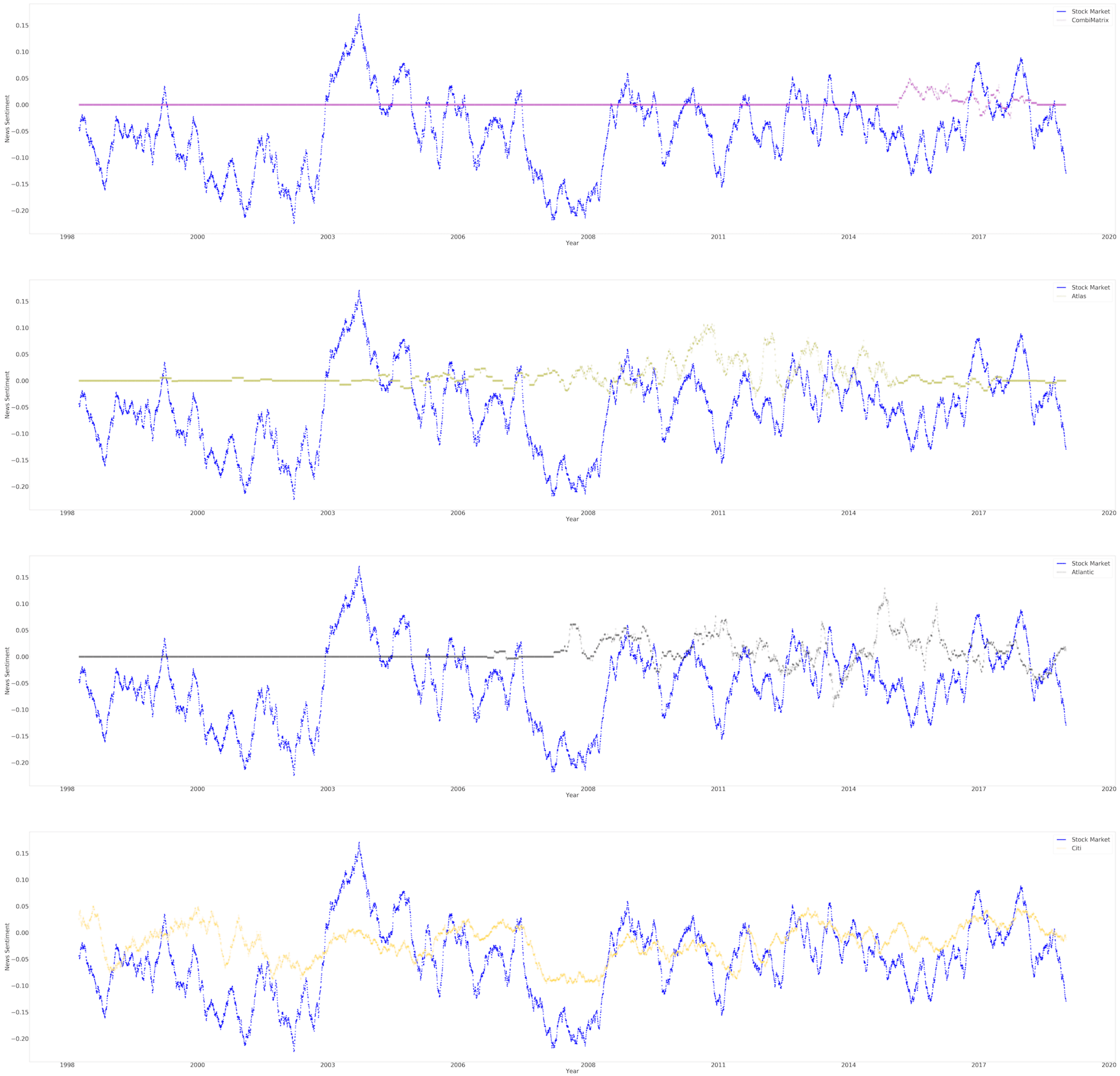


Figure 2.4: Stock Market News Sentiment vs. Firm News Sentiment II



All daily measures from TRMI are calculated from newsfeeds before 3:30 PM ET each day. The key variables used in the present study are: U.S. stock market news sentiment S_m as a proxy for biased tone in the market news received by investors; public company news sentiment S_i as a proxy of biased tone in the firm-specific news received by investors; and the sum (absolute value) of all relevant references to an asset extracted by the algorithm called *Buzz*. *Buzz* can be thought of as a measure of the intensity of media coverage. The higher the value of *Buzz*, the more the firm is discussed in news articles or online media.

Sentiment is calculated by taking overall positive references net of negative references to an asset or market:

$$Sentiment_j = \frac{Positive - Negative}{Total\ Buzz}, \quad \text{where} j \in \{m, i\} \tag{2.17}$$

where $Positive$ is the sum of the count of all $Positive$ terms and phrases, and $Negative$ is the sum of the count of all $Negative$ terms and phrases; $Total\ Buzz$ is the sum of total $Positive$ and $Negative$ counts of terms and phrases; m and i denote market and a particular firm respectively.

News sentiment varies daily, and the following empirical tests are based on firms' quarterly events and key financial variables from yearly data estimation. Therefore, this study follows the TRMI instruction in the user guide to aggregate the news sentiment into longer frequencies such as quarterly or yearly by $Buzz$-weighted average:

$$Sentiment_{j,T} = \frac{Buzz_{j,t} Sentiment_{j,t}}{\sum_t^T Buzz_{j,t}}, \quad \text{where} j \in \{m, i\} \tag{2.18}$$

Intuitively, the higher $Buzz$ at day t the more weight will be assigned to the sentiment at day t; as a result, sentiment values with high $Buzz$ are more influential in contributing to the mean sentiment in a particular period. In addition, sentiment measures from news released during weekends or U.S. Federal holidays are aggregated into the next trading day.[15]

Figure 2.3 and 2.4 plot time series firm news sentiment shown in different colors versus stock market news sentiment shown in blue. To visualize the noisy daily data, I smooth the sentiment indices by calculating the 90-day moving average for both the market and specific firms. On the one hand, the stock market news sentiment is an aggregated index that is calculated from all U.S. listed companies scaled by the $Buzz$. Specifically, if a company has positive news sentiment and contributes a large amount of texts measured by the Buzz, this company increases the overall stock market news

[15] In fact, the empirical study is not sensitive to how news is aggregated for non-trading days.

sentiment. On the other hand, I choose representative companies covering different industries. For example, Apple, General Electric, Walmart, Combimatrix, Atlas Pipeline Partners, Atlantic Power and Citigroup are from the technology, industrial, consumer, healthcare, energy, utilities and financial industries respectively. In addition, these exemplary companies include both large and small firms: large companies generate news almost every day, whereas small firms infrequently appear in the news, so the full spectrum of news frequency is covered.

First, all firms have a relatively higher news tone than the market. On the one hand, this is consistent with the findings in finance studies that report positive hype in firm news. For example, Jiang et al. (2019b) argue that managers tend to deliver a positive tone in corporate financial disclosures. Bajo and Raimondo (2017) explore the relation between positive news sentiment and the IPO under-pricing effect. Additionally, Gurun and Butler (2012) find that local media positively slants the local firms in news reports. On the other hand, the calculation of the market is aggregated by all companies, which is mathematically averaging down the overall market news sentiment in comparison to a particular firm. In fact, the average numerical market news sentiment is slightly negative, but positive for firms shown in Table 2.1 below. Second, the news sentiment of large firms such as Apple, GE, Walmart and Citigroup show clear co-movements with the market news sentiment. Unsurprisingly, these firms contribute large amounts of news to the entire market. Second, even though small companies do not generate news as frequently as the larger firms, small companies' news sentiment exhibit relative co-movement with the market. For example, Atlantic news sentiment shows approximate mimicking of the broader market's fluctuations. However, CombiMatrix shows movements that diverge from those of the market. A possible reason for this is that the company is in the healthcare industry, which accounts for a smaller market share of the U.S. stock market compared to other industries, such as consumer retail or finance and banking.

The TRMI contains about 4036 U.S. listed companies and the sample period is from 1998 to 2018 in this empirical study. Daily stock returns are taken from the Center for Research in Security Prices (CRSP) and financial fundamentals data are taken from the CRSP/Compustat merged database. I retain all U.S.-based common stocks with share code (SHRCD) value 10 or 11 listed on the NYSE, AMEX, and NASDAQ with exchange code (EXCHCD) 1 or 31, 2 or 32 and 3 or 33 respectively. I exclude stocks priced at less than $5 for consideration of illiquid stocks bias. Analyst forecast information is collected from the Institutional Brokers' Estimate System (IBES) and institutional ownership data are retrieved from Thomson Reuters Institutional

(13f) Holdings data file. I consider two measures as benchmark systematic uncertainty proxies: VIX and the Economic Policy Uncertainty Index (EPU) introduced by Baker et al. (2016). VIX data are obtained from the WRDS CBOE index, and EPU data are assembled from the Baker et al. (2016) research lab website. Additionally, the Generalized Probability of Informed (GPIN) Trading data from NYSE stocks are gathered from the Duarte et al. (2020) website. Finally, Fama–French asset pricing factors are downloaded from the Kenneth R. French - Data Library.

Panel A in Table 2.1 provides summary statistics of the key daily news variables and stock financial fundamental variables in the sample. *Buzz*, market value of equity, book value to market value and Amihud's (2002) illiquidity are positively skewed and are taken as the natural log to reduce positive skewness in the subsequent regressions. Sentiment ranges from -1 (most pessimistic) to 1 (most optimistic) with a score of 0 indicating perfectly neutral sentiment. The average of sentiment in stock market news is slightly negative in tone. However, the sentiment mean in the firm-specific news is slightly positively biased.[16] Notably, the firm-specific news sentiment is much more varied than the market news sentiment. The difference between the 75$^{\text{th}}$ and 25$^{\text{th}}$ percentile in the firm-specific news sentiment is about 0.5, which is almost twice as much as the spread of sentiment in the stock market news, which is 0.28 between the percentiles. Intuitively, this is not surprising because idiosyncratic news about a variety of companies from a wide variety of news reports should understandably be divergent when compared to news about the market, which is very standardized. Hence, the variety of firm-specific news has an anticipated large spread of biased tones.

Panel B in Table 2.1 shows the Pearson correlation between stock market news sentiment and systematic uncertainty. First, the systematic uncertainty measures, VIX and EPU, have the expected positive significant correlation and incorporate information to represent uncertainty in economic conditions. Second, the stock market sentiment from news has a significant negative correlation with both of the systematic uncertainty proxies, and this negative relationship is consistent with extant uncertainty studies in economics.[17] The negative correlation between stock market sentiment and the VIX is even more compelling, as it is approximately -0.32. More importantly, the

[16] The average positively biased tone in the firm-specific news is consistent with Berger and Milkman (2012) and Hirshleifer (2020), who assert that $E[b] > 0$ indicating media content is more likely to be positively than negatively biased.

[17] Chernenko et al. (2016) study investors' over-optimism in credit markets and under-perception of the downside risk - a combination that amplifies credit booms. Baker et al. (2016) find evidence of a negative correlation (-0.742) between their uncertainty index and the Michigan Consumer Sentiment index. Da et al. (2015) construct a FEARS index as a proxy for time varying parameter uncertainty to capture investors' pessimism about market recession.

TABLE 2.1: Summary Statistics and Correlations

This table presents summary statistics and correlations for sample variables. Panel A reports descriptive statistics used in empirical studies that test the impact of news sentiment on information acquisition and cross-sectional stocks returns. Panel B reports Pearson correlations (significant at the 1% level) between stock market news sentiment and proxies of economic uncertainty. Panel C reports Pearson correlations between firm-specific news sentiment and other financial fundamental variables. Detailed definitions of all variables are available in Appendix B.8.

Panel A Summary Statistics

	Mean	Std	Min	25%	50%	75%	Max	Count
$Buzz_{m,t}$	5408.554	5948.982	0.000	1230.225	3490.450	8008.125	123305.100	7670
$Sentiment_{m,t}$	-0.053	0.183	-0.870	-0.192	-0.049	0.085	0.714	7668
$Sentiment_{i,t}$	0.074	0.394	-1.000	-0.172	0.049	0.328	1.000	3458582
$Buzz_{i,t}$	223.457	1145.504	0.100	12.000	34.000	114.600	183978.300	3458582
VIX	20.208	8.500	9.140	13.885	18.540	24.035	80.860	5283
EPU	100.487	68.106	3.320	53.850	83.245	128.820	719.070	7670
$ME_{i,t}$	16701.942	43494.512	1.968	1038.649	3302.520	12375.910	867506.995	2867485
$BM_{i,t}$	0.553	2.690	0.000	0.244	0.430	0.708	359.622	2867485
$Illiquidity_{i,t}$	0.073	2.632	0.000	0.000	0.001	0.003	813.735	3451226
$OP_{i,t}$	0.457	14.556	-331.333	0.147	0.237	0.360	9423.750	2867105
$INV_{i,t}$	0.152	0.601	-0.933	-0.003	0.062	0.164	55.264	2797908
$RV_{i,t}$	0.026	0.016	0.001	0.016	0.021	0.030	1.019	3451228
$MOM_{i,t}$	0.153	0.624	-1.000	-0.109	0.100	0.318	98.571	3385220
$ST_{i,t}$	0.012	0.139	-1.000	-0.049	0.010	0.066	13.495	3451138
$AbRet_{i,t}$	0.001	0.030	-1.012	-0.009	0.000	0.009	6.979	2867485
$AbRet_{i,t-5,t-1}$	0.002	0.061	-1.077	-0.021	0.000	0.022	13.630	2867103
$AbTurn_{i,t}$	10.593	38.642	-4.304	2.905	5.707	11.337	25084.092	2867092

Panel B Systematic Variable Correlations

	Stock Market Sentiment	VIX	EPU
Stock Market Buzz	-0.154	0.012	0.079
Stock Market Sentiment		-0.319	-0.096
VIX			0.406

Panel C Idiosyncratic Variables Correlations

	$Buzz_{i,t}$	$ME_{i,t}$	$BM_{i,t}$	$Illiquidity_{i,t}$	$OP_{i,t}$	$INV_{i,t}$	$RV_{i,t}$	$MOM_{i,t}$	$ST_{i,t}$	$AbRet_t$	$AbRet_{t-5,t-1}$	$AbTurn_{i,t}$
$Sentiment_{i,t}$	-0.028	-0.048	-0.009	0.003	-0.002	-0.007	-0.046	0.045	0.047	0.079	0.070	-0.022
$Buzz_{i,t}$		0.427	-0.008	-0.004	0.002	0.004	-0.041	-0.002	-0.005	-0.001	-0.006	0.031
$ME_{i,t}$			-0.022	-0.007	0.011	0.001	-0.170	-0.039	-0.021	-0.006	-0.009	-0.045
$BM_{i,t}$				0.004	-0.004	-0.013	0.035	0.010	0.001	0.001	0.000	0.002
$Illiquidity_{i,t}$					0.000	-0.003	0.051	0.018	0.007	0.004	0.006	0.005
$OP_{i,t}$						-0.004	-0.019	-0.004	-0.002	-0.002	0.000	-0.004
$INV_{i,t}$							0.108	-0.005	-0.015	-0.003	-0.004	0.029
$RV_{i,t}$								0.008	0.049	0.022	0.045	0.156
$MOM_{i,t}$									0.009	0.001	0.000	0.034
$ST_{i,t}$										0.010	0.098	0.008
$AbRet_t$											-0.008	0.142
$AbRet_{t-5,t-1}$												0.070

negative relationship between stock market news sentiment and the systematic uncertainty measures confirms the assumption in the theoretical model that more positive sentiment in the market news ($S_m \uparrow$) biases investors to understate the uncertainty of market component ($\sigma^2_{b,m} \downarrow$) and vice versa.[18]

Finally, Panel C in Table 2.1 shows the Pearson correlation coefficients in stock level. In general, the correlation between sentiment and other variables does not yield a significant economic relationship. However, the *Buzz* measure is positively correlated with firm size and trading turnover, but negatively correlated with illiquidity. This evidence is consistent with existing textual studies,[19] which find that larger and more liquid firms tend to be better covered in the media and thus attract more investor attention. Therefore, the *Buzz* of both the stock market and firm-specific news are important controls for the news coverage (attention) effect in the subsequent empirical tests. Finally, since there is a very high negative correlation between the size variable and the illiquidity measure after taking natural logs, to alleviate the potential multicollinearity problem in the regression analysis, only one of them is included, usually the size, as one of the control variables.[20]

2.4 Empirical Results

By using this novel news dataset, I first validate the proposed channel of irrationality. This particularly applies to firm-specific news sentiment as the proxy for biased public information about firm-specific condition negatively predicting firm-specific uncertainty. Next, I conduct empirical tests to verify the theoretical results including the biased effect of investors' acquisition decisions about firm-specific information resulting from either market or firm-specific news sentiment. Lastly, I verify the proposition that the deviation of information risk leads to investors' requirement for a risk premium, which is in line with the cross-sectional variation of stock returns caused by firm-specific news sentiment.

2.4.1 Firm-Specific Uncertainty and Firm-Specific News Sentiment

As argued in section 2.2.5, investors' beliefs about the firm-specific uncertainty ($\sigma^2_{b,e}$) is biased by the sentiment in the firm-specific news. Therefore, it is important to

[18]In Appendix B.9.1, I also conduct a fixed effect regressions test to verify the negative relationship between market news sentiment and systematic uncertainty.

[19]For example, Fang and Peress (2009) argue that large firms are much more likely to be covered in the media. Engelberg and Parsons (2011) study the local media impact on local trading about S&P500 index firms.

[20]In fact, all the results are unchanged, regardless of size or illiquidity.

verify this theoretical presumption before showing the evidence of biased information acquisition.

I use three measures as proxies for uncertainty in the firm-specific component. I take companies' quarterly earnings to stand for e_1 in the theory model; thus the uncertainty about quarterly earnings per share (EPS) represents the firm-specific uncertainty. With a minor abuse of notation, in the following tests, I denote $\sigma^2_{e,t}$ as the proxy for firm-specific uncertainty with investors' rational perception when $S_e = 0$ in the firm-specific news. First, I start with a simple model to estimate the uncertainty of e_1 by following the time series of firm earnings in the accounting literature. Specifically, the non-Martingale process of firm quarterly earnings has been addressed by Griffin (1977), who proposes several models to illustrate how a stationary first-order autoregressive process can be found in the data. I assume that the firm's earnings follow a simple AR(1) process; therefore, the mean squared errors (MSE) from the regression model yield firm earnings uncertainty.[21] I then conduct the AR(1) regression for company quarterly earnings as follows:

$$\begin{aligned} EPS_{i,t+1} &= \gamma_0 + \gamma_1 EPS_{i,t} + \epsilon_{i,t} \\ \hat{\sigma}^2_{e,t} \text{ for firm } i &= \frac{\sum_{t=1}^{T} \epsilon^2_{i,t}}{T-2} \end{aligned} \tag{2.19}$$

For each firm, I conduct rolling regressions to estimate the $\hat{\sigma}^2_{b,e}$ as the first proxy of firm-specific uncertainty. I require companies to have at least 16 quarters of earnings for the estimation.

Second, the unexpected earnings (SUE) has been broadly addressed in the literature[22] and captures realized firms' fundamental performance. However, instead of using the traditional measure of SUE, I follow Hirshleifer et al. (2008) to measure the absolute value of SUE and use $Abs(SUE_{i,t})$ to identify the intensity of the seasonal random walk of unexpected earnings. Intuitively, the large SUE with a significant seasonal difference indicates a seasonal drift that is significantly different from zero between past earnings or expected earnings and future earnings. Accordingly, regardless of the seasonal difference being negative and positive, the greater the magnitude of $Abs(SUE_{i,t})$, the more difficult it is for investors to forecast either unexpectedly favorable or unfavorable company earnings using available information such as past earnings

[21]The higher the MSE from the regression, the more uncertain the forecast earnings from the model by assuming the AR(1) process. Additionally, this AR(1) process is also in the spirit of the theoretical model setting from the study of Veldkamp (2006).

[22]Livnat and Mendenhall (2006) review related studies of SUE in accounting and corporate finance literature.

or other forecasts.[23] Therefore, I first measure the unexpected earnings, SUE, following Livnat and Mendenhall (2006) as:

$$\begin{aligned} Compustat : SUE_{i,t} &= \frac{X_{i,t} - X_{i,t-4}}{P_{i,t}} \quad (a) \\ IBES : SUE_{i,t} &= \frac{X_{i,t} - E[X_{i,t}]}{P_{i,t}} \quad (b) \end{aligned} \tag{2.20}$$

where $SUE_{i,t}$ (a) is calculated by using Compustat quarterly earnings data while adjusting for stock splits on $X_{i,t-4}$ and $SUE_{i,t}$ (b) is calculated by using IBES investors forecast data for robustness purposes. The $E[X_{i,t}]$ is the most recent month's median earnings forecast by analysts for the quarter t. I then take the absolute value of each measure of $SUE_{i,t}$ as the second proxy of $\sigma^2_{e,t}$.

Importantly, Bali et al. (2018) develop a new measure of idiosyncratic volatility shock, arguing that such shock is more appropriate than the level of volatility in the identification of unusual news events. Instead of arguing for the utility of measuring unusual news flow, I investigate the relationship between news sentiment and idiosyncratic volatility shock as another proxy of firm-specific uncertainty. In fact, idiosyncratic volatility shock measures the difference between future idiosyncratic risk and expected idiosyncratic risk. Intuitively, where investors use expected idiosyncratic volatility (risk) to infer future firm idiosyncratic uncertainty (risk), increased or decreased certainty in the firm-specific component will yield a smaller or higher unexpected idiosyncratic volatility respectively. As a result, the more optimistically biased tone in firm-specific news predicts a lesser volatility shock. This is because positive news sentiment induces investors to believe there will be less idiosyncratic risk in the firm-specific business condition relative to their expectation. Following Bali et al. (2018), I estimate the idiosyncratic shock as:

$$\begin{aligned} R^e_{i,t} &= \alpha_i + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + \epsilon_{i,t}, \\ IVOL_{i,t} &= \sqrt{var(\epsilon_{i,t}) * \text{no. of trading days}} \end{aligned} \tag{2.21}$$

where $f_{i,m}$ is the benchmark pricing factor. I begin by estimating the Fama–French five factor and momentum factor model for each stock. I require a firm to have had at least 60 daily returns. I then conduct daily cross-sectional regressions for each firm to

[23] Karampatsas et al. (2018) measure firm-specific sentiment from social media and argue that it has a significant impact on stock price and negative earnings surprise. Their study shares a similar view with Tetlock's Tetlock (2007) study. However, my study focuses on the impact of firm-specific news sentiment on the magnitude of the earnings surprise regardless of whether it is positive or negative.

estimate the idiosyncratic shock as:

$$IVOL_{i,t} = \phi_{0,t} + \phi_{1,t}\overline{IVOL}_{i,t-1} + \sum_{j=1}^{10} \Phi_{j,t} D_{i,j} + v_{i,t} \tag{2.22}$$

where $IVOL_{i,t}$ from (2.21) and $\overline{IVOL}_{i,t-1}$ is the past average stock idiosyncratic volatility as investors' expectation about firms' idiosyncratic risk calculated by the moving average window between $t-24$ and $t-4$. $D_{i,j}$ is the 10 industry classifications dummy from Kenneth French's Data Library. Thus, the daily unexpected shock to idiosyncratic volatility is defined as: $IDIO^{shock}_{i,t} \equiv v_{i,t}$.

Finally, I use the three measures of firm-specific uncertainty to conduct the test as follows:

$$\begin{aligned} \hat{\sigma}^2_{e,t} &= \beta_0 + \beta_1 Sentiment_{i,[t-30,t-1]} + X\delta + \epsilon_{i,t} \qquad (a) \\ IDIO^{shock}_{i,t} &= \beta_0 + \beta_1 Sentiment_{i,t-1} + X\delta + \epsilon_{i,t} \qquad (b) \end{aligned} \tag{2.23}$$

where $\hat{\sigma}^2_{e,t}$ is the proxy from (2.19) or (2.20) as representing the firm-specific uncertainty. The model (a) in (2.23) is based on quarterly earnings data and the firm-specific news sentiment , $Sentiment_{i,[t-30,t-1]}$, is in the most recent month before quarter t calculated by the $Buzz_{i,t}$-weighted average as equation (2.18) from daily data. The model (b) is based on daily idiosyncratic volatility shock analysis. The X in both (a) and (b) is a vector of control variables (see Appendix B.8 for details) and δ as the coefficient vector. I use fixed effect regression for model (a) and daily Fama–Macbeth (1973) cross-sectional regressions for model (b) to test whether firm-specific news sentiment negatively predicts the proxy of firm-specific uncertainty and idiosyncratic shock respectively. In sum, the β_1 in both model (a) and (b) is expected to be both significant and negative.

Table 2.2 summarizes the regression results from models (a) and (b). All proxies of firm-specific uncertainty variables are winsorized at the 1% level to reduce the impact of extreme outliers. Additionally, I take the natural log of regression variance from equation (2.19) to reduce extreme positive skewness. Columns (1)-(3) are fixed effect regressions with standard errors clustered by firm and year-quarter. It should be noted that I use regression variance as the dependent variable in the model. Chen et al. (2018) use residuals as the dependent variable in the second step regression, and they argue that estimation of the interest explanatory variable (β_1 here) might be biased if the independent variable (sentiment) is correlated with the variables used in the first step regression. Therefore, it is necessary to include the independent variables used in the first step regression in the second step regression. I then include the $EPS_{i,t-1}$ as an additional control variable.[24]

First, column (1) clearly shows that firm-specific news sentiment negatively predicts the firm earnings AR(1) regression variance $\sigma^2_{e,t}$ estimated from equation (2.19). At an increase of two standard deviations of firm-specific news sentiment $(2 * 0.2871)$ the firm earnings which can not be explained by the AR(1) decreased by about 0.8%. This is strong evidence for the claim that a more optimistic tone in firm-specific news may induce investors to believe that quarterly company earnings are less uncertain by applying the AR(1) model to the forecast. Second, and unsurprisingly, columns (2) and (3) show that $Abs(SUE_{i,t})$ is significantly negatively predicted by sentiment in firm-specific news. An increase in firm-specific news sentiment by two standard deviations, $Abs(SUE_{i,t})$ decreases by about 0.6% and 4.6% of its mean value, respectively, to two measures of SUE. The more optimistic tone in the news causes investors to be more confident about expected earnings or about past earnings as a reliable forecast for future earnings either up or down; thus, they feel less uncertain about the company's earnings performance, and perceive less dispersion of unexpected earnings. The reverse is also true in relation to a more pessimistic tone in the news.

There is an intriguing finding that the IBES measured $Abs(SUE_{i,t})$ has much more economic significance - about 7.6 times larger than Compustat-measured $Abs(SUE_{i,t})$. The large impact that arises from applying IBES data is consistent with studies in the accounting and corporate finance literature, which argue that the analyst earnings forecasts are more likely to be subject to bias due to irrationality from optimism.[25]

Third, column (4) shows daily cross-sectional Fama–Macbeth (1973) regressions of $IDIO^{shock}_{i,t}$ on firm-specific news sentiment; standard errors are Newey-West corrected. The regression coefficient on firm-specific news sentiment shows consistent results with columns (1)-(3). The more optimistic tone in the daily firm-specific news leads investors to believe that their understanding of firm idiosyncratic risk is less uncertain. This negatively biased firm-specific uncertainty causes investors to perceive less future idiosyncratic risk in the firm, which results in them perceiving a lower value in the unexpected idiosyncratic volatility. As a consequence, a lesser idiosyncratic volatility shock is predicted where there is more positive sentiment in the firm-specific news and vice versa.

In sum, if we assume that the econometric model uses the correct fundamental variables which are widely considered to be rational or objective, then the model should be impartial in predicting future firm-specific uncertainty (risk) of earnings. However,

[24] By the same token, I also include $\overline{IVOL}_{i,t-1}$ in the model (b) from equation (2.23).

[25] See relevant studies by De Bondt and Thaler (1990), Abarbanell and Bernard (1992), and Easterwood and Nutt (1999), which argue that analysts are more likely to give optimistic forecasts.

Table 2.2: Firm-Specific News Sentiment and Firm-Specific Uncertainty

This table reports the results of regressions of proxies for firm-specific uncertainties on firm-specific news sentiment. Columns (1)–(3) are based on a quarterly data fixed effect regression model from equation (2.23)-(a): $\hat{\sigma}^2_{e,t} = \beta_0 + \beta_1 Sentiment_{i,[t-30,t-1]} + X\delta + \epsilon_{i,t}$ and column (4) conducts daily cross-sectional Fama–Macbeth (1973) regressions from equation (2.23)-(b): $IDIO^{shock}_{i,t} = \beta_0 + \beta_1 Sentiment_{i,t-1} + X\delta + \epsilon_{i,t}$. For regressions in column (1)–(3), I calculate $Sentiment_{i,[t-30,t-1]}$ as daily $Buzz_{i,t}$-weighted average in the last month before quarterly earnings announcements. Control variables include: lagged one period of dependent variable, forecast revision, forecast dispersion, size, book-to-market, turnover, return volatility, idiosyncratic volatility, absolute value of last month return, absolute value of cumulative abnormal returns, and institutional ownership for quarterly data regressions in columns (1)–(3). In addition, for the regression in column (1), an additional control EPS_{t-1} from the first step regression to estimate σ^2_e is also included. Control variables in in the column (4) regression include size, book to market, turnover, firm news $Buzz_{i,t-1}$, moving average of idiosyncratic volatility $\overline{IDIO}_{i,t-1}$ from the first step regression to estimate idiosyncratic volatility shock, operating profitability, firm investment, momentum return, return volatility and short term reversal return. Detailed definitions of all variables are available in Appendix B.8. Standard errors are clustered by both firm- and time-fixed effects in columns (1)–(3). Newey-West standard errors in column (4) are robust to heteroskedasticity and twelve days of autocorrelation. $***$,$**$,$*$ indicate statistical significance at the two-sided 1%,5%,10% levels, respectively.

	(1)	(2)	(3)	(4)
Dependent Variable	AR(1) $\hat{\sigma}^2_e$	$Abs(SUE_{i,t})$	$Abs(SUE^{IBES}_{i,t})$	$IDIO^{shock}_{i,t}$
$Sentiment_{i,[t-30,t-1]/t-1}$	−0.013***	−0.006***	−0.0004***	−0.0003***
	(0.004)	(0.002)	(0.000)	(0.0001)
$LagDep$	0.926***	0.379***	0.317***	0.967***
	(0.006)	(0.018)	(0.012)	(0.0005)
$ForecastRevision_{i,t-1}$	−0.027*	−0.024	−0.001	
	(0.014)	(0.032)	(0.002)	
$ForecastDispersion_{i,t-1}$	−0.046	0.463***	0.017***	
	(0.033)	(0.064)	(0.005)	
$LogME_{i,t-1}$	0.041***	−0.004***	−0.0004***	−0.0003***
	(0.003)	(0.001)	(0.000)	(0.0002)
$LogBM_{i,t}$	0.018***	0.010***	0.001***	−0.0003***
	(0.003)	(0.001)	(0.000)	(0.0004)
$LogTurn_{i,t-1}$	0.006*	−0.010***	−0.002***	
	(0.003)	(0.001)	(0.000)	
$ReturnVolatility_{i,t-31}$		0.201***	0.018***	
		(0.019)	(0.002)	
$Idiosyncratic\ Volatility_{i,t-31}$	11.897***	4.446***	0.356***	
	(2.415)	(1.667)	(0.123)	
$Abs(Return_{i,t-31})$	0.188***	0.010	0.002***	
	(0.041)	(0.013)	(0.001)	
$Abs(FFCAR_{i,t-30,t-3})$	−0.057**	0.006	0.002*	
	(0.027)	(0.010)	(0.001)	
$Abs(FFCAR_{i,t-2})$	−2.805	0.685	0.056	
	(2.694)	(0.534)	(0.097)	
$Institutional\ Ownership_{i,t-1}$	−0.00021***	-0.000	−0.000	
	(0.000)	(0.000)	(0.000)	
$EPS_{i,t-1}$	−0.0462***			
	(0.0035)			
$LogBuzz_{i,t-1}$				−0.00001
				(0.00002)
$\overline{IVOL}_{i,t-1}$				−0.001***
				(0.0001)
$OP_{i,t-1}$				−0.0004***
				(0.0001)
$INV_{i,t-1}$				0.0002***
				(0.0001)
$MOM_{i,t-1}$				−0.0004***
				(0.0001)
$RV_{i,t}$				0.03***
				(0.0115)
$ST_{i,t-1}$				−0.0023***
				(0.0003)
FE Firm	Yes	Yes	Yes	
FE Year-Quarter	Yes	Yes	Yes	
Fama-Macbeth				Yes
Constant				0.0035***
				(0.0004)
Observations	61,393	89,973	89,973	2,847,177
R-squared	0.925	0.234	0.155	0.939
Number of Firms	2,589	3,042	3,042	3,592

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

all three tests using either quarterly earnings data or daily idiosyncratic volatility data show strong evidence that by conditioning on biased tone in the firm-specific news, the sentiment negatively predicts every proxy of firm-specific uncertainty. Therefore, as investors read the news prior to making investment or trading decisions, their beliefs about future firm-specific uncertainty are biased either upward from negative sentiment or downward from positive sentiment in the news. This biased belief, caused by news sentiment transmits to the biased effect on investors' decision to acquire firm-specific information.

2.4.2 Firm-Specific Information Acquisition and News Sentiment

As the model predicts, the biased beliefs about uncertainty shift investors' acquisition of firm-specific information, component e_1, in comparison to the acquisition decision under rational expectations. However, investors' information acquisition, e_1, is not directly observed, so I conduct an event study, based on existing literature, of earnings announcements to test the inverse relationship between news sentiment and firm-specific information acquisition predicted by the theory.

I follow a novel measure of firm-specific information acquisition developed by Weller (2018) to estimate a jump ratio which is calculated within a certain window before and after companies' quarterly earnings announcements.[26] First, I define the pre-announcement window, as starting from 21 ($a = 21$) trading days before the announcement, as the period of identification of earnings-related information entering into the price before the announcement. Second, the identification of earnings-related information incorporated into prices when the earnings information is released spans two trading days ($b = 2$) after the announcement.[27] Based on the defined study windows, I first estimate the ACAR for both pre- and post-announcement as the price drift net of predicted returns from the Fama–French five-factor model. The momentum factor is also included:

$$CAR_{i,t}^{j_1,j_2} = \sum_{t=j_1}^{j_2} \left(R_{it}^e - \alpha_i - \sum_{m=1}^{M} \beta_{i,m} f_{m,t} \right) = \sum_{t=j_1}^{j_2} \epsilon_{i,t} \tag{2.24}$$

where $R_{i,t}^e$ is stock excess return and $f_{m,t}$ is the Fama-French and momentum factors. The α_i and $\beta_{i,m}$ is estimated by using 252 daily return data points and 90 days before

[26] In Appendix B.9.2, I use EDGAR SEC file searching volume to measure investors' acquisition of firm-specific information. The results are consistent with this price jump ratio proxy.

[27] The additional two days are to accommodate for the post-earnings announcement drift effect.

the earnings announcement. I require stocks that have observations on at least 63 trading days to estimate the factor model.

The jump ratio is estimated by using the post-announcement ACAR scaled by the total ACAR including before and after the earnings announcement as:

$$Jump_{i,t}^{a,b} = \frac{CAR_{i,t}^{T-1,T+b}}{CAR_{i,t}^{T-a,T+b}} \tag{2.25}$$

where $a = 21$ and $b = 2$ as the pre-announcement and post-announcement window respectively. As indicated in Weller (2018), the denominator $CAR_{i,t}^{T-a,T+b}$ may be close to zero. Therefore, I follow the instruction proposed by Weller (2018) to set up a threshold as $|CAR_{i,t}^{T-21,T+2}| > \sqrt{24}\hat{\sigma}_{i,t}$ where $\hat{\sigma}_{i,t}$ is daily return volatility during the 24-day event window.

Intuitively, if investors decide to acquire earnings-relevant information before the announcement day, the price incorporates more information about earnings. As a consequence, informed traders drive a greater price drift ($CAR_{i,t}^{T-a,T+b}$) than that which would be expected when earnings information becomes public. Conversely, if few investors decide to acquire information about firm earnings before the information is publicly revealed, on the announcement day the price drift jumps to incorporate the newly released earnings information once it becomes available. As in the model setting, if more informed investors conduct trading before the earnings announcement, the price is more informative and reflects earnings information (e_1 in the model) which can be partially gleaned by uninformed investors as well. Hence, when the earnings information becomes available, as price has reflected the earnings information before it is revealed, the price will not jump as much as in the case in which no or few investors are informed about the earnings before the information is released. Thus, the higher the price jump ratio, the less information is incorporated in the price (less information acquisition, e_1) relative to the post-announcement information set and vice versa (Weller, 2018). Therefore, aggressive and informed traders who trade before the earnings announcement drive the price jump close to 0, while an absence of informed trading drives the price jump towards 1. To test how news sentiment biases investors' firm-specific information acquisition, I conduct a fixed effect regression as follows:

$$jump_{i,t} = \beta_0 + \beta_{j,1} Sentiment_{j,[t-21,t-1]} + X\delta + \epsilon_{i,t}, \quad \text{where} j \in \{m, i\}. \tag{2.26}$$

where $Sentiment_{j,[t-21,t-1]}$ is the $Buzz_{i,t}$-weighted average news sentiment from 21 trading days up to 1 day before the earnings announcement. X is a vector of control variables (see detailed definitions in Appendix B.8) and δ as the coefficient vector. The theoretical model in Andrei et al. (2019) indicates that economic uncertainty in the fundamental payoff matters for investors' information acquisition decision. Therefore, I add customary systematic uncertainty measures, either VIX or EPU, as an additional control variable to identify the impact of news sentiment more clearly. From the theory model predictions in Corollary 1, the $\beta_{j,1}$ is expected to be positively significant to indicate a more positive or optimistic tone in market or firm-specific news, predicting a higher price jump which implies less information acquisition, e_1, and vice versa.

Panel A in Table 2.3 shows the results from equation (2.26) regarding the impact of stock market news sentiment on firm-specific information acquisition. First, the specification in column 1 only controls for month- and firm-fixed effects, and indicates that sentiment from stock market news strongly predicts positive price jumps. A one-unit increase in the optimism of market news sentiment causes the price jump to increase by 0.089 (relative to the median jump ratio of 0.3365). In line with firm-specific information acquisition interpretation, an increase in stock market news sentiment by one standard deviation (0.09) is associated with a 2.38% decrease in the proportion of earnings announcement-related price impact that arises pre-announcement.

Columns (2) and (3) in Panel A include additional controls to identify the impact of stock market news sentiment on firm-specific information acquisition. In addition to fundamental controls, I also add the $Buzz_{j,t}$ variable to control for a potential asymmetric information reduction effect as stated in Tetlock (2010). As he argues, public information from news can dissipate private information held by informed investors. $Buzz_{j,t}$ is the proxy of intensity of news coverage; therefore, based on the findings from Tetlock (2010), a higher value for $Buzz_{j,t}$ indicates there is more public information available to investors and may resolve information asymmetry. Because VIX and EPU may be strongly correlated, I control for each of the measures one at a time.

Including additional controls, the second specification in column (2) – the impact of market news sentiment on the jump ratio – shows very similar results. The VIX, as expected, is negatively significant in predicting the jump ratio, which is consistent with the model under rational expectations: as systematic uncertainty increases, firm-specific information acquisition increases (Andrei et al., 2019). Column (3) shows a slightly higher magnitude of impact from stock market news sentiment on information acquisition. The EPU index has an expected negative sign as consistent with the VIX implication, but is not statistically significant.

Panel B in Table 2.3 shows empirical results from equation (2.26) with respect to

Table 2.3: News Sentiment Impact on Information Acquisition

This table presents the results of regressions of the price jump ratio as the proxy for firm-specific information acquisition on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (2.26): $jump_{i,t} = \beta_0 + \beta_{j,1} Sentiment_{j,t-21,t-1} + X\delta + \epsilon_{i,t}$, where $j \in \{m, i\}$ and $jump_{i,t}$ is estimated as $CAR_{i,t}^{T-1,T+2}/CAR_{i,t}^{T-21,T+2}$ and $CAR_{i,t}^{T-a,T-b}$, the cumulative abnormal return is calculated from Fama–French five factor plus momentum factor model. The news sentiment variable $Sentiment_{j,[t-21,t-1]}$ and $Buzz_{j,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{j,t}$-weighted average in the study window. Control variables include: $Buzz_{j,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, economic uncertainty proxies (VIX and EPU) and the numbers of analyst coverage is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.8. Standard errors are clustered by both firm- and time- fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	Panel A Stock Market News Sentiment			*Panel B Firm-Specific News Sentiment*		
	(1)	(2)	(3)	(1)	(2)	(3)
Dependent Variable	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$
$Sentiment_{i,t-21,t-1}$				0.057***	0.050**	0.051***
				(0.015)	(0.020)	(0.020)
$Sentiment_{m,t-21,t-1}$	0.089***	0.091***	0.116***		0.120*	0.124**
	(0.010)	(0.011)	(0.011)		(0.063)	(0.063)
$Buzz_{m,t-21,t-1}$		0.027***	0.029***			
		(0.002)	(0.002)			
$Buzz_{i,t-21,t-1}$					-0.013***	-0.012**
					(0.005)	(0.005)
$VIX_{t-21,t-1}$		-0.002***			-0.001	
		(0.000)			(0.001)	
$EPU_{t-21,t-1}$			-0.0002			-0.0001
			(0.000)			(0.0001)
$Size_{i,t-42,t-21}$		0.006**	0.007***		0.040***	0.041***
		(0.002)	(0.002)		(0.016)	(0.015)
$Turn_{i,t-42,t-21}$		0.002	0.003**		0.018	0.019
		(0.001)	(0.001)		(0.012)	(0.012)
$Price_{i,t-42,t-21}$		-0.011***	-0.010***		-0.032**	-0.034**
		(0.003)	(0.003)		(0.016)	(0.016)
$RV_{i,t-42,t-21}$		-0.021***	-0.028***		-0.042**	-0.047***
		(0.002)	(0.002)		(0.017)	(0.015)
$NUMEST_{i,t-21,t-1}$		0.001***	0.001***		0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$		0.056***	0.053***		0.050	0.048
		(0.007)	(0.007)		(0.036)	(0.036)
FE Month	Yes	Yes	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93,198	91,873	91,873	3,550	3,521	3,521
R-squared		0.021	0.020		0.033	0.033
Number of Firms	10,329	10,241	10,241	1,891	1,880	1,880

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the impact of firm-specific news sentiment on information acquisition about e_1. Column (1) is the specification-only controls for month- and firm-fixed effects. The biased tone in firm-specific news shows significantly positive predictive power on the jump ratio. With a one-unit increase in firm-specific news sentiment, the price jump ratio increases by 0.057 (relative to the median value of jump ratio 0.393). With regard to firm-specific information acquisition, an increase in the optimism of firm-specific news sentiment by one standard deviation (0.31) causes investors to acquire less earnings-related information by 4.5% before the earnings announcement. Columns (2) and (3) are specifications including additional controls. I also control stock market news sentiment in specification 2 and 3. In fact, the magnitude of economic significance from firm-specific news sentiment is not compromised after adding additional control variables. Although stock market news sentiment maintains its explanatory influence on the jump ratio, it is promising that bias in the firm-specific news is inversely related to investors' firm-specific information acquisition. Moreover, there are intriguing findings between Panel A and B in Table 2.3. The two $Buzz_{j,t}$ controls, stock market and firm-specific news, have entirely opposite impacts on the jump ratio. In fact, more $Buzz_{i,t}$ from firm-specific news significantly decreases the jump ratio, which implies an increase in firm-specific information acquisition before the earnings announcement. However, in the case of stock market news, $Buzz_{m,t}$ has the reverse effect. This intriguing evidence is consistent with the key argument of Tetlock (2007; 2008; 2010) that market news sentiment does not contain value-relevant information, unlike firm-specific news sentiment regarding firms' fundamentals.

2.5 Information Risk from Firm-Specific News Sentiment

2.5.1 Probability of Informed Trading and Firm-Specific News Sentiment

As the model proposed that information risk is affected by variations in the proportion of informed investors as a result of firm-specific news sentiment, I investigate this proposition by testing the relationship between the probability of informed trading (PIN) developed by Easley et al. (1996) and news sentiment from particular firms. The PIN has been empirically tested as a proxy of information risk and the risk premium can be found in cross-sectional asset returns.[28]

Intuitively, as more investors choose to become informed about e_1 and trade in the market, the equilibrium price becomes more informative and is of more utility to uninformed investors. Thus, to hold the indifference condition as proposed by Grossman and Stiglitz (1980), informed investors are not willing to trade aggressively by submitting additional more-informed orders (i.e. buying when asset value is high and selling when asset value is low). In fact, when price is informative, there is a high proportion of informed traders in the market, which leads to a decline in the knowledge disparity between informed and uninformed investors. Correspondingly, submitting more informed orders does not contribute extra benefits to informed investors, since they do not want uninformed investors to gain a free ride by learning from the equilibrium price, which is itself an increment of uninformed utility. As a consequence, a

[28] For a comprehensive study and review, see Duarte et al. (2020). See studies by Easley et al. (2002); Easley and O'hara (2004) and Easley et al. (2010) for information risk premium implied by PIN.

reduction in aggressively informed orders submitted by informed traders decreases the order arrival rate of informed traders, the μ in the PIN model. Therefore, as informed order arrival rate decreases, the PIN value decreases and there is less information risk in the asset.

As stated in Corollary 2 and 3, a more optimistic tone in firm-specific news decreases firm-specific information demand by investors and results in more information risk in the biased belief equilibrium and vice versa. Therefore, following the literature that argues that PIN can be seen as a proxy for information risk, I conduct a hypothesis test on the relationship between PIN and sentiment from firm-specific news. However, the traditional measure of PIN is subject to bias, which is that it cannot match a large amount of variation in turnover initiated by noise trade (Duarte et al., 2020).[29] Hence, in considering the limitations of the PIN model, which may result in inaccurate statistical inference, I use Generalized Probability of Informed Trade (GPIN) from Duarte et al. (2020) as an information risk proxy. It should be noted that GPIN data are only available for NYSE stocks. Consequently, the empirical results are intended to be very conservative and understate the impact of news sentiment on information risk, because companies traded on NYSE are, in general, large liquid stocks presumed to have fewer information asymmetry problems. I conduct the fixed effect regression as :

$$GPIN_{i,t} = \beta_0 + \beta_1 Sentiment_{i,t-1} + X\delta + \epsilon_{i,t}, \tag{2.27}$$

where $GPIN_{i,t}$ is the stocks' generalized probability of informed trade, estimated with year t daily trade tick data, and $Sentiment_{i,t-1}$ is $Buzz_{i,t}$-weighted average firm-specific news sentiment in year $t-1$. The X includes a bundle of control variables (see Appendix B.8 for details) and δ as the coefficient vector. Since the news data begins in 1998, the regression starts from 1999. The reason I use a lagged sentiment variable as the explanatory variable is due to a concern about potential inverse causality in contemporaneous periods. More specifically, since firm-specific news comes randomly throughout the year, a contemporaneous regression cannot be guaranteed to be free of endogenous issues about companies' information asymmetry, which may potentially affect sentiment in the firm-specific news. All in all, I expect a positively significant β_1 in equation (2.27), indicating that positive firm-specific news sentiment predicts high information risk.

[29] Duarte et al. (2020) state that the implied variability of buys and sells from the PIN model, in general, is 550 times smaller than the realized variability in the data. The biased estimation from PIN derives from the failure of the model to capture large amounts of variability in noise trading.

TABLE 2.4: Firm-Specific News Sentiment Impact on Probability of Informed Trading

This table presents the results of regressions of Generalised Probability of Informed Trading (GPIN) as a proxy for information risk for all stocks listed on the NYSE. Columns (1)–(3) are based on fixed-effect regression from equation (2.27): $GPIN_{i,t} = \beta_0 + \beta_1 Sentiment_{i,t-1} + X\delta + \epsilon_{i,t}$. The GPIN is estimated yearly and the regression starts from 1999 to 2018. News sentiment from either firm-specific news or stock market news is the $Buzz$-weighted average within a year. Control Variables include : $Buzz_{j,t-1}$ where $j \in \{m, i\}$ proxies coverage about firm-specific and stock market news, Size, Book to Market, Trading Volume, Idiosyncratic Volatility and Institutional Ownership. All independent variables are lagged for one year. Detailed definitions of all variables are available in Appendix B.8. Standard errors are clustered by both firm- and time-fixed effect in columns (1)–(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively.

Dependent Variable	(1) $GPIN_{i,t}$	(2) $GPIN_{i,t}$	(3) $GPIN_{i,t}$
$Sentiment_{i,t-1}$	0.017***	0.014***	0.014***
	(0.005)	(0.005)	(0.005)
$Sentiment_{m,t-1}$			0.118
			(0.087)
$Buzz_{m,t-1}$			-0.022
			(0.014)
$Buzz_{i,t-1}$		-0.002	-0.002
		(0.001)	(0.001)
$ME_{i,t-1}$		-0.003	-0.003
		(0.002)	(0.002)
$BM_{i,t-1}$		0.001	0.001
		(0.002)	(0.002)
$Turn_{i,t-1}$		-0.0019	-0.0018
		(0.002)	(0.002)
$IDIOVOL_{i,t-1}$		0.012	0.012
		(0.009)	(0.009)
$ITOW_{i,t-1}$		0.015**	0.015**
		(0.006)	(0.006)
FE Year	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes
Observations	15,571	13,551	13,551
R-squared	0.150	0.148	0.148
Number of Firms	1,434	1,355	1,355

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.4 presents results from equation (2.27). The specification in column (1) only controls for year- and firm-fixed effects, and it confirms that biased tones in firm-specific news predict positive GPIN. With a one-unit increase in the optimism of sentiment, information risk as measured by GPIN increases by approximately 0.017 relative to its mean value (0.26). By adding more controls in columns (2) and (3), which are also variables with considerable explanatory power in respect to information asymmetry, it still maintains strong positive significance in explaining the variation of GPIN. In fact, a two-standard deviation (0.18) increase in the positivity of news sentiment concerning a particular firm increases the GPIN by about 2%. This indicates that the buy or sell orders are 2% more likely to be from informed traders who hold private information about the risky asset. Therefore, information risk in risky assets increases as tone, reported in the firm-specific news, becomes more optimistic. This high information risk caused by positive news sentiment implies a reduction of the benefit of price informativeness gained by the uninformed traders to alleviate the information asymmetry risk trading against informed traders and vice versa.

O'Hara (2003) proposes that information asymmetry existed in risky assets as the disparity in the information held by informed and uninformed investors. This information risk is perceived by traders who require compensation to hold risky assets. As shown by the results in Table 2.4, sentiment in the firm-specific news affects information risk as measured by GPIN in risky assets. Next, I investigate the variation in cross-sectional asset returns (the risk premium) using the deviation in information risk caused by biased tones in the firm-specific news through the biased information acquisition in equilibrium.

2.5.2 Firm-Specific News Sentiment Impact on Cross-Sectional Stock Returns

In the model in section 2.2, Corollary 2 and 3 indicate a monotonic relationship between firm-specific news sentiment and expected returns on the risky assets. This reflects the variation in information risk. Hence, to evaluate whether firm-specific news sentiment induces a deviation in information risk, which causes variation in the expected returns of assets, I examine whether sentiment from firm-level news on day t predicts positive firm excess return on $t+1$. In addition, since I argue that this predictability of cross-sectional returns stems from the investors' biased firm-specific information acquisition in equilibrium rather than from the mispricing of fundamental value, I expect that the positive predictability is not reversed and is persistent in cumulative returns for the subsequent trading days. Therefore, I conduct tests on cumulative returns up to 5 and 10 days after the firm-specific news is released.

The dependent variable in the regression model is day $t+1$ stock excess return (R^e_{t+1}) and either 5 or 10 days' worth of cumulative returns ($R^e_{t+2,t+5/10}$), where the day $t+1$ return is omitted from cumulative return as a consideration of bid-ask bounce. The control variables, all firm characteristics that affect the predictability of expected returns, include measures of company size[30] ($Size_{i,t}$), book to market ratio ($BM_{i,t}$), operating profitability ($OP_{i,t}$), investment ($IVN_{i,t}$), yearly return momentum ($MOM_{i,t}$) excluding the most recent month, the last month return volatility ($RV_{i,t}$) and the last month return ($ST_{i,t}$) as short-term reversal effects. To consider return reversal predictability,[31] I add day t abnormal return $AbRet_{i,t}$ defined as the raw return minus the value-weighted market return from CRSP and cumulative abnormal returns from the

[30] Since Amihud (2002) illiquidity measure is highly negatively correlated with the size measure about -0.92, I use both of them one at a time and the results are unchanged.

[31] See related studies by Roll (1984); Jegadeesh (1990); and Lehmann (1990).

past five trading days ($AbRet_{t-5,t-1}$). As demonstrated in the model of Llorente et al. (2002), if stock trading volumes are aligned with daily returns, this strongly predicts future returns.[32] Hence, additional controls regarding the trading volume effect include firms' abnormal trading volume ($AbTurn_{i,t}$), defined as log turnover on trading day t net of its average of log turnover from $t-5$ to $t-1$ and the interaction between day t abnormal return and trading volume ($AbRet_{i,t} * AbTurn_{i,t}$).

The main test is on the firm-specific news sentiment on day t ($Sentiment_{i,t}$) to predict day $t+1$ or cumulative returns in the following days. There are two major concerns in the identification of the effect of firm-specific news sentiment on information risk premium. First, Tetlock (2010) proposes that public information from news resolves information asymmetry by testing the reduction of return reversal and volume-induced return on firm news days. Because of the definition of 'sentiment,'- which I argue is the tone of public information in the news, tending to induce a deviation in the proportion of information asymmetry in risky assets - it is necessary to control for the impact of news on resolving asymmetric information as stated by Tetlock (2010). Therefore, I use $Buzz_{i,t}$ as a proxy for the intensity of firm-specific news coverage interacting with firm-abnormal returns on day t. The rationale for controlling $Buzz_{i,t}$ is that, as a company is widely discussed in the news or there is more public information available to uninformed investors, it is easier for uninformed investors to infer superior information about the firm from the news and become less reluctant to provide liquidity to informed investors (Tetlock, 2010). As a consequence, when investors have more relevant public information about a firm on day t, the abnormal return at day t is conditional on the availability of firm-level news information, and is expected to lead return momentum as liquidity shock is dissipated gradually after the news is released (Tetlock, 2010).

Second, there is a growing number of studies[33] using textual data to assess the effect of sentiment in firm-specific news or online media platforms containing value-relevant information about companies' fundamentals. For example, Tetlock et al. (2008) and Chen et al. (2014) argue that negative sentiment in the firm-level news or media is instructive to investors regarding unfavorable earnings information from companies. Additionally, in a recent study by Aboody et al. (2018) measures firm-specific investor sentiment with overnight return and argues for the existence of an inverse relation with subsequent returns. However, the predictability of firm-specific

[32] See related studies of trading volume impact on return predictability by Campbell et al. (1993) and Lee and Swaminathan (2000).

[33] Comprehensive survey studies can be found in Tetlock (2014b) and Loughran and McDonald (2016).

news sentiment, as argued in this research, mainly relates to the risk premium of information asymmetry, which is distinct from the argument regarding genuine information in extant studies. In fact, the predictability of the effect of news sentiment impact on cross-sectional stock returns, which is the main relationship evaluated in this study, is in addition to the predictability found in the growing literature.

Thus, a thorough consideration of the genuine information effect is necessary, as an essential control to conduct a return predictability test in the subsequent main regressions. If the genuine information effect dominates predictability from the news sentiment, the empirical results would not show a significant predictive power from the firm-specific news sentiment after controlling for the genuine information effect. Therefore, it is indispensable to disentangle the effect of genuine information contained in the firm-level news sentiment from the sentiment variable ($Sentiment_{i,t}$) for each firm. There is a valuable measure provided by TRMI: it is $EmotionVsFact_{i,t}$ and ranges from -1 to 1. This index measures the proportion of emotional references net of the factual reference from news articles. The emotional reference counts subjective words in the news article such as people's expressed opinions or feelings about the news stories. The factual reference counts objective words or fundamental firm information from the news stories, such as content related to operation, earnings, merging or accounting (see Appendix B.8 for details).

Intuitively, the closer to 1 in the $EmotionVsFact_{i,t}$ measure is, the more subjective opinion there is in the news story about a company. Conversely, the closer to -1 in the $EmotionVsFact_{i,t}$ measure is, the more factual, objective or fundamental material is in the news story about a company. In line with the evidence of firm-specific news sentiment containing genuine information about firms' fundamental payoff, we should expect the more factual (lower number of $EmotionVsFact_{i,t}$) reference to interact with $Sentiment_{i,t}$ to predict positive stock future returns.[34]

I therefore interact $EmotionVsFact_{i,t}$ and $Sentiment_{i,t}$ to control for the potentially genuine information contained in $Sentiment_{i,t}$. Furthermore, I add an interaction between $EmotionVsFact_{i,t}$ and $AbRet_{i,t}$ as another control for the effect of news resolving information asymmetry. For example, the greater the proportion of factual information in the company news, as investors read the news, the more likely they are to infer the private information from factual information in the news and vice versa.[35]

[34]For example, a negative number of $EmotionVsFact_{i,t}$ and a negative sentiment indicate negative value-relevant information for the firm fundamentals and vice versa.

[35]Intuitively, if the news contains more factual information about firm fundamentals, there should be a reduction on daily return reversal. In other words, one would expect a return momentum as more factual information in the news is reported in day t.

Finally, all independent variables are standardized by each day before computing interaction terms for easy interpretation. I require at least 100 firms to have some news and non-missing independent variables each day.[36] For all firms with news sentiment, I estimate daily cross-sectional Fama–Macbeth (1973) regressions to evaluate whether the positively biased tone in the news predicts future returns either on day $t+1$ or the cumulative return in the following days. The cross-section regression specifications are:

$$DepVar_{i,t+1} = \beta_0 + \beta_1 Sentiment_{i,t} + \delta X + \epsilon_{i,t} \text{ for each trading day } t. \tag{2.28}$$

where $DepVar_{i,t+1}$ is either R^e_{t+1} or $R^e_{t+2,t+5/10}$ and the X is a vector of control variables and δ as the coefficient vector. The main purpose of this test is to determine whether β_1 is significantly different from 0. More importantly, as per the theoretical predictions argued in section 2.2, it is expected that β_1 will have a positive value, indicating that a more optimistic tone in the news will bring about a higher return, as investors expect to be compensated for higher information risk in the risky asset and vice versa.

Column (1) in Table 2.5 is the results of day $t+1$ return prediction. As all independent variables are standardized, the regression coefficients interpret a change in the dependent variable as a change of one standard deviation on the predictors. Notably, the firm-specific news sentiment ($Sentiment_{i,t}$) at day t significantly predicts positive stock return on the next day, even after controlling for other important effects implied by news information. With an increase in firm-specific news sentiment by one standard deviation, the next day's return increases by about 3.1 basis points, which is equivalent to a 0.65% monthly return. Surprisingly, this increment in the next day return is economically significant; in fact, R^e_{t+1} increases by approximately 110% relative to its mean (2.83 basis points) in the sample period. More precisely, I calculate the marginal effect of sentiment by netting the effect of predictability from genuine information within sentiment, which is controlled by the interacted variable $EmotionVsFact_{i,t} * Sentiment_{i,t}$. Its regression coefficient is consistent with the literature. For instance, sentiment extracted from subjective references in the news causes a reversal prediction. On the contrary, sentiment about factual or fundamental references in the firm-specific news positively predicts future returns. This evidence provides an important contribution to the debate in the behavioral finance literature

[36] Because the regression model contains about 20 predictors, the minimum observation is a consideration of sufficient degrees of freedom and the statistical power of the tests. However, the results are insensitive to this requirement.

TABLE 2.5: Cross-Sectional Return Predictability from Firm-Specific News Sentiment

This table presents the results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$. Variables measured by news content and all other control variables are known by day t. Columns (1)–(3) report the time-series average of the coefficients based on the model in equation (2.28): $DepVar_{i,t+1} = \beta_0 + \beta_1 Sentiment_{i,t} + \delta X + \epsilon_{i,t}$ for each trading day t, where $DepVar_{i,t+1}$ is R^e_{t+1}, $R^e_{t+2,t+5}$, and $R^e_{t+2,t+10}$, respectively. The variable $Sentiment_{i,t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i,t} * Sentiment_{i,t}$, $EmotionVsFact_{i,t} * AbRet_{i,t}$, and $Buzz_{i,t} * AbRet_{i,t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals from $Sentiment_{i,t}$. Additionally, abnormal return $AbRet_{i,t}$ at day t and its related interactions such as $AbRet_{i,t} * Size_{i,t}$ and $AbRet_{i,t} * AbTurn_{i,t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the last five days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.8. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

	(1)	(2)	(3)
Dependent Variable	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	3.089	3.764	4.341
	(8.188)	(5.084)	(3.799)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-2.673	-1.743	-1.966
	(-5.109)	(-1.743)	(-1.326)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-3.908	1.459	-1.311
	(-3.302)	(0.799)	(-0.484)
$Buzz_{i,t} * AbRet_{i,t}$	3.505	3.895	5.850
	(7.019)	(5.318)	(5.296)
$Buzz_{i,t} * ME_{i,t}$	-0.034	0.755	0.700
	(-0.122)	(1.200)	(0.635)
$Buzz_{i,t}$	-0.180	-0.279	0.869
	(-0.643)	(-0.459)	(0.866)
$EmotionVsFact_{i,t}$	-0.348	-2.006	-0.715
	(-0.543)	(-1.604)	(-0.376)
$AbRet_{i,t}$	-3.383	-6.727	-6.839
	(-4.533)	(-5.006)	(-3.643)
$ME_{i,t}$	-1.550	-5.509	-12.437
	(-3.189)	(-3.692)	(-4.279)
$BM_{i,t}$	-0.541	-2.279	-2.959
	(-1.094)	(-1.545)	(-1.023)
$OP_{i,t}$	0.014	0.348	0.619
	(0.038)	(0.359)	(0.349)
$IVN_{i,t}$	0.017	-2.278	-4.974
	(0.052)	(-2.329)	(-2.638)
$RV_{i,t}$	-0.086	-0.494	-1.081
	(-0.118)	(-0.200)	(-0.217)
$MOM_{i,t}$	-0.534	1.218	3.208
	(-0.904)	(0.673)	(0.888)
$ST_{i,t}$	-0.844	-1.706	-3.518
	(-1.569)	(-1.119)	(-1.220)
$AbRet_{i,t} * Size_{i,t}$	-2.774	-6.171	-8.232
	(-5.862)	(-7.780)	(-7.060)
$AbTurn_{i,t}$	-5.638	-1.079	-4.525
	(-4.302)	(-0.484)	(-1.251)
$AbRet_{i,t-5,t-1}$	-2.710	-4.787	-5.356
	(-4.651)	(-3.749)	(-2.525)
$AbRet_{i,t} * AbTurn_{i,t}$	0.324	-0.708	-1.309
	(1.128)	(-1.441)	(-1.907)
$Constant$	3.395	15.841	33.927
	(1.939)	(2.462)	(2.561)
Daily Average Observations	540	540	539
Adjusted R-squared	0.141	0.133	0.129
Observations	2,842,780	2,840,509	2,838,805

by using textual analysis of whether sentiment is a form of bias affecting investors' valuation of an asset or contains genuine information about the firm fundamentals. Therefore, the marginal effect of sentiment predictability on $t+1$ return is about 0.42 basis point (representing a 15% increase from its mean), increasing on the next day return in line with a one standard deviation increase in sentiment and net of the effect of one standard deviation in factual reference in the firm-specific news $(3.1*1-2.67*1)$.[37]

By disentangling the effect of news sentiment that may cause investors either to estimate firms' value in a biased way or to obtain firm value-relevant information, the sentiment maintains significant positive predictability on firm future returns. This effect is both statistically and economically significant. Hence this additional cross-sectional return predictability implies variation in information risk through firm-specific news sentiment, thus causing firm-specific information acquisition to deviate from the rational expectations equilibrium. Additionally, the control variables $Buzz_{i,t} * AbRet_{i,t}$, and $EmotionVsFact_{i,t} * AbRet_{i,t}$ are all consistent with the findings in the literature on capturing the effects of asymmetric information mitigation.[38]

Columns (2) and (3) are results regarding cumulative returns at 5 and 10 days respectively after firm-specific news is released. The firm-specific news sentiment remains positively significant on 5- and 10-day cumulative returns. With a one standard deviation increase in the optimism of news sentiment, the 5- or 10-day cumulative returns increase by about 26.83% and 13.15% respectively relative to their mean values (14.03 and 33.01 basis points). However, the control variable $EmotionVsFact_{i,t}*Sentiment_{i,t}$ is no longer significant. The insignificance of $EmotionVsFact_{i,t} * Sentiment_{i,t}$ on cumulative returns infers that the market is efficient as one-day turnaround to either incorporate valuable information about firms from news or to correct the mispricing resulting from investors' irrational response to public news containing more subjective references.

[37] A one standard deviation increase on factual references is -1, in line with one standard deviation increase on sentiment of +1. Therefore, the predictability is contributed by genuine information from news sentiment is 2.67 $(-2.67*-1*1)$. Thus, the net effect is calculated by subtracting 2.67 from 3.1.

[38] For example, the positive significant coefficient on $Buzz_{i,t}*AbRet_{i,t}$ is consistent with the argument of Tetlock (2010) that information from news reduces daily return reversal, just as the release of public information in the news resolves information asymmetry in the assets. There is an intriguing finding on $EmotionVsFact_{i,t}*AbRet_{i,t}$ with a negative significant effect on the next day's returns. Intuitively, the negative significant effect on $EmotionVsFact_{i,t}*AbRet_{i,t}$ captures both effects about the potential value-relevant information contained in the firm-specific news and investors' irrational response to news information. As news articles contain more subjective reference information, this could bias investors' valuation on firm fundamental payoffs,causing negative serial correlation of returns and vice versa. See studies related to biased valuation of sentiment by Baker and Wurgler (2006) and Baker et al. (2012). Additionally, Holden and Subrahmanyam (2002) stress that as information asymmetry is dissipated through public information from news, there is a positive serial correlation in stock returns.

More importantly, the empirical evidence of persistent return predictability from firm-specific news sentiment is similar to that found in information diffusion studies, although the rationale is somewhat distinct. In general, information diffusion studies such as that of Hong and Stein (1999) argue that boundedly rational investors cause gradual information diffusion and their simple trading strategy causes short-run momentum and long-run overreaction on returns. However, in my model, the key issue is that public information from news is biased in the way in which it is reported; thus, the bias stems from the news supplier, not from the investors themselves, particularly their assumed irrationality.

The consequence is that, to some extent, investors are forced into being biased in their beliefs about the uncertainty of fundamental payoff, due to their being unduly influenced by the firm-specific news. Investors form a biased belief equilibrium regarding firm-specific information acquisition, e_1. In sum, as long as the tone of news information is biased (either positively or negatively), there is always a deviation in the acquisition of firm-specific information in equilibrium. This implies either higher or lower information risk in the asset compared to the rational expectations model. As proposed by O'Hara (2003), traders require compensation to hold risky assets containing more information risk, which, in my study, varies with sentiment in firm-specific news.

2.5.3 Firm-Specific News Sentiment Portfolio Analysis

The cross-sectional variation of stock returns predicted by firm-specific news with optimistic and pessimistic tones suggests news sentiment may result in variation in information risk across assets. Therefore, I conduct a portfolio formation analysis sorted by daily firm-specific news sentiment by following Fama and French (1992) to verify whether the risk premium of information risk triggered by firm-specific news sentiment can be captured by traditional asset pricing factors.

At the end of each trading day (3:30 PM EST), I first use monthly NYSE breakpoints of the last month median market capitalization from the Kenneth R. French Data Lab to split stocks into two portfolio sizes: small (S) and large (B). Independently, I rank stocks based on day t firm-specific news sentiment into three sentiment portfolios: pessimistic (N), neutral (M), optimistic (P). Stocks within the lowest 30^{th} percentile (N) have more negative sentiment (pessimistic tone) from their daily news articles; stocks within the highest 30^{th} percentile (P) have more positive sentiment (optimistic tone) in the daily news stories; and the stocks within the middle 40% (M) contain relatively neutral tones in the news discussion. The six interacted portfolios, value-weighted with respect to size and firm-specific news sentiment, are: $N/S; N/B; M/S; M/B; P/S;$ and P/B, sorted by portfolio size and news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios ($P/S; P/B$) and the average of short position in the two negative sentiment portfolios ($N/S; N/B$) each day and I calculate the next day ($t + 1$) value-weighted portfolio returns from this zero-cost trading strategy.

The firm-specific news sentiment zero-cost portfolio generates significant positive average daily returns of about 6.6 basis points (which equates to a 16.63% annualized return) shown in Table 2.6. It should be noted that there is a concern that illiquidity could play a role in information risk, and that could itself explain the news sentiment pricing effect, as this study proposes that it can be seen as a trigger of information risk across assets. I calculate daily Pastor and Stambaugh liquidity factors (PSLIQ) by following Pástor and Stambaugh (2003)[39] as an additional important pricing factor to test the pricing capability of news sentiment. Panel A in Table 2.6 is the Pearson correlations between the firm-specific news sentiment factor and other customary pricing factors. In fact, the news sentiment factor has a weak negative correlation with the market factor (-0.106), the value factor (-0.102), and the short-term reversal factor (-0.142), and a weak positive correlation with the momentum factor (0.202). The remaining factors have correlations with the news sentiment factor roughly close to zero. Next, I investigate whether these existing factors can explain the abnormal return from the pricing factor constructed by daily firm-specific news sentiment.

Panel B in Table 2.6 shows the risk-adjusted daily returns from the zero-cost portfolio based on a trading strategy informed by news sentiment. I use the CAPM, Fama–French three factors (1993), and Fama–French five factors (2015) models in line with illiquidity factor to adjust the returns of the zero-cost portfolio. I also include additional momentum factor, short-run reversal and long-run reversal factors as a consideration of behavioral pricing effect in the news sentiment trading strategy portfolio. Columns (2)-(6) clearly show that none of the models fully explain the abnormal returns generated by the zero-cost portfolio that is based on firm-specific news sentiment. The average abnormal daily return ranges from 6.1 to 6.8 basis points across different pricing models. Notably, the liquidity factor does not contribute any significant effect

[39] See studies by Easley et al. (2010) and Kelly and Ljungqvist (2012) argue the relationship between liquidity risk and information asymmetry. I use the Fama–French five factors to estimate illiquidity beta for each stock.

to the value of the abnormal return from the zero-cost news sentiment portfolio.[40]

Interestingly, the full model in column (6) shows that the news sentiment factor portfolio return is negatively significant in relation to both the value and the short-term reversal factors. Additionally, it has positive factor loadings on the momentum and investment factors. Intuitively, the significance of the momentum and short-term reversal factors captures potential behavioral effects from news sentiment affecting investors' valuation of stock performance. For example, if return on a stock is high on day t, and, in the meantime, there is positive news about the firm, investors tend to under-react to this information and generate a momentum effect (Hong and Stein, 1999). Factor loading on the investment factor may be explained by news sentiment regarding companies' fundamentals, such as reporting on a firm's investment plan. For instance, if news stories report pessimistically about a company that plans to shrink its future investment, investors analyzing the company may suffer a high leverage issue in the firm to reduce investment and require a high expected return, and vice versa.

In sum, existing pricing factors have some explanatory power, either from behavioral finance perspectives, such as momentum and short-term reversal, or from fundamental interpretations such as value or investment factors. However, these baseline asset pricing factors' effects on the firm-specific news sentiment zero-cost portfolio are not economically significant and they only capture 1.5% to 7.5% of the variation in the daily zero-cost portfolio return.

This study proposes a novel interpretation of the zero-cost news sentiment portfolio's risk-adjusted abnormal return; it can be seen as an information risk premium resulting from the biased tone in firm-specific news. The abnormal return from the news sentiment zero-cost portfolio offers empirical evidence to verify the theoretical study in section 2.2 for the argument of biased tone in the firm-specific news leading to a deviation in information risk. One could question whether the firm-specific news sentiment trading strategy can generate considerable profits. In fact, taking a moderate round-trip transaction cost, such as 5 basis points, the rough calculation for daily return (including the trading cost for the risk-adjusted abnormal daily return from the zero-cost portfolio) is about 1.17 basis points. Obviously, the profit will be lost by increasing the round-trip trading cost, since a daily-basis formation is too frequent in reality. Of course, the trading cost could be reduced through a weekly re-balance or

[40]The information risk premium raised by firm-specific news sentiment, which causes a deviation in the information incorporated into the price, is distinct from the illiquidity effect and the role of this information risk cannot be precluded by risk premium from the illiquidity factor.

Table 2.6: Firm-Specific News Sentiment Factor Risk Premium-Fama-French Factor Model Testing

This table shows daily risk-adjusted returns (α) from firm-specific news sentiment zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment are: $N/S; N/B; M/S; M/B; P/S; P/B$ sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% ($P/S; P/B$) and the average of short position in the two negative sentiment portfolios 30% ($N/S; N/B$) each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows Pearson correlation between the news sentiment portfolio return and conventional factors. Panel B presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	MKT_t	SMB_t	HML_t	RMW_t	CMA_t	UMD_t	ST_t	LT_t	$PSLIQ_t$
$Sentiment_t$	-0.106	0.029	-0.102	0.071	0.084	0.202	-0.142	0.025	0.010
MKT_t		0.070	-0.012	-0.425	-0.333	-0.257	0.355	-0.084	0.085
SMB_t			0.052	-0.298	0.055	0.029	0.014	0.283	0.042
HML_t				0.088	0.483	-0.344	-0.097	0.477	0.098
RMW_t					0.280	0.151	-0.245	-0.161	0.040
CMA_t						0.065	-0.283	0.520	0.025
UMD_t							-0.126	0.030	-0.066
ST_t								-0.138	0.061
LT_t									-0.029

Panel B Risk-Adjusted News Sentiment Zero-Cost Portfolio Return

	$Sentiment_t$	$CAPM$	$FF3$	$FF5$	$FF5+UMD$	$FF5+Full$
α	0.066	0.067	0.068	0.064	0.061	0.065
t_α	(6.397)	(6.588)	(6.756)	(6.390)	(6.143)	(6.640)
MKT_t		-0.065	-0.069	-0.031	-0.016	0.000
t_{MKT}		(-4.795)	(-5.349)	(-2.582)	(-1.382)	(0.039)
SMB_t			0.051	0.056	0.041	0.035
t_{SMB}			(1.991)	(2.304)	(1.732)	(1.430)
HML_t			-0.123	-0.202	-0.126	-0.135
t_{HML}			(-4.263)	(-7.421)	(-4.851)	(-5.091)
RMW_t				0.058	0.038	0.029
t_{RMW}				(1.670)	(1.125)	(0.784)
CMA_t				0.235	0.185	0.147
t_{CMA}				(5.430)	(4.438)	(3.196)
$PSLIQ_t$			0.021	0.018	0.020	0.024
t_{PSLIQ}			(1.151)	(1.026)	(1.166)	(1.425)
UMD_t					0.115	0.109
t_{UMD}					(6.780)	(6.806)
ST_t						-0.088
t_{ST}						(-4.669)
LT_t						0.019
t_{LT}						(0.568)
$\bar{R}^2$	0.007	0.011	0.024	0.038	0.055	0.064
$Days$	5241	5241	5241	5241	5241	5241

tailoring of extreme sentiment stocks.[41] Nevertheless, the main purpose in this mimicking (zero-cost) portfolio factor analysis is to investigate the validity of the implications of news sentiment leading to a deviation in information risk, for which investors require high expected returns as compensation. The firm-specific news sentiment trading strategy leaves room for future study from a behavioral finance perspective.

In the Appendix B.9, I conduct several robustness tests - for example: conducting fixed effect regressions test to verify the negative relationship between the market news sentiment and measures of economic uncertainty; using the count of SEC EDGAR file searching volume as a direct measure of information acquisition; excluding data from earnings announcement days; sorting data into sub-samples based on financial characteristics; choosing an alternative asset pricing model (q-factor model by Hou et al. (2015)); and utilizing an innovative news pricing factor to control for a genuine or mis-valuation effect from firm-specific news sentiment. The empirical results are robust to all of the tests.

2.6 Conclusions

In this chapter, I developed a theoretical model and empirically tested the predictions implied by the model to demonstrate that biased public information from news media gives rise to investors' biased acquisition of firm-specific information. First, the static information acquisition model derives several theoretical predictions, by introducing a channel via which clostless but biased public information is exogenously distributed to investors before they make investment decisions. Because investors naively do not adjust for the bias in public news information, their beliefs about the systematic and firm-specific uncertainties included in the risky asset's payoff are biased. Thus, investors' acquisition of private information about the risky asset is subject to their biased beliefs.

Second, the empirical tests I conducted, where sentiment in the news is used as a proxy for biased public information in the model, yielded results that are consistent with my theoretical predictions. Investors' acquisition of firm-specific information is significantly inversely related to the tone (sentiment) in the news about the stock market or particular firms. In addition, firm-specific news sentiment in the model causes a

[41] As Table 2.5 shows, news sentiment can predict a positive cross-section stock cumulative return of up to 10 days. Alternatively, one could construct a trading strategy by tailoring for firm news sentiment, for example (as Ke et al. (2019) proposed) by adopting a strategy of buying the 50 stocks with the most positive sentiment and selling the 50 stocks with the most negative sentiment.

deviation in firm-specific information acquisition from the rational expectations equilibrium. This causes the degree of information risk to deviate as well. Empirically, the Fama–Macbeth (1973) regression verifies the positive predictability of firm-specific news sentiment on expected returns, as the theoretical model predicts. Also, by constructing a daily zero-cost portfolio return factor for firm-specific news sentiment, the annualized risk premium is about 17% per year. This result is robust to the addition of additional traditional pricing factors and a novel news effect pricing factor as controls, and switching to alternative asset pricing model such as the q-factor model.

In sum, this study introduces a new understanding of the channel of irrationality in economic activity, specifically, information acquisition in investment. In most of the behavioral studies in finance and economics, researchers relax the assumption of rational economic agents and argue that psychological irrationality in humans plays an important role in economic study. This study does not oppose this classical theory. The key claim in this study is based on the perspective of the behavioral studies, but challenges assumptions about how bias arises. It is difficult to claim that economic agents are rational all the time, as an advocate of behavioral economics would believe, but it is also difficult to accept that investors intend to make important decisions, particularly investment decisions, from an irrational or psychologically-biased standpoint. As emphasized by Tirole (2002), the enrichment derived by the incorporation of psychological factors in economics models should focus on parsimony and normative analysis rather than the impulsive framework of psychology. In this study, I keep the view aligned with behavioral finance to argue for the role of irrationality in conducting economic activities. Instead of stressing human behavioral irrationality, the trigger-biased information percolation proposed by Hirshleifer (2020) discussed in this study conceptualizes irrationality within economic agents as social transmission bias through the distribution of news. In particular, irrationality forced by the biased information transmission through news has a significant impact on investors' decisions concerning further information acquisition. As the theoretical model demonstrates, investors' sub-optimal choices come down to thinking and decision-making that is affected by the transmission of biased information from sources upon which they may rely, such as the news media.

Chapter 3

Factor Structure in Cryptocurrency Returns and Volatility

3.1 Introduction

Cryptocurrencies have caught the eye of individual and institutional investors, primarily because of the exceptional returns they have offered. Though they have been in existence since 2008, the year Bitcoin was invented by Nakamoto,[1] the most critical period in the history of cryptocurrencies is the so-called Bitcoin bubble. Between April 2017 and December 2017, the dollar price of Bitcoin rose from \$600 to \$19,815. On December 16, 2017, as the Bitcoin price reached a historical high, *The Wall Street Journal* published an article entitled "Is Bitcoin a Bubble? 96% of Economists Say 'Yes'". From January to February 2018, the Bitcoin price fell by 65%.

Despite this enormous Bitcoin price fall, which was shared by many cryptocurrencies, the total crypto market capitalization remains substantial; in September 2018, it was around \$208 billion. As cryptocurrency trading has become more popular, finance academics have been drawn to examine the market, starting with Yermack (2015).

Based on the dramatic pricing behaviors in the cryptocurrency market, we explore the characteristics of cryptocurrencies as financial assets and investigate whether a factor structure contains information to explain the cross-sectional variations in the returns and volatility of cryptocurrencies. On the basis of high-frequency quote and transactions data, we are able to estimate realized volatility and returns for the nine

[1]See the study by Nakamoto (2019)

most liquid cryptocurrencies quoted against Bitcoin between October 2016 and November 2018. We go on to test whether the common components of returns and volatilities are driven by major macroeconomic factors, and how the crypto factor structures were affected by the Bitcoin pricing bubble. Finally, we test whether Bitcoin acts as a fundamental market factor in the cryptocurrency market.

We demonstrate nine stylized facts:

Fact 1: *Daily realized cryptocurrency volatility has high persistence.*

Fact 2: *The distribution of the logarithm of realized volatility of cryptocurrencies is close to normal.*

Fact 3: *The factor structure in daily cryptocurrency volatility is stronger than the factor structure in returns.*

Fact 4: *Economic and financial factors do not have strong explanatory power on the common factors of cryptocurrency return and volatility and there is a weak inverse relationship between cryptocurrency risk and macroeconomic indices.*

Fact 5: *Bitcoin can be considered for most cryptocurrencies as a fundamental factor able to explain a small proportion of the variations in return and volatility.*

Fact 6: *The Factor Structure model is more powerful in explaining variation in returns and volatilities during the Bitcoin bubble period and this explanatory power persists - and for volatilities actually increases further - after the Bitcoin bubble burst.*

Fact 7: *There is heterogeneity in the relationship between Bitcoin and other cryptocurrencies for both returns and volatility after the Bitcoin pricing bubble burst.*

Fact 8: *Cryptocurrency betas with Bitcoin were negative before the Bitcoin bubble burst but became positive after the bubble burst.*

Fact 9: *The fraction of variance of cryptocurrency explained by the Bitcoin variance is high during the bubble period, and the explained fraction remains at an elevated level in the post-bubble period.*

This chapter is inspired by the trending studies in the blockchain economic literature. Researchers develop theoretical models to develop economic insights based on the application of blockchain to economic transactions (Abadi and Brunnermeier, 2018; Huberman et al., 2019; Cong and He, 2019; Schilling and Uhlig, 2019; Cong et al., 2019). For example, Biais et al. (2019) develop a stochastic game to model the logic of the blockchain working protocol. Sockin and Xiong (2020) argue that the efficiency of tokenization in a decentralized economy such as that of Bitcoin users can alleviate central bankers' delegation issue. In fact, their study builds a foundation on which to develop a theoretical pricing model for cryptocurrency. As the cryptocurrencies represent the value of their blockchain technology, we uncover the characteristics of cryptocurrencies by treating them as financial assets. Specifically, we mainly document a horse race between cryptocurrenices and commodities.

Our study makes a particular contribution from the asset pricing perspective by discovering the existence of a factor structure (Weber, 2016; Chiu and Koeppl, 2017). In a seminal study, Bianchi (2020) conducts an empirical study on the returns of cryptocurrencies and traditional financial assets. His main finding is that there is no significant correlation between cryptocurrency returns and the return of traditional financial assets. Only gold and crude oil have weak correlations with cryptocurrency. A recent study by Liu and Tsyvinski (2018) finds similar results, concluding that cryptocurrency prices contain no information related to other financial assets or pricing factors. However, Liu and Tsyvinski explore the theory that the returns of cryptocurrency are predicted by factors that are specific to cryptocurrencies. This may imply that the information explaining the pricing behaviors of cryptocurrency is not shared with traditional financial assets in the financial market. In other words, questions relating to the cryptocurrency market should be investigated by focusing on the inherent characteristics of cryptocurrency instead of naively borrowing from studies on traditional financial assets. We use high-frequency tick data to construct factor structures for both returns and volatilities. The empirical results we find in this study are comparable to a similar study in commodity market (Christoffersen et al., 2019). The factor structure constructed by the principal component analysis serves as a pricing model to explain variations in the cross-sectional cryptocurrencies' return and volatility.

In addition, we contribute to the studies of cryptocurrency by dating the price bubble of Bitcoin. Taking Bitcoin as a financial asset at this moment, we define its price bubble following econometric literature (Phillips and Yu, 2011; Phillips et al., 2011) as the stochastic process of Bitcoin price becomes explosive, such that the bubble period is the one for which the augmented Dickey–Fuller (ADF) statistics are greater than the right-tailed critical values. We find that the bubble period extends from May 24^{th} 2017 to January 28^{th} 2018, which coincidentally covers one third of the sample period in our study. As the cryptocurrencies are quoted against Bitcoin, its price bubble makes the relation between Bitcoin and other cryptocurrencies unstable and characterized by heterogeneity after the bubble burst. We believe that if a bubble inflates in the cryptocurrency market, the price movement should exhibit explosive characteristics, consistent with the rational bubble explanation in the stock market. To the best of our knowledge, we are the first to date the Bitcoin bubble by mimicking the definition and methodology of dating pricing bubbles in the financial market.

The remainder of the chapter is structured as follows. In section 3.2, we describe our data and methodology for computing returns and estimating realized volatility. We construct factor structure models in section 3.3. The explanatory powers of economic factors on the common components in cryptocurrencies are tested in section 3.4. Section 3.5 detects the timing of the Bitcoin price bubble and the impact of the bubble on the factor structure of other cryptocurrencies. In section 3.6, we estimate cryptocurrency market betas and compute systematic risk ratios contributed by Bitcoin. We draw our conclusions in Section 3.7. Supplementary figures and tables can be found in the Appendix C.

3.2 Construction of Returns and Realized Volatility

3.2.1 Cryptocurrency Data

We obtain intraday trading data on cryptocurrency from Kaiko, a company that collects tick data pertaining to cryptocurrencies. Kaiko provides tick by tick data on more than 200 cryptocurrencies traded on 15 large and liquid cryptocurrency exchanges.[2] As explained below, we augment the Kaiko data with similar data from CoinAPI.io, a company that provides a similar service to Kaiko, providing cryptocurrency data accessed through querying APIs from multiple exchanges.

We make several methodological decisions regarding use of the source data. First, we analyse cryptocurrency exchange rates against Bitcoin (Cryptos/BTC) rather than crypto rates against fiat currencies such as the U.S. dollar. The extreme price changes of Bitcoin versus the dollar noted above were mirrored by most other cryptocurrencies. Studying crypto exchange rates against the dollar would have inevitably uncovered enormous common structures, as cryptos first rose and then fell against the dollar - or any other non-crypto base price. While this is an important issue to consider it is not what we wished to examine in this study. Instead, we focus on testing for common structures *between* cryptocurrencies and instead use the BTC/USD boom and bust episodes as sub-periods for our tests. Since Bitcoin is the headline cryptocurrency, we use it as the base price against which all crypto exchange rates are measured.

Second, and following from the decision to focus on BTC-cross crypto rates, we take data from the Bittrex exchange, a leading exchange located in Seattle that mainly facilitates trades of cryptos against Bitcoin. Makarov and Schoar (2020) have noted

[2] The exchanges are Bitstamp, Kraken, BTCC, Bittrex, Coinbase, OkCoin, Bitfinex, Poloniex, Bithumb, Gemini, Quoine, bitFlyer, Huobi, Binance and Zaif.

Table 3.1: Summary of Cryptocurrency

The table shows all cryptocurrency used in our study. For each asset, we report the trading symbol time-zone, market capitalization, close price, circulating supply, and percentage of total market capitalization in the cryptocurrency market. The summary data is from https://coinmarketcap.com. All statistical data is up to November 2018 which is the last month in our sample period.

Currency	Symbol	Time Zone	Market Cap	Price	Circulating Supply	% Total Market Cap
Bitcoin	BTC	UTC	$65,549,846,077.00	$3,768.79	17392787	54.28%
Ripple	XRP	UTC	$13,998,356,446.00	$0.35	40327341704	11.59%
Etherrum	ETH	UTC	$11,158,159,719.00	$107.90	1866712302	9.24%
Litecoin	LTC	UTC	$1,692,307,423.00	$28.54	59229875	1.40%
Monero	XMR	UTC	$929,735,016.00	$56.02	16596133	0.77%
Dash	DASH	UTC	$743,512,468.00	$87.85	8463191	0.62%
Ethereum Classic	ETC	UTC	$478,701,141.00	$4.50	106284797	0.40%
Zcash	ZEC	UTC	$339,981,605.00	$64.03	5309689	0.28%
Lisk	LSK	UTC	$146,100,728.00	$1.30	112501790	0.12%
Stratis	STRAT	UTC	$64,322,236.00	$0.06	99106480	0.05%

that cryptos often trade at markedly different prices on different exchanges; hence to ensure comparability it is important that all rates come from the same exchange.[3]

Finally, though many cryptos are traded at the same time, many do not survive long, and many others have only recently been introduced. We select nine cryptocurrencies that have had data available throughout the sample period from October 2016 to November 2018. These nine currencies are Ethereum (ETH), Ethereum Classic (ETC), Ripple (XRP), Litecoin (LTC), Dash (DASH), Zcash (ZEC), Lisk (LSK), Monero (XMR), Stratis (STRAT). There is a clear and conscious selection bias inherent in this decision. Our results pertain only to this set of relatively long-lived cryptocurrencies selected for the very reason that they have survived.

Table 3.1 summarizes the cryptocurrencies' overall market capitalization, volume and circulating supply at the end of November 2018. Including Bitcoin, the cryptocurrencies we study in this chapter represent almost 79% of the total market capitalization of the cryptocurrency market.

3.2.2 Other Data

We collect commodity futures and foreign exchange spot data from Thomson Reuters Tick History (TRTH) at the minute frequency. Specifically, we use commodity futures on crude oil, gold, S&P500 E-mini, and CBOE SPX VIX, and foreign exchange spot data for CNY/USD and EUR/USD. Our cryptocurrency study focuses on the interval between October 2016 and November 2018 and makes use of this market's 24-7 continuous trading feature. Analysis using the foreign exchange factor is from February 2017

[3]Several other papers document potential problems of investment in cryptocurrencies, including Borri and Shakhnov (2018), Hu et al. (2019) and Borri (2019).

to November 2018 since offshore trading in the Chinese currency started in February 2017, as discussed further in section 3.4.

We follow the standard high-frequency data cleaning process to remove bad data points. To be exact, we follow the first three steps of the quote data cleaning processes described by Barndorff-Nielsen et al. (2009). We do not conduct the fourth step that eliminates extreme quotes because we want to preserve the nature of cryptocurrency trading as much as possible. Nevertheless, the results of our analysis are insensitive to the removal of extreme quotes.

All cryptocurrency and financial products' daily realized volatilities are calculated from minute-sampled mid-quote data after the data cleaning procedures.

3.2.3 Return and Realized Volatility Calculations and Data Cleaning

We analyze daily return and realized volatility measures for our set of nine cryptocurrencies. In theory, given that the crypto market trades continuously over seven days per week, calculating these measures should be straightforward. Unfortunately, the data are imperfect and there are intervals where relevant observations are missing. We first explain the methods used to calculate our key measures on the assumption of perfect data and then detail how we deal with the missing data.

We construct realized volatility following Christoffersen et al. (2019) and Zhang et al. (2005). Each day has an $(n + 1)$ 1-minute time-grid price. The n 1-minute time-grid returns at day t are calculated as:

$$r_{t_j} = log(Mid_{t_j}) - log(Mid_{t_{j-1}}) \tag{3.1}$$

where $t_j - t_{j-1}$ is equal to one minute and $log(Mid_{t_j})$ is the mid quote of logarithm of ask price and logarithm of bid price. We then calculate each five-minute return by summing the five one-minute returns:

$$\widetilde{r}_{t_k} = \sum_{k=j}^{j+4} r_{t_k} \tag{3.2}$$

Each day will have $(n - 4)$ five-minute return. Finally, the daily measure of 5-minute realized volatility calculated with 1-minute subsampling is defined as :

$$RV_t^{oc} = \frac{n}{5(n-4)} \sum_{k=1}^{n-4} (\widetilde{r}_{t_k})^2 \tag{3.3}$$

Using subsampling techniques to calculate 5-minute returns reduces market microstructure noise in the volatility estimate.[4]

The Kaiko data provide minute snapshots of the crypto orderbook up to ten levels from the best bid and ask prices (giving both price and depth data). In theory, since cryptos trade around the clock, seven days per week, we should observe 1440 snapshots of the data throughout our sample. Unfortunately, this is not the case. We therefore check whether the missing prices can be filled in with data from the CoinAPI.io database. In theory, this should be a reasonable solution since when both Kaiko and CoinAPI provide data for the same crypto from the same exchange, the data are exactly comparable. Nevertheless, even after filling in all possible missing observations, data are still sometimes missing, particularly in the April-August 2017 interval. That data are missing in this interval is probably not random. The Bitcoin price was rising rapidly at this time and trading was extremely active. We suspect that data providers struggled to keep up with orderbook developments leading to data problems.

As a result of this problem, we encounter some days with intervals during which no orderbook data are available. We adopt two methods to solve this issue. Our first approach is to follow Müller et al. (1990) Dacorogna et al. (1993) and Andersen et al. (2001). This involves simply interpolating in a linear fashion across intervals in the data as long as the interval is small enough for this to be reliable. To decide what constitutes a small enough interval we run the following test.

For each currency, we extract those days with the full 1440 minutes of data. We randomly delete observations within the day creating missing data intervals of length j-minutes. These intervals are then re-filled by linearly interpolating across the gap. We then calculate the daily realized volatility as discussed below. One realized volatility is calculated for the original full data set and the other is calculated using the data set containing j minutes of interpolated data. Finally, we compute the correlation between the two computed realized volatilities. We deem the interpolation to be acceptable if the correlation is greater than or equal to 0.98. In practice, we conclude that data can be linearly interpolated up to $j = 200$ minutes without loss of accuracy. Above that, the correlation is unacceptably low and a second econometric method has to be employed.

We also test the power of linear interpolation on days with multiple missing data intervals (for example, we may have ten missing intervals in the data during a day,

[4]The market microstructure noise issue on high frequency data has been well discussed by Campbell et al. (1997), Andersen et al. (2005) and Ait-Sahalia et al. (2005).

each 50 minutes long but separated in each case by ten minutes of observed data). Perhaps unsurprisingly, the problem of several relatively small gaps in a day is far less severe than the problem arising from one long missing interval. The example of one day with ten 50-minute intervals is acceptably corrected by using linear interpolation across each gap in the data, even though the intervals total some 500 minutes. This is much longer than the single interval that can be successfully interpolated. In summary, as long as the data missing between two timestamps do not exceed 200 minutes, we use linear interpolation.[5]

For intervals longer than 200 minutes, we use a second procedure in line with that used by Hansen and Lunde (2005). Their method is designed to account for systematic breaks in trading as is typically observed in stock markets. Hansen and Lunde (2005) propose that both the realized variance computed from high-frequency data during trading hours and the squared close-to-open return (r^{co}) over an inactive period contain information relevant to computing the integrated variance (IV) of an asset.[6]

To minimize the difference between realized variance and integrated variance,[7] Hansen and Lunde (2005) develop optimal weights for r_t^{co} and RV^{oc}, which remove much of the noise due to using high-frequency data:

$$RV_t(w) = w_1(r_t^{co})^2 + w_2 RV_t^{oc} \tag{3.4}$$

The Hansen and Lunde (2005) technique can be applied easily if trading breaks are of equal duration and occur each trading day, since the parameters driving the optimal weights can be estimated from simple sample averages. However, in our cryptocurrency data, the trading breaks occur at different points during the day, are of different lengths and only occur sporadically. We therefore adapt the Hansen and Lunde (2005) approach accordingly as follows.

For days with a single trading break (longer than 200 minutes) on day t, we simulate all other days with full data availability to have the exact same trading break (occurring at the same time, and for the same interval). We then apply the Hansen and Lunde (2005) technique outlined above (and described in further detail in their paper) to calculate the optimal weights for the close to open squared return and the open to close realized volatility.

[5] We also check whether other methods such as Spline or Lagrange interpolation perform better than linear interpolation. The results are very similar to linear interpolation.

[6] See the detailed analysis by Andersen and Bollerslev (1998).

[7] That is, $\min_{\omega} E[RV_t(\omega) - IV_t]^2 = 0$

For days having more than two breaks we adapt the Hansen and Lunde (2005) method and apply:

$$RV_{i,t}(w) = \hat{w}_1 \sum_{b=1}^{B} (r_{b,t}^{co})^2 + \hat{w}_2 \sum_{b=1}^{B+1} RV_{b,t}^{oc} \tag{3.5}$$

where B is the number of breaks, $RV_{i,t}^{oc}$ is the realized volatility between each break calculated from equation (3.3), and $r_{i,t}^{oc}$ is close to open returns between breaks. We again create simulated data with exactly matching breaks from those days with complete data and proceed as usual.

The combination of simple linear interpolation across small gaps in the data and the Hansen and Lunde weighting when there are longer gaps allows us to compute daily realized volatilities for all currencies. To compute daily returns, as used in the analysis below, we need a price at midnight each day. On some days, there are trading gaps spanning midnight. We therefore linearly interpolate between the last available mid-price on day t and the first available price on day $t + 1$ to obtain the midnight price.

3.2.4 Properties of the Cryptocurrency Daily Returns and Volatilities

Table 3.2 provides descriptive statistics of the daily log returns of the nine cryptocurrencies. All nine exhibit positive skewness, and the extreme values - both maxima and minima - are dramatic. The first order autocorrelation does not show strong persistence at the 1% level except for ZEC and LSK, and the Ljung-Box test shows no significant persistence across 5 to 21 lags. Figure 3.1 plots the autocorrelation function up to 60 lags confirming that cryptocurrency daily returns do not show high persistence.[8]

Figure 3.2 plots the daily realized volatility (RV_t) of the nine cryptocurrencies and Panel A in Table 3.3 reports descriptive statistics. As expected, the cryptocurrencies are very volatile and the RVs of all cryptocurrencies also have high positive skewness and kurtosis. The maximum daily volatility in our sample period is extremely large, even compared to commodities (see the study by Christoffersen et al. (2019)). More importantly, the first-order autocorrelation is large and significant at the 1% level for all cryptocurrencies, and the Ljung-Box test statistics are also strongly significant across both 5 and 21 lags. It is clear that the realized volatilities of cryptocurrencies are highly persistent and this is the first stylized fact we report:

Fact 1: *Daily realized cryptocurrency volatility has high persistence.*

[8]The price level and return figures of the nine cryptocurrencies can be found in Appendix C

Table 3.2: Sample Statistics of Cryptocurrency Return

The table shows sample statistics for daily log return for all cryptocurrency during the October 2016-November 2018 period. ACF(1) denotes the first-order autocorrelation. Q(L) is the Ljung-Box test of zero autocorrelation in lags 1 through L. An asterisk indicated at the 1% level.

Statistics	ETH	XRP	STRAT	LTC	ETC	DASH	ZEC	LSK	XMR	BTC
Mean	0.03	0.25	0.16	0.03	-0.07	0.02	-0.61	-0.03	0.02	0.23
Std.	5.22	8.76	7.68	5.44	6.25	5.69	7.23	7.26	5.59	4.42
Min	-25.88	-68.49	-27.11	-23.12	-28.38	-28.85	-51.78	-38.00	-26.36	-17.14
25%	-2.25	-2.87	-3.98	-2.04	-2.89	-2.61	-3.30	-3.67	-2.62	-1.45
50%	-0.29	-0.55	-0.62	-0.46	-0.59	-0.37	-0.79	-0.72	-0.36	0.37
75%	1.52	1.78	3.40	1.30	1.70	1.96	1.74	2.66	2.08	2.31
Max	32.78	101.20	48.91	56.32	54.23	46.46	46.96	40.69	36.96	23.82
Skewness	1.21	2.85	1.07	2.80	1.43	1.27	0.02	0.75	1.24	-0.04
Kurtosis	10.04	36.38	7.83	24.78	13.58	13.84	13.44	8.18	10.67	6.28
ACF(1)	0.07	-0.02	-0.004	0.02	-0.0002	0.04	0.18*	0.09*	-0.06	0.01
Q(5)	16.06*	12.07	6.58	6.20	13.74	2.02	51.19*	15.97*	11.31	8.35
Q(21)	43.55*	44.68*	55.29*	34.44	52.79*	30.57	77.07*	38.58	27.96	24.97

Table 3.3: Sample Statistics of Cryptocurrency Realized Volatility

Panel A : Sample Statistics of Cryptocurrency RV_t

This panel shows sample statistics for daily realized volatility for all cryptocurrency during the October 2016-November 2018 period. ACF(1) denotes the first-order autocorrelation. Q(L) is the Ljung-Box test of zero autocorrelation in lags 1 through L. An asterisk indicated at the 1% level.

Statistics	ETH	XRP	STRAT	LTC	ETC	DASH	ZEC	LSK	XMR	BTC
Mean	4.02	5.82	7.48	4.46	5.57	5.16	6.31	7.96	5.52	4.02
Std.	3.12	5.51	5.62	3.60	3.57	3.89	5.17	5.65	3.72	3.33
Min	0.37	0.60	0.86	0.47	0.72	0.93	0.45	0.94	1.04	0.40
25%	1.71	2.83	3.59	2.04	2.95	2.37	3.01	3.73	2.68	1.96
50%	3.30	4.33	6.26	3.55	4.80	4.11	4.92	6.36	4.98	3.17
75%	5.40	6.74	9.76	5.55	7.25	6.74	8.00	11.08	7.24	5.08
Max	29.80	65.14	70.14	35.79	25.18	31.54	49.42	42.19	45.48	31.09
Skewness	2.02	4.04	3.08	2.45	1.47	2.13	3.26	1.62	2.68	2.93
Kurtosis	10.72	31.04	25.36	13.62	6.60	10.33	21.04	7.44	21.68	17.43
ACF(1)	0.72*	0.68*	0.62*	0.72*	0.72*	0.72*	0.68*	0.68*	0.58*	0.57*
Q(5)	1427.45*	948.94*	1096.56*	1564.51*	1264.85*	1338.11*	965.29*	1211.05*	868.75*	671.98*
Q(21)	3841.71*	1875.95*	2696.37*	4257.24*	3330.80*	4547.58*	2208.61*	3755.61*	2190.55*	1491.79*

Panel B: Sample Statistics of Cryptocurrency Log RV_t

This panel shows sample statistics for log daily realized volatility for all cryptocurrency during the October 2016-November 2018 period. ACF(1) denotes the first-order autocorrelation. Q(L) is the Ljung-Box test of zero autocorrelation in lags 1 through L. An asterisk indicated at the 1% level.

Statistics	ETH	XRP	STRAT	LTC	ETC	DASH	ZEC	LSK	XMR	BTC
Mean	1.12	1.48	1.78	1.23	1.52	1.40	1.60	1.83	1.51	1.12
Std.	0.74	0.73	0.69	0.73	0.66	0.68	0.69	0.71	0.65	0.75
Min	-0.99	-0.51	-0.15	-0.75	-0.33	-0.08	-0.80	-0.06	0.04	-0.91
25%	0.54	1.04	1.28	0.71	1.08	0.86	1.10	1.32	0.99	0.67
50%	1.19	1.47	1.83	1.27	1.57	1.41	1.59	1.85	1.61	1.15
75%	1.69	1.91	2.28	1.71	1.98	1.91	2.08	2.41	1.98	1.63
Max	3.39	4.18	4.25	3.58	3.23	3.45	3.90	3.74	3.82	3.44
Skewness	0.00	0.16	-0.10	0.04	-0.28	0.18	0.11	-0.13	-0.13	-0.17
Kurtosis	2.31	3.24	2.67	2.65	2.72	2.36	2.97	2.33	2.45	3.10
ACF(1)	0.81*	0.80*	0.78*	0.81*	0.77*	0.82*	0.77*	0.80*	0.76*	0.75*
Q(5)	1893.77*	1751.34*	1789.71*	1965.42*	1666.36*	2037.87*	1571.92*	1956.61*	1636.39*	1570.17*
Q(21)	5552.37*	5050.21*	5386.38*	6020.58*	4680.87*	6897.97*	4237.17*	6592.26*	4734.48*	3826.03*

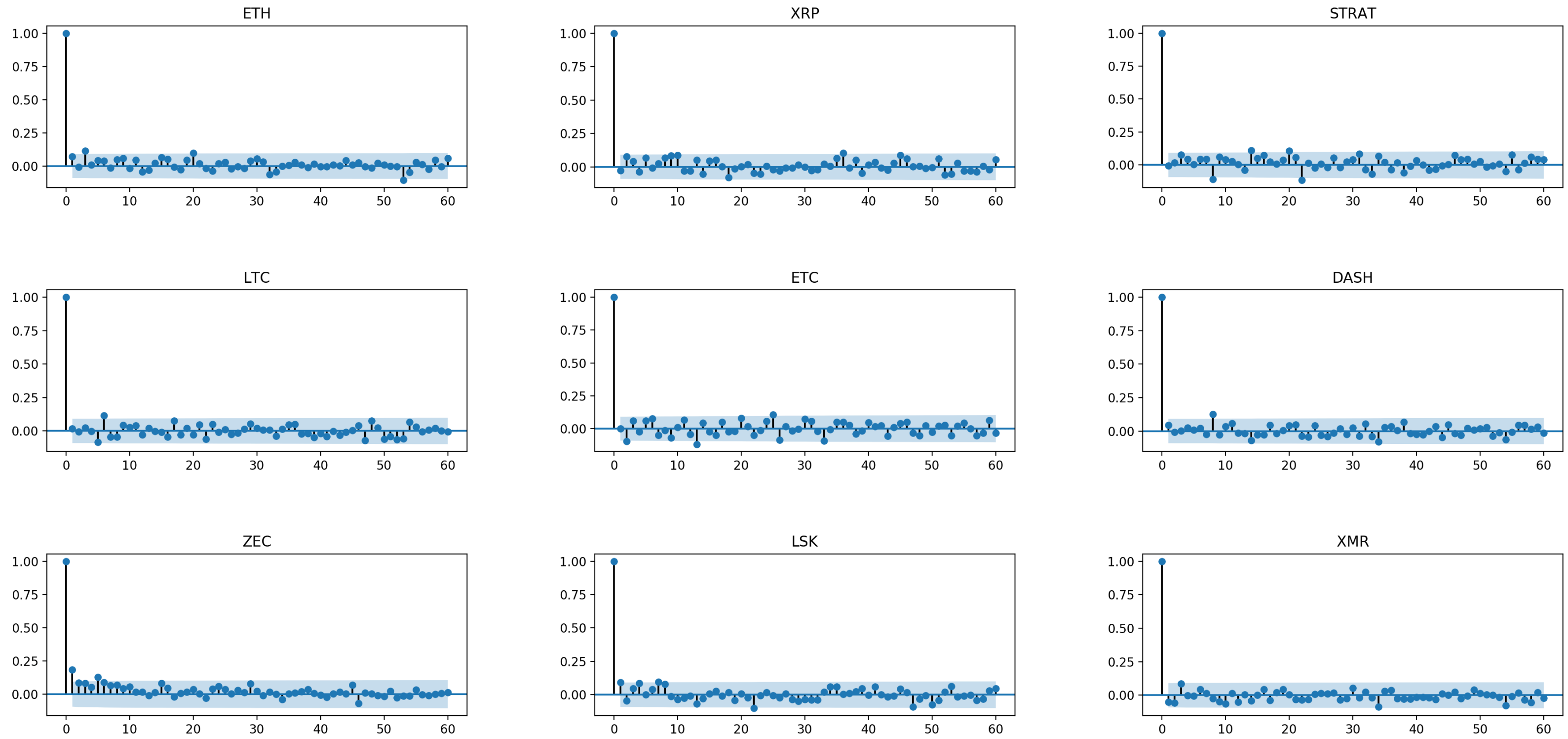


Figure 3.1: Cryptocurrency Daily Return Autocorrelation

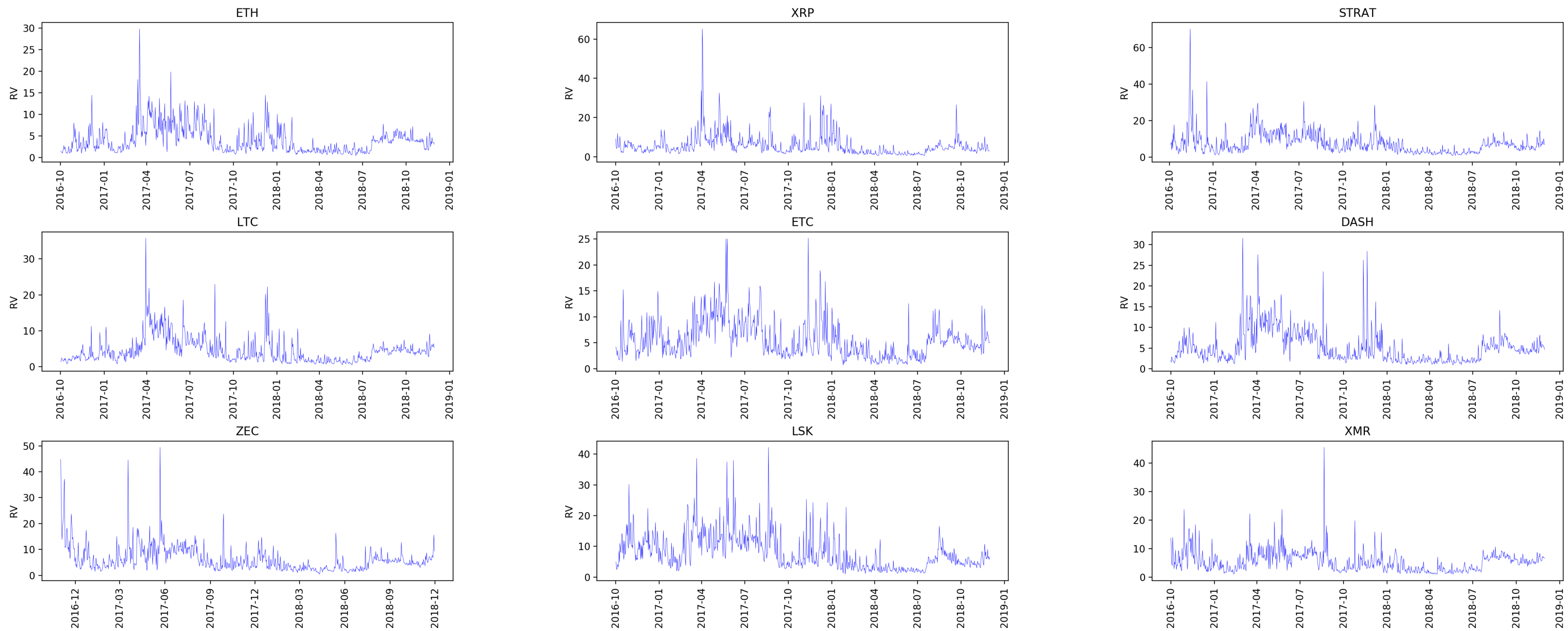


FIGURE 3.2: Cryptocurrency Daily Realized Volatility

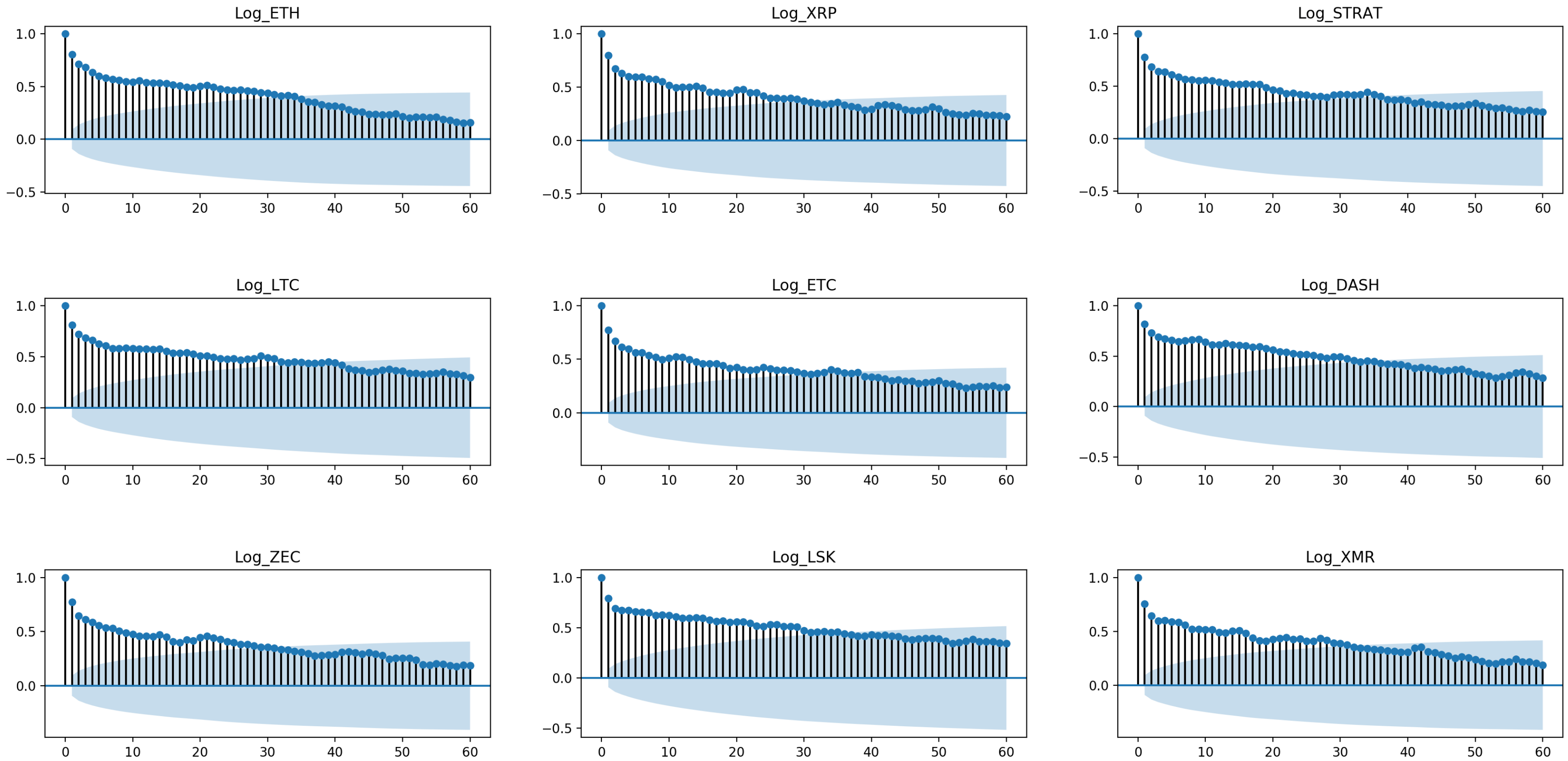


FIGURE 3.3: Log RV_t Autocorrelation

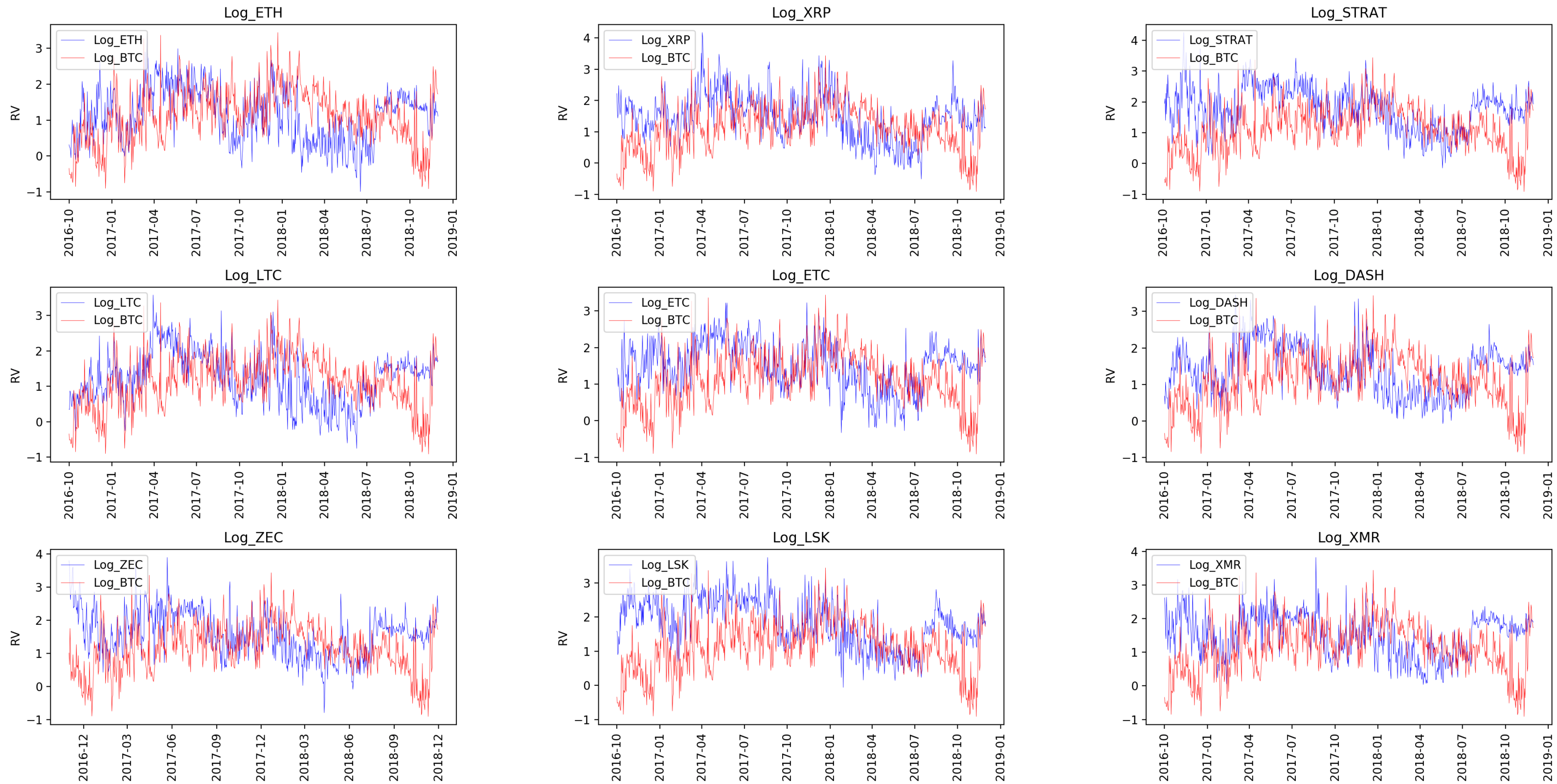


FIGURE 3.4: Log RV_t

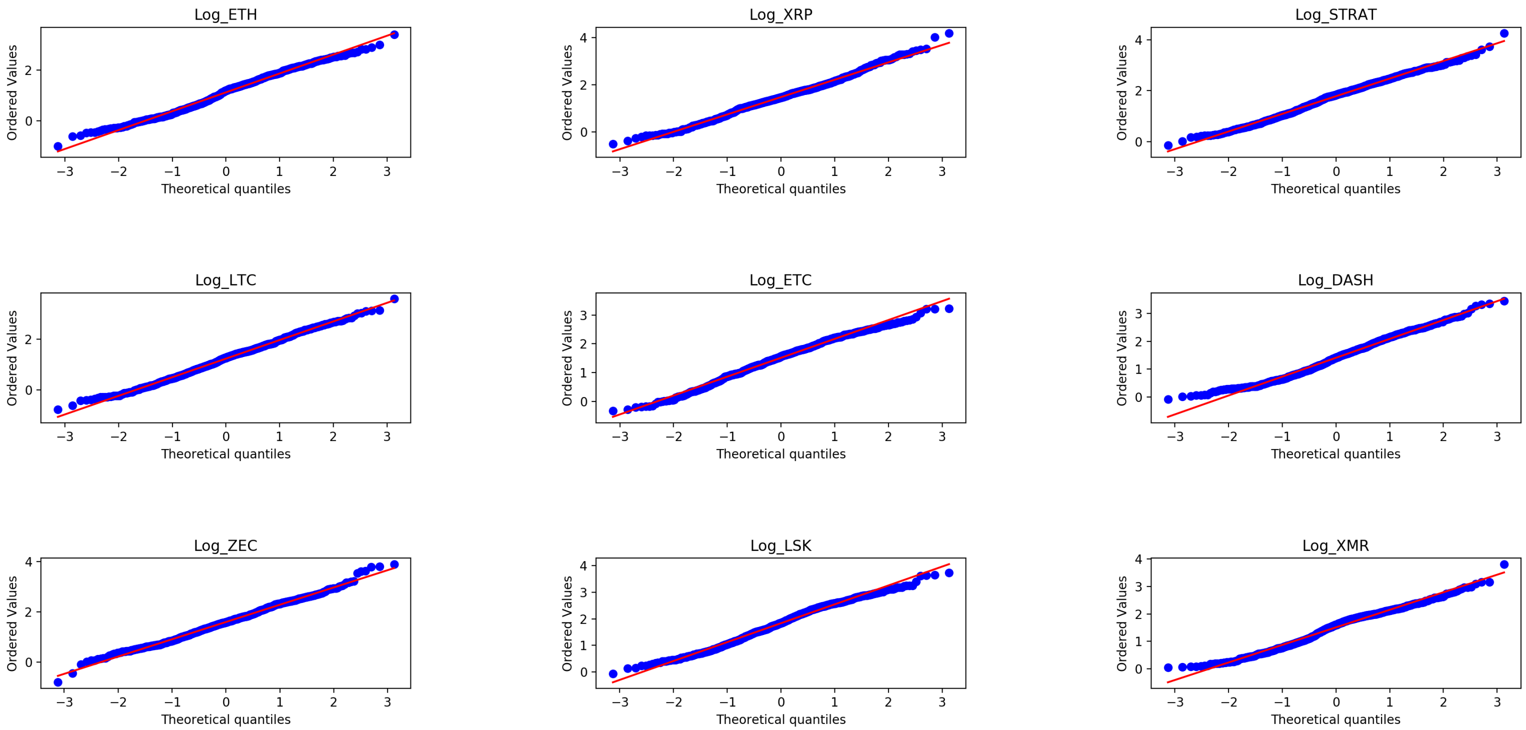


FIGURE 3.5: Log RV_t QQ Plots

Panel B in Table 3.3 reports sample statistics of the natural logarithm of realized volatilities. This does not alter our conclusions regarding the persistence of realized volatilities (see Figure 3.3) but the log transformation changes the data distribution dramatically (see Figure 3.4). The skewness of log realized volatilities are all close to zero, much reduced from levels reported in Panel A. All log realized volatilities have kurtosis levels close to three. Figure 3.5 gives the QQ plot of $log(RV_t)$ for the cryptocurrencies, demonstrating the near normality of log realized volatilities. The effects of taking logarithms of realized volatility have been addressed for equities by Andersen et al. (2001), for the foreign exchange market by Andersen et al. (2001) and in commodity markets by Christoffersen et al. (2019). To the best of our knowledge, we are the first to document that:

Fact 2: *The distribution of the logarithm of realized volatility of cryptocurrencies is close to normal.*

3.3 Factor Structure in Cryptocurrency Returns and Volatility

We investigate the multivariate properties of cryptocurrency returns and volatilities by constructing a factor structure model in the cross-section of cryptocurrencies. Cross sectional common factors in cryptocurrencies in either returns or volatility have not been addressed in the literature. Following Liu and Tsyvinski (2018), who find that cryptocurrency returns are not exposed to stock market or macroeconomic factors, we test whether cross-currency structures in the cryptocurrency market can be explained by factors derived from the cryptocurrency rather than these exogenous factors.

3.3.1 A Common Factor in Cryptocurrency Returns?

To get a first impression of cross-sectional cryptocurrency dependence, Table 3.4 presents the correlation matrix of daily returns across cryptocurrencies in our sample period. The pairwise correlation between two cryptocurrencies' daily return ranges from 15% to 52%. It should be noted that the XRP and STRAT have relatively low average correlations, 22% and 28% respectively. The other cryptocurrencies have a similar average correlation of around 35%. The average return across all pairs of cryptocurrencies is 32%. The correlation of daily returns between each of the nine cryptocurrencies against Bitcoin and the BTC-USD return is always negative and relatively small, ranging from -2% to -19%. The negative correlation between Bitcoin and the other nine cryptocurrencies is not surprising due to Bitcoin being the counter currency of each of the

Table 3.4: Correlation Matrix of Cryptocurrency Return

The table shows Pearson correlations for all cryptocurrency daily log returns during October 2016 – November 2018 sample period. We also report the average pair correlation across each cryptocurrency and the average correlation across all pairwise correlation between two cryptocurrencies.

	ETH	XRP	STRAT	LTC	ETC	DASH	ZEC	LSK	XMR	BTC
ETH		0.26	0.31	0.33	0.52	0.38	0.40	0.37	0.41	-0.19
XRP			0.18	0.30	0.20	0.15	0.23	0.17	0.26	-0.16
STRAT				0.24	0.33	0.27	0.24	0.33	0.36	0.04
LTC					0.40	0.28	0.24	0.28	0.29	-0.10
ETC						0.31	0.32	0.47	0.30	-0.12
DASH							0.44	0.35	0.47	-0.17
ZEC								0.31	0.39	-0.15
LSK									0.31	-0.02
XMR										-0.10
Average	0.37	0.22	0.28	0.30	0.36	0.33	0.32	0.32	0.35	-0.11
All Pair Average	0.32									

cryptos. Therefore, the higher the value of Bitcoin, the higher the Bitcoin return and, since a base cryptocurrency uses Bitcoin as the counter currency, the lower the return of the crypto.

We next conduct principal component analysis to look for evidence of a common factor in our nine cryptocurrency returns. Figure 3.6 plots the first four principal components (PCs). These components explain 39.81%, 10.64%, 10.05%, 8.67% respectively. Figure 3.7 is the plot of cumulative explained ratio by the first four PCs for a total 69.17% of the cross-sectional variation in the nine cryptocurrency returns. Recent studies find evidence of a factor structure in the returns of a cross-section of commodities. For instance, Szymanowska et al. (2014) and Bakshi et al. (2019) work on the portfolio level of commodity futures and find a factor structure, arguing that the major principal components can explain the variation of commodity portfolio return and risk premia from different sorting strategies.

Christoffersen et al. (2019) also look at commodity futures and find relatively weak evidence of a factor structure in daily commodity future returns. In their study, the first four PCs can explain 65.3% variation of the cross-section of 15 commodities' daily return, which is close to our cryptocurrency finding of about 70%. Nevertheless, the first principal component from the cryptocurrencies' returns is almost 40%, which is 10% higher than in their commodity universe. We interpret this as evidence of a factor structure in daily cryptocurrency returns and propose that a factor structure in cross-sectional cryptocurrency return has a considerable amount of pricing explanatory power in this market.

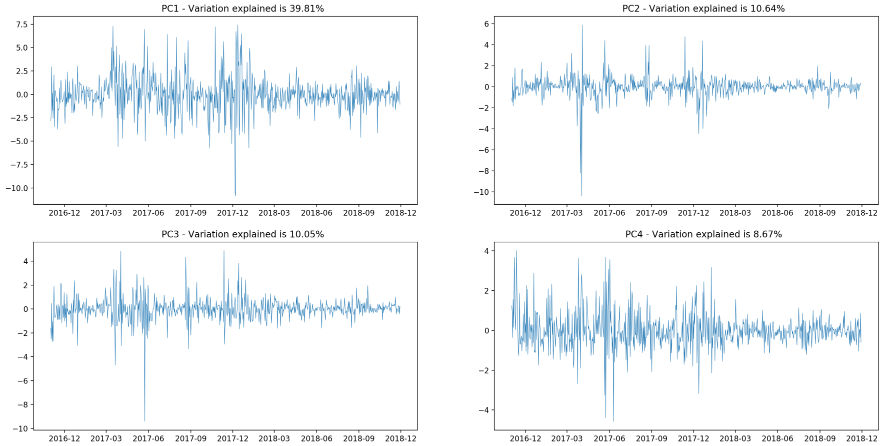


FIGURE 3.6: First Four Principle Components of Cryptocurrency Return

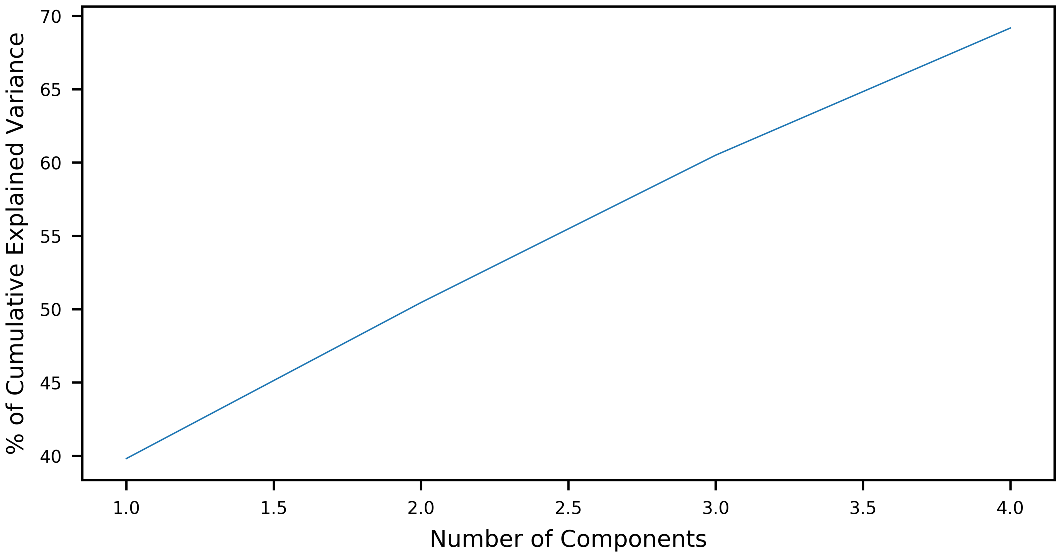


FIGURE 3.7: Cumulative Explained Variance for the First Four Principle Components of Cryptocurrency Return

TABLE 3.5: Correlation Matrix of Cryptocurrency Log RV_t

The table shows Pearson correlations for all cryptocurrency daily log realized volatility during October 2016 – November 2018 sample period. We also report the average pair correlation across each cryptocurrency and the average correlation across all pairwise correlation between two cryptocurrencies.

	ETH	XRP	STRAT	LTC	ETC	DASH	ZEC	LSK	XMR	BTC
ETH		0.71	0.64	0.75	0.74	0.74	0.63	0.67	0.70	0.16
XRP			0.63	0.70	0.67	0.63	0.55	0.66	0.60	0.23
STRAT				0.65	0.65	0.65	0.61	0.67	0.62	0.07
LTC					0.75	0.74	0.57	0.64	0.65	0.12
ETC						0.70	0.65	0.66	0.64	0.12
DASH							0.67	0.69	0.72	0.07
ZEC								0.63	0.67	0.06
LSK									0.64	0.10
XMR										0.06
Average	0.70	0.64	0.64	0.68	0.68	0.69	0.62	0.66	0.65	0.11
All Pair Average	0.66									

3.3.2 A Common Factor in Cryptocurrency Volatility?

Evidence that the factor structure of volatility is stronger than the factor structure of returns has been addressed in finance studies.[9] Therefore, we now question whether a factor structure of volatility exists in cryptocurrencies and, if so, whether it is more powerful than the factor structure in cryptocurrency returns.

We investigate the multivariate properties of nine cryptocurrencies' $log(RV_t)$. Table 3.5 gives the correlations for log volatility of cryptocurrency. There is clear evidence that volatility has much higher correlations compared to returns. In particular, XRP and STRAT have the lowest average correlations of returns, but have an appreciable correlation of $log(RV_t)$, averaging 0.64 for both of them. The average correlations across different cryptocurrencies range from 62% to 70%. The average all pair correlation of $log(RV_t)$ is 66% compared with just 32% for returns. In addition, we check the correlation between the nine cryptocurrencies and Bitcoin. The correlation ranges from 6% to 23%, averaging 11%. In summary, there is weak correlation of log realized volatility between cryptocurrencies and their counter currency Bitcoin. Nevertheless, the weak positive correlations lead us to question the explanatory power of Bitcoin on common factors of cryptocurrency realized volatility and returns.

Figure 3.8 shows that the first four principal components of nine cryptocurrencies' $log(RV_t)$ capture 70.15%, 5.93%, 4.85%, and 3.87% respectively, for a total of 84.8% of the total variation as shown in Figure 3.9. A closer look reveals that the first principal component of $log(RV_t)$ in Figure 3.8 mirrors closely the time series of ETH $log(RV_t)$ in the top left panel of Figure 3.4.

[9]Factor structure of idiosyncratic volatility in the equity market had been addressed by Chen and Petkova (2012), Duarte et al. (2014) and Herskovic et al. (2016); Factor structure of volatility in the commodity market is touched upon by Christoffersen et al. (2019).

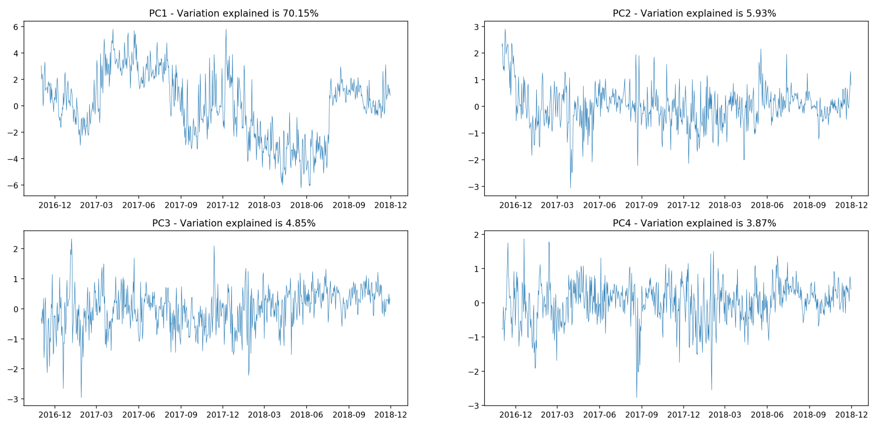


FIGURE 3.8: First Four Principle Components of Log RV_t

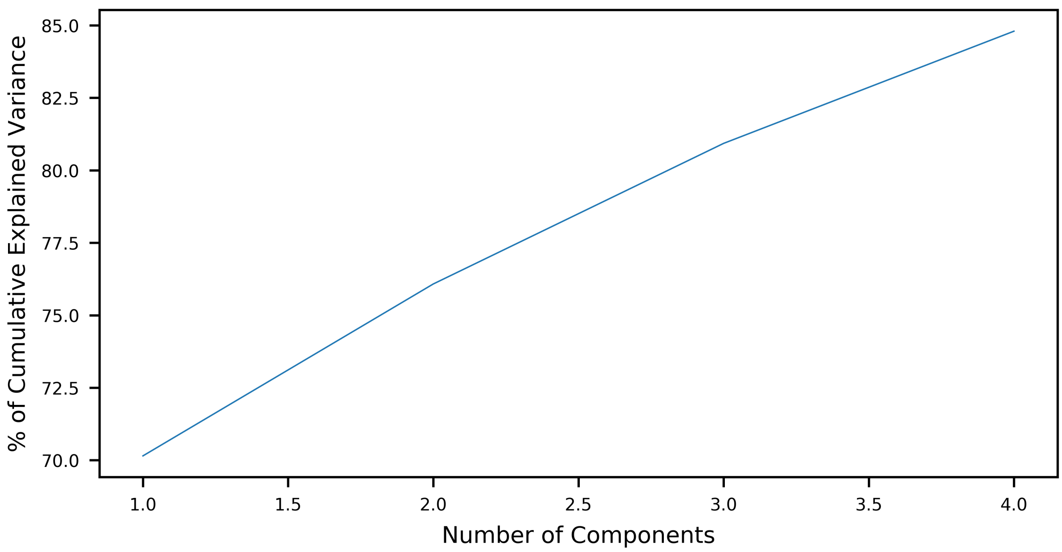


FIGURE 3.9: Cumulative Explained Variance for the First Four Principle Components of Log RV_t

Further, to investigate the factor structure of cryptocurrency returns and volatility, we conduct regression analyses of returns and volatility on their respective PCs. Panel A in Table 3.6 is a regression of each cryptocurrency return on the first four PCs. For each cryptocurrency, we re-conduct a principal component analysis based only on the other eight cryptocurrencies, to avoid mechanical correlations in the regressions. The first PC captures the most variation of cryptocurrency returns, and the other PCs are either marginally significant or insignificant in explaining the commonality of returns. The average of R^2 is about 28%.

Table 3.6: Factor Structure of the First Four Principle Components

Panel A: Regression of Daily Log Return on Principal Components

The panel shows parameter estimates of daily return regressed on principal components of 9 cryptocurrencies during the October 2016 - November 2018. For each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Robust t statistics for each principle component are shown as t_{PC}.

	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	1.790	13.219	-0.059	-0.241	-0.231	-1.034	-0.364	-1.325	0.366
XRP	1.578	6.355	-0.295	-0.492	-0.488	-1.192	-1.299	-2.032	0.127
STRAT	1.900	8.045	0.109	0.317	-0.271	-0.654	-0.025	-0.056	0.201
LTC	1.396	7.831	-0.264	-1.047	-0.765	-2.395	-0.225	-0.734	0.230
ETC	2.011	16.986	-0.146	-0.523	0.905	2.725	-1.374	-2.329	0.377
DASH	1.695	11.668	0.337	0.718	1.064	2.999	-0.603	-1.755	0.323
ZEC	2.048	18.334	0.346	1.411	0.892	2.736	-0.841	-2.908	0.289
LSK	2.092	10.378	0.146	0.345	-1.081	-2.271	-0.034	-0.094	0.287
XMR	1.699	15.134	0.236	0.676	-0.767	-2.343	0.766	1.891	0.335

Panel B: Regression of Daily Log RV_t on Principal Components

The panel shows parameter estimates of daily log realized volatility regressed on principal components of 9 cryptocurrencies during the October 2016 - November 2018. For each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Robust t statistics for each principle component are shown as t_{PC}.

	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.263	29.441	-0.073	-1.834	0.113	3.112	-0.013	-0.303	0.712
XRP	0.237	20.306	0.129	3.628	-0.071	-1.786	0.005	0.111	0.601
STRAT	0.221	22.391	-0.011	-0.253	0.086	2.031	-0.007	-0.191	0.578
LTC	0.251	24.905	-0.138	-3.187	0.118	3.862	-0.009	-0.243	0.693
ETC	0.227	23.229	-0.052	-2.188	0.038	1.179	0.008	0.239	0.673
DASH	0.241	30.340	-0.041	-1.444	0.075	2.885	0.008	0.209	0.698
ZEC	0.210	16.858	-0.153	-3.248	0.068	2.046	-0.001	-0.026	0.567
LSK	0.235	24.460	0.023	0.752	-0.086	-2.312	-0.002	-0.040	0.618
XMR	0.210	25.509	-0.080	-2.169	0.075	2.202	-0.005	-0.111	0.615

Panel B in Table 3.6 shows the regression of each cryptocurrency volatility on the first four PCs, which are again recomputed using the other eight cryptos. While the first PC captures the most variation of cryptos' volatility, the second and third PCs also capture appreciable amounts of volatility variation. The fourth PC is insignificant. All R^2 calculations from each crypto volatility regression are much higher than the R^2 in the return analysis. Noticeably, the average R^2 in volatility analysis is about 64% compared with just 28% in returns. In sum, commonality in volatility is much greater than commonality in returns. We conclude that:

Fact 3: *The factor structure in daily cryptocurrency volatility is stronger than the factor structure in returns.*

3.4 Economic Factors and Cryptocurrency Commonality

In this section, we investigate whether the common factors of cryptocurrency return and volatility are related to economic and financial factors. In particular, we study return and volatility from S&P500 E-mini futures, Gold futures, Crude Oil futures, CBOE SPX VIX and the spot rate of foreign exchange currencies including CNH/USD and EUR/USD. The calculation of returns and volatilities on economic and financial factors is discussed in section 3.2 above.

3.4.1 Impact of Economic Factors on Cryptocurrency Return and Volatility

In section 3.3, we studied the factor structure of cryptocurrency returns and realized volatility. While much more pronounced for volatility, there is still a clear factor structure in crypto returns. We now investigate whether or not the time-series of the key principal components of cryptocurrency return and volatility can be explained by fundamental economic and financial factors. For this study, we regress each PC on each economic factor as follows:

$$PC_{i,t} = \alpha + \beta_1 X_t + \beta_2 PC_{i,t-1} + \epsilon_{i,t} \tag{3.6}$$

Considering potential spurious regression problems, we also add the lagged PC as an additional controlled regressor in the model. We seek to determine whether β_1, the estimated regression coefficient on the economic factor, provides a significant explanatory power for the variation in the PCs. Since cryptocurrency has a 24-7 continuous trading pattern and products of economic factors are not traded over the weekend and on Federal holidays, we merge data which are subject to economic factor trading rules.[10] Except for the two foreign exchange currencies, all economic factor data are available from October 2016 to November 2018 and all factors are available with daily frequency.

Table 3.7 reports regression results from equation (3.6). Panel A in Table 3.7 reports results for the PCs of cryptocurrency returns on economic factors. The first principal component is marginally significantly related to returns on the S&P500 E-mini future, but the R^2 value is quite low. The other economic factors have no significant relationship to the time series variation of cryptocurrency return PCs, and all regression R^2 values are low. This lack of a relationship between the key components of crypto returns and economic or financial factors is consistent with extant studies (Yermack, 2015; Liu and Tsyvinski, 2018; Biais et al., 2019).

Panel B in Table 3.7 repeats the analysis using PCs of cryptocurrency realized volatility. The first key finding is the significant negative relationship between the

[10] The regression result is not sensitive to the data merging method.

Table 3.7: Economic Factors Impact on Principle Components of Cryptocurrency

Panel A: Regression of Principal Components of Cryptocurrency Return on Economic Factors

The table shows output from the regression of the first four principal components of cryptocurrency return on its lags and different economic return factors. S&P500 is daily log return calculated from S&P500 E-Mini Futures. Gold is daily log return calculated from Gold Futures traded. Oil is daily log return calculated from Crude Oil WTI Futures. CNHUSD is daily log return calculated from the daily spot rate of CNH/USD. EURUSD is daily log return calculated from the daily spot rate of EUR/USD. VIX is the daily log return calculated from COBE SPX Volatility Index. All factor minutely data is downloaded from TRTH with sample period from October 2016 – November 2018 except CNHUSD and EURUSD which are not available before February 2017. The principal components are constructed as the matrix of the log return for all 9 cryptocurrencies multiplied by the eigenvectors of the covariance matrix. Robust t statistics for each principle component and economic factors are shown as t_{PC} and $t_{EconomicFactor}$.

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$S\&P500$	$t_{S\&P500}$	R^2
$PC1$	-0.003	-0.035	0.104	2.014	0.142	1.945	0.014
$PC2$	0.004	0.112	-0.017	-0.274	-0.013	-0.289	0.000
$PC3$	0.001	0.036	0.114	1.538	-0.021	-0.516	0.013
$PC4$	-0.004	-0.093	0.003	0.036	0.017	0.364	0.000

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$Gold$	t_{Gold}	R^2
$PC1$	0.006	0.076	0.105	2.060	0.034	0.285	0.011
$PC2$	0.003	0.087	-0.016	-0.257	-0.009	-0.153	0.000
$PC3$	0.001	0.015	0.113	1.559	0.008	0.145	0.013
$PC4$	-0.004	-0.105	0.006	0.065	-0.051	-0.589	0.002

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	Oil	t_{Oil}	R^2
$PC1$	0.004	0.053	0.107	2.056	0.041	0.854	0.012
$PC2$	0.003	0.077	-0.018	-0.284	0.034	1.595	0.004
$PC3$	0.001	0.014	0.113	1.556	-0.011	-0.581	0.013
$PC4$	-0.003	-0.068	0.007	0.073	-0.001	-0.062	0.000

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$CNHUSD$	t_{CNHUSD}	R^2
$PC1$	0.004	0.053	0.107	2.056	0.041	0.854	0.012
$PC2$	0.000	0.011	0.046	0.724	-0.196	-1.640	0.005
$PC3$	-0.003	-0.066	0.053	0.953	-0.030	-0.244	0.003
$PC4$	-0.001	-0.023	-0.088	-1.314	-0.077	-0.719	0.009

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$EURUSD$	t_{EURUSD}	R^2
$PC1$	-0.004	-0.043	0.090	1.717	0.100	0.556	0.009
$PC2$	0.000	-0.001	0.041	0.619	0.010	0.111	0.002
$PC3$	-0.004	-0.099	0.053	0.950	0.104	1.041	0.005
$PC4$	-0.001	-0.025	-0.089	-1.325	-0.005	-0.062	0.008

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	VIX	t_{VIX}	R^2
$PC1$	0.005	0.067	0.100	1.885	0.001	0.079	0.010
$PC2$	0.005	0.120	-0.027	-0.473	0.003	0.700	0.001
$PC3$	0.001	0.028	0.093	1.394	0.000	-0.034	0.009
$PC4$	-0.004	-0.104	-0.019	-0.223	-0.004	-0.832	0.002

Panel B: Regression of Principal Components of Cryptocurrency Log RV_t on Economic Factors

The table shows output from the regression of the first four principal components of cryptocurrency log realized volatility on its lags and different economic volatility factors. S&P500 is daily log realized volatility calculated from S&P500 E-Mini Futures. Gold is the daily log realized volatility calculated from Gold Futures. Oil is daily log realized volatility calculated from Crude Oil WTI Futures. CNHUSD is daily log realized volatility calculated from daily spot rate of CNH/USD. EURUSD is daily log realized volatility calculated from daily spot rate of EUR/USD. VIX is the daily log realized volatility calculated from COBE SPX Volatility Index. All factor minutely data is downloaded from TRTH with sample period from October 2016 – November 2018 except CNHUSD and EURUSD which are not available before February 2017. The realized volatility calculation is subject to Hansen and Lunde (2005) method described under section 3.2. The principal components are constructed as the matrix of the log realized volatility for all 9 cryptocurrencies multiplied by the eigenvectors of the covariance matrix. Robust t statistics for each principle component and economic factors are shown as t_{PC} and $t_{EconomicFactor}$.

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$S\&P500$	$t_{S\&P500}$	R^2
$PC1$	-0.182	-2.703	0.868	45.488	-0.281	-3.330	0.789
$PC2$	0.028	0.677	0.610	10.189	0.051	1.000	0.379
$PC3$	0.014	0.289	0.456	8.433	0.023	0.349	0.211
$PC4$	0.018	0.474	0.448	9.343	0.028	0.689	0.201

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$Gold$	t_{Gold}	R^2
$PC1$	0.009	0.111	0.883	47.442	0.028	0.178	0.781
$PC2$	0.080	1.139	0.597	11.097	0.194	1.473	0.382
$PC3$	-0.048	-0.800	0.453	8.530	-0.113	-1.011	0.213
$PC4$	-0.064	-1.290	0.446	9.541	-0.150	-1.672	0.209

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	Oil	t_{Oil}	R^2
$PC1$	-0.017	-0.232	0.883	47.089	0.033	0.217	0.781
$PC2$	-0.081	-1.925	0.596	10.695	0.190	2.071	0.382
$PC3$	0.028	0.629	0.460	8.731	-0.067	-0.741	0.212
$PC4$	-0.004	-0.096	0.452	9.883	0.012	0.149	0.204

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$CNYUSD$	t_{CNYUSD}	R^2
$PC1$	-0.283	-1.669	0.868	39.438	-0.208	-1.682	0.781
$PC2$	-0.121	-1.372	0.500	9.903	-0.088	-1.385	0.257
$PC3$	-0.150	-1.819	0.396	6.925	-0.109	-1.961	0.169
$PC4$	0.021	0.255	0.472	7.768	0.014	0.243	0.224

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	$EURUSD$	t_{EURUSD}	R^2
$PC1$	-0.265	-1.659	0.878	42.832	-0.323	-1.791	0.781
$PC2$	-0.132	-1.437	0.502	9.909	-0.158	-1.384	0.257
$PC3$	-0.183	-1.893	0.394	6.899	-0.220	-2.001	0.172
$PC4$	-0.097	-1.108	0.471	7.743	-0.118	-1.195	0.226

	$Constant$	$t_{Constant}$	PC_{t-1}	$t_{PC_{t-1}}$	VIX	t_{VIX}	R^2
$PC1$	0.402	2.164	0.876	45.798	-0.210	-2.247	0.780
$PC2$	-0.156	-1.113	0.611	10.245	0.079	1.153	0.379
$PC3$	-0.078	-0.614	0.454	8.625	0.041	0.659	0.209
$PC4$	-0.011	-0.128	0.471	11.332	0.006	0.157	0.223

first PC of crypto volatility and both the volatility of the S&P500 E-mini future and the CBOE SPX VIX. In addition, the first principal component is also marginally negatively related to the volatility of CNY/USD and EUR/USD exchange rates.

The reason for the negative relationship between cryptocurrency volatility factors and macroeconomic indices (S&P500, VIX) is not clear. One potential explanation is that cryptocurrency is more susceptible to investor sentiment than macroeconomic factors, although the latter may influence the former (see for example Chuen et al. (2017), Corbet et al. (2018), and Drobetz et al. (2019)). High macroeconomic risk leads to more caution amongst investors and, as a consequence, less trading activity. As less trading activity results in less irrational trading, the volatility of cryptocurrencies in particular tends to decline. We leave the true underlying reason for a negative relationship between crypto volatility and macro volatility for a future study.

The negative relationship between commonality cryptocurrency volatility and that of foreign exchange is also not clear-cut. One possible reason is that cryptocurrency ultimately needs to be converted to fiat currency for at least some investors. If the major foreign exchange rates are highly volatile, cryptocurrency traders are reluctant to trade more. As a result, cryptocurrency becomes less volatile for the same reasons as outlined above.

In sum, there is strong evidence to show that both daily returns and realized volatilities of cryptocurrency cannot be explained by traditional economic factors. It is not surprising that there is almost no significant relation between cryptocurrency return and benchmark economic factor returns, as this has been addressed in extant studies of cryptocurrency returns. The lack of explanatory power for realized volatilities in cryptocurrencies contrasts with findings in other financial markets. Christoffersen et al. (2019) state that most of the macro factors they consider have a strong relation to the first component of cross-section commodity futures realized volatility. Based on their empirical evidence, the R^2 in the regression of the first PC of $log(RV_t)$ is around 70%. Comparable regressions in cryptocurrency show that the PCs can only be explained by their first lags, and not by the economic factors. The relatively high R^2 in the realized volatility PCs regression derives mainly from the lagged variable and the factor structure itself, with very little contributed by the economic or financial factors.

Therefore, despite the presence of some significant correlations, the overall relationship between macro factors and the PCs of cryptocurrency return and volatility remains relatively weak and we conclude that:

Fact 4: *Economic and financial factors do not have strong explanatory power on* *the common factors of cryptocurrency return and volatility and there is a weak inverse relation between risk of cryptocurrency and macroeconomic indices.*

3.5 Bitcoin Impact on Cryptocurrency Return and Volatility

3.5.1 Bitcoin as a Fundamental Factor in the Cryptocurrency market

In this section, we investigate whether the behavior of Bitcoin can be thought of as a fundamental factor to explain the time series variation of the PCs of cryptocurrency return and volatility. Products traded on Bittrex are mainly cryptocurrencies quoted against Bitcoin. This trading feature is very similar to a foreign exchange, as one is the base currency and the other is the counter currency. Bitcoin as a counter currency is a very liquid product in the crypto markets and it is reasonable to hypothesize that fluctuations in the Bitcoin price against the dollar have an impact on other cryptocurrencies.

On average, Bitcoin return has a weak negative correlation with all other cryptocurrency returns shown in Table 3.4. In the meantime, the $log(RV_t)$ of Bitcoin is positively correlated with other cryptocurrency $log(RV_t)$ shown in Table 3.5. We first run regressions of the time series of Bitcoin returns and realized volatility on time series of PCs (from returns and volatility). Table 3.8 shows the results. The sign of regression coefficient on both first principal components is as expected and significant at the 5% level. Other PCs are also statistically significant; however, the R^2 values are quite small at just 7.6% for the return regression and 6.2% for the volatility regression, which does not suggest a strong relationship. It appears the common components of returns (or realized volatilities) in the cryptocurrencies are weakly related to Bitcoin returns (and volatility).

As an alternative approach, we test whether Bitcoin adds explanatory power over and above the principal components for the returns and volatilities of individual cryptocurrencies. For each cryptocurrency, we regress its return (or volatility) on the first four PCs and the return (or volatility) of Bitcoin. As usual, the first four PCs are computed by taking the other eight cryptocurrency returns or volatilities. We orthogonalize each PC by regressing it on the relevant Bitcoin variable and taking the residuals. This gives Bitcoin the maximum possible chance of explaining the returns and volatilities of the individual cryptocurrencies. The regression model is as follows:

$$log(Return/RV_{i,t}) = \alpha + \beta_1 PC_{i,t} + \beta_2 log(Return/RV_{BTC,t}) + \epsilon_t \tag{3.7}$$

TABLE 3.8: Regression of BTC Return and RV_t on Principal Components

Panel A in the table shows the regression of Bitcoin return and volatility on principal components of the other 9 cryptocurrencies. Panel B is the same regression with Bitcoin bubble detection. The principle components are constructed as the matrix of the log realized volatility or returns for all 9 cryptocurrencies multiplied by the eigenvectors of the covariance matrix. Robust t statistics for each principle component are shown as t_{PC}.

Panel A : Pooled Regression of BTC return and RV on PCs from the other 9 Cryptocurrency

	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
BTCReturn	-0.4089	-2.554	0.2288	0.975	-0.6063	-2.901	0.8494	3.659	7.60%
BTC RV_t	0.0375	2.297	-0.1698	-3.26	-0.007	-0.095	-0.1489	-2.567	6.20%

Panel B Sub-group Regression of BTC return and RV on PCs from the other 9 Cryptocurrency

	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
Pre-Bubble									
BTCReturn	-0.719	-3.59	-0.424	-1.221	0.0406	0.0105	0.4883	2.001	16.80%
BTC RV_t	0.0539	3.515	-0.2608	-4.425	0.0813	1.272	-0.0989	-1.41	12.30%
Bubble									
BTCReturn	-0.9221	-3.917	-0.8835	-2.347	0.7326	1.148	0.0793	0.205	15.80%
BTC RV_t	0.0703	2.792	-0.1337	-2.579	0.0286	0.481	-0.0084	-0.115	13.20%
Post Bubble									
BTCReturn	0.7384	5.352	-0.8717	-3.361	0.8405	2.8	1.0039	3.524	24.80%
BTC RV_t	-0.028	-1.152	-0.2156	-3.884	-0.1365	0.126	-0.1194	-1.594	8.10%

Table 3.9 gives the regression results. Panel A of Table 3.9 shows all cryptocurrency returns are significantly negatively related to the return of Bitcoin with the exceptions of STRAT and LSK. Panel B of Table 3.9 reports the regression results of each cryptocurrency volatility. All cryptocurrencies are significantly positively related to $log(RV_t)$ of Bitcoin, though the level of significance differs across the cryptocurrencies. Each of ETH, XRP, LTC, ETC, DASH and LSK are significant at the 1% level, XMR is significant at 5%, while STRAT and ZEC are only significant at the 10% level. Nevertheless, the goodness of fit statistics for each regression are only slightly increased from those reported in Table 3.6. It appears that while Bitcoin captures some information relevant to explaining returns and volatilities of cryptocurrency i, the other eight cryptocurrencies themselves already contain much of this information already.

So far, we conclude that:

Fact 5: *Bitcoin can be considered for most cryptocurrencies as a fundamental factor able to explain a small proportion of the variations in return and volatility.*

3.5.2 Bitcoin Bubble Impact in Cryptocurrency Return and Volatility

In our sample period, the Bitcoin price from the Coinbase exchange climbed from \$615.65 on October 1, 2016 to the peak price of \$19650.01 on December 16, 2017 (see Figure 3.10). The Bitcoin price increased almost 32-fold in only 6 months. After reaching its peak price, Bitcoin tumbled until February 2018. On February 5, 2018 the

TABLE 3.9: Regression of Cryptocurrency Return and RV_t on BTC Return and RV_t

Panel A in the table shows parameter estimates of return regressed on principal components of 9 cryptocurrencies and the Bitcoin daily return during the October 2016 - November 2018. Panel B in the table shows parameter estimates of log realized volatility regressed on principal components of 9 cryptocurrencies and the Bitcoin daily log realized volatility during the October 2016 - November 2018. Noted that, for each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Besides, we take residuals from the equation (3.7) as principal components from the other 8 cryptocurrencies and orthogonalize it to data from Bitcoin. Robust t statistics for each principle component and Bitcoin are shown as t_{PC} and t_{BTC}.

Panel A: Regression of Cryptocurrency Returns on BTC and PCs

	$Constant$	$t_{Constant}$	BTC	t_{BTC}	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.116	0.758	-0.229	-5.523	1.741	13.264	-0.030	-0.128	-0.316	-1.402	-0.275	-1.023	0.377
XRP	0.353	1.163	-0.315	-5.178	1.496	6.060	-0.361	-0.606	-0.285	-0.645	-1.297	-2.079	0.138
STRAT	0.092	0.346	0.076	1.261	2.012	9.032	0.073	0.205	-0.170	-0.400	0.010	0.022	0.218
LTC	0.065	0.358	-0.129	-2.539	1.386	7.769	-0.249	-0.972	-0.768	-2.407	-0.204	-0.657	0.230
ETC	0.026	0.147	-0.163	-3.425	1.987	16.486	-0.128	-0.457	0.961	2.930	-1.365	-2.297	0.378
DASH	0.124	0.780	-0.217	-3.615	1.651	11.920	0.405	0.892	1.069	3.065	-0.509	-1.557	0.330
ZEC	-0.524	-1.798	-0.246	-6.149	2.012	18.854	0.390	1.597	0.873	2.746	-0.756	-2.533	0.292
LSK	0.051	0.222	-0.034	-0.458	2.145	10.959	0.132	0.308	-1.026	-2.217	-0.128	-0.341	0.292
XMR	0.131	0.792	-0.120	-4.011	1.702	15.376	0.233	0.665	-0.772	-2.290	0.763	1.867	0.335

Panel B: Regression of Cryptocurrency Log RV_t on BTC and PCs

	$Constant$	$t_{Constant}$	BTC	t_{BTC}	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.953	18.301	0.162	4.483	0.261	28.889	-0.066	-1.726	0.113	3.151	-0.006	-0.151	0.713
XRP	1.193	27.488	0.236	6.547	0.233	19.311	0.115	3.197	-0.072	-1.904	0.011	0.253	0.619
STRAT	1.703	32.504	0.063	1.694	0.222	23.052	-0.004	-0.097	0.090	2.068	-0.007	-0.187	0.580
LTC	1.109	26.095	0.126	3.866	0.251	24.184	-0.141	-3.321	0.120	3.929	-0.009	-0.263	0.693
ETC	1.399	35.105	0.107	3.807	0.227	22.804	-0.052	-2.226	0.038	1.181	0.008	0.242	0.673
DASH	1.330	30.397	0.070	2.416	0.243	29.497	-0.036	-1.269	0.075	2.881	0.004	0.107	0.699
ZEC	1.535	28.570	0.054	1.701	0.211	16.328	-0.151	-3.281	0.067	1.965	-0.003	-0.072	0.567
LSK	1.708	27.419	0.098	2.564	0.235	23.938	0.023	0.748	-0.086	-2.305	-0.002	-0.038	0.618
XMR	1.428	32.615	0.055	1.956	0.211	24.333	-0.076	-2.094	0.074	2.170	-0.010	-0.206	0.616

price closed at \$6905.19, its lowest point after the price peak. It is important to examine the impact of the Bitcoin bubble on the behavior of other cryptocurrencies. The weak relationship highlighted above between Bitcoin and other cryptocurrencies could be due to shifts in the underlying relationships across the different Bitcoin periods.

Our initial aim is to find the bubble's origin and burst dates. We follow the approach of Phillips et al. (2011) and Phillips and Yu (2011) using forward recursive regressions to calculate Dickey-Fuller (DF) t statistics that can then be compared to the critical value of the DF test defined in their paper. Figure 3.11 plots the DF t statistics and critical values (see Appendix C.1 for detailed calculations). We adopt the definition of the burst of the bubble from the Phillips et al. (2011) paper and define the bubble burst date to be the last date on which the DF statistic is greater than the DF critical value. We keep the origin of the bubble as the earliest date on which the DF statistic is greater than the DF critical value.[11] Based on this method, the Bitcoin bubble began on May 24, 2017 and ended on January 28, 2018. Our bubble period naturally includes the price peak of December 16, 2017. The bubble lasted 250 days, and contains around one-third of the data in the sample period, allowing us to

[11]The test statistics often dropped below or rose above the relevant critical values between these dates. In real time, dating the bubble with this approach would have been difficult, but our interest is in historically dating the bubble solely in order to split our sample into pre-bubble, bubble and post-bubble periods.

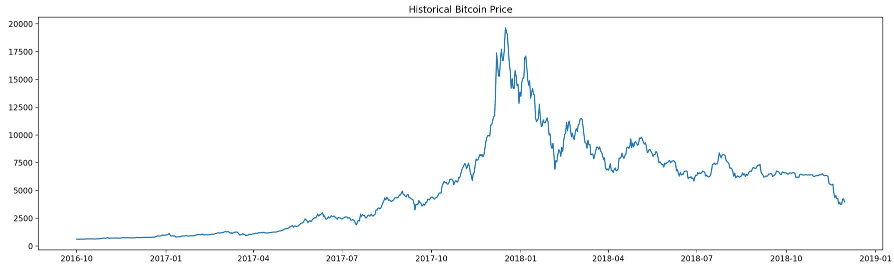


FIGURE 3.10: Historical Price of Bitcoin from Oct. 2016 to Nov. 2018

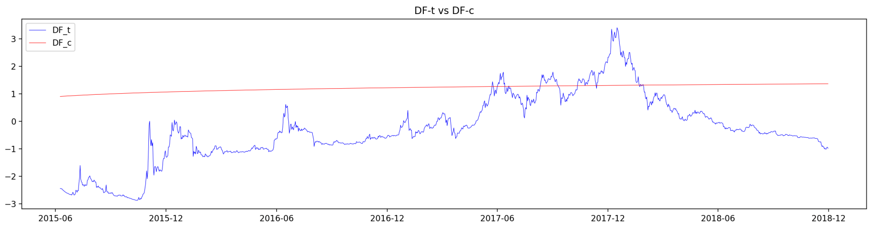


FIGURE 3.11: Bitcoin Price Bubble Test

define three sub-periods: the period before May 24, 2017 is defined as the pre-bubble period; May 24, 2017 through January 28, 2018 is the bubble period; and the interval after January 28, 2018 is the post-bubble period.

We now re-conduct regression analysis of Bitcoin factors in these three sub-periods. Table 3.10 reports the results of regressing cryptocurrency returns on the first four PCs.[12] There are clear indications that the factor structure is a powerful way to explain the variation in cryptocurrency returns and that bubble-related dynamics are important. All R^2 values from the bubble period are significantly higher than in the pre-bubble period. It should be noted that the XRP and STRAT R^2 values are only 5% and 8.6% respectively in the pre-bubble period and that these both increase to 25% during the bubble. Once the bubble had burst, the R^2 figures remain close to or, in some cases, above the same statistics from the bubble interval. The simple average R^2 values across nine cryptocurrencies are 18%, 40%, 36% from the pre-bubble, bubble and post-bubble respectively. In sum, the commonality in cryptocurrency returns is stronger during and, to a large extent, after the Bitcoin bubble.

Table 3.11 reports the regression of each cryptocurrency's volatility on the first four volatility PCs. The explanatory power of PCs in the volatility regressions using the full sample were higher than for returns, and this survives splitting the sample into sub-periods. Explanatory power again increases from the pre-bubble to the bubble

[12]Again, the PCs are computed by the other eight currencies to avoid an endogeneity issue.

Table 3.10: Regression of Cryptocurrency Return on PCs during Pre-Bubble, Bubble, and Post Bubble Periods

The table shows parameter estimates of daily return regressed on principal components of 9 cryptocurrencies conditional on Bitcoin bubble problem during the October 2016 - November 2018. For each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Robust t statistics for each principle component are shown as t_{PC}.

Panel A: Pre-Bubble	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	2.073	5.939	-0.143	-0.416	0.551	1.147	-0.955	-1.842	0.255
XRP	0.498	0.742	0.487	0.478	1.372	1.761	2.302	1.923	0.050
STRAT	1.507	3.577	0.129	0.307	-0.207	-0.493	0.794	1.328	0.086
LTC	1.095	3.291	0.669	1.052	1.082	2.421	0.405	0.764	0.103
ETC	2.229	7.872	0.103	0.204	-0.860	-2.103	-1.484	-3.936	0.348
DASH	1.529	4.268	-0.567	-0.688	0.560	1.536	1.406	2.271	0.189
ZEC	2.064	6.368	-0.310	-0.545	0.842	1.078	-1.887	-2.606	0.132
LSK	1.869	6.094	-0.078	-0.166	-0.681	-1.572	-1.466	-2.518	0.232
XMR	1.791	4.899	-0.088	-0.237	0.619	1.560	0.346	0.462	0.225

Panel B: Bubble	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	2.231	12.279	-0.523	-1.703	-0.218	-0.437	0.209	0.509	0.499
XRP	2.320	6.446	0.716	0.775	0.480	0.524	-1.181	-1.073	0.251
STRAT	2.428	5.290	-0.486	-0.540	-0.416	-0.447	-0.116	-0.128	0.248
LTC	1.841	6.386	-0.552	-1.071	0.342	0.661	-1.222	-2.362	0.337
ETC	2.548	14.651	1.105	1.592	1.958	2.500	-1.705	-1.216	0.465
DASH	2.210	10.537	0.968	1.631	0.325	0.863	-1.420	-2.202	0.422
ZEC	2.796	19.108	0.986	2.179	0.337	0.961	-0.432	-1.011	0.617
LSK	2.816	8.419	-0.600	-0.525	0.481	0.623	0.954	0.831	0.343
XMR	2.234	11.247	1.283	1.639	-1.134	-1.540	0.782	1.023	0.436

Panel C: Post-Bubble	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	1.009	12.325	-0.346	-2.724	-0.635	-4.363	-0.249	-1.684	0.529
XRP	1.280	8.669	-0.318	-1.782	0.714	2.362	-0.263	-0.932	0.364
STRAT	1.313	13.503	-0.205	-1.064	0.401	1.545	0.448	1.819	0.379
LTC	0.838	12.442	0.171	1.162	-0.351	-2.308	0.282	1.739	0.351
ETC	0.979	9.476	0.337	1.467	-0.557	-2.053	-0.506	-1.753	0.268
DASH	1.053	9.927	0.481	2.151	-0.128	-0.881	-0.220	-1.383	0.425
ZEC	1.029	11.937	-0.317	-1.669	0.040	0.224	-0.415	-1.987	0.292
LSK	1.192	8.734	0.008	0.032	0.507	1.803	0.764	3.123	0.281
XMR	0.763	8.456	0.200	1.036	-0.603	-3.152	-0.443	-1.911	0.278

period for volatility (from an average of 48 % to 56%) and continues to rise in the post-bubble period (averaging 71%). Consequently, based on the findings here, we conclude that:

Fact 6: *The Factor Structure model is more powerful in explaining variation in returns and volatilities during the Bitcoin bubble period and this explanatory power persists - and, for volatilities actually increases further - after the Bitcoin bubble burst.*

Table 3.11: Regression of Cryptocurrency Log RV_t on PCs during Pre-Bubble, Bubble, and Post Bubble Periods

The table shows parameter estimates of log realized volatility regressed on principal components of 9 cryptocurrencies conditional on Bitcoin bubble problem during the October 2016 - November 2018. For each cryptocurrency, we re-conduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Robust t statistics for each principle component are shown as t_{PC}.

Panel A: Pre-Bubble									
	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.263	12.558	0.092	1.588	-0.131	-2.18	-0.017	-0.334	0.595
XRP	0.218	8.77	0.174	3.626	-0.078	-1.193	-0.041	-0.757	0.54
STRAT	0.172	5.784	0.075	1.158	0.035	0.383	-0.119	-1.248	0.27
LTC	0.236	11.948	-0.253	-6.622	0.157	3.386	0.097	2.369	0.647
ETC	0.17	9.468	-0.062	-2.074	-0.06	-1.285	0.031	0.756	0.457
DASH	0.281	19.436	-0.025	-0.866	-0.032	-0.673	-0.024	-0.428	0.64
ZEC	0.146	5.608	-0.242	-3.707	-0.049	-0.718	-0.141	-1.928	0.32
LSK	0.139	6.077	0.068	2.028	-0.027	-0.578	0.002	0.034	0.343
XMR	0.196	10.549	0.158	2.801	-0.036	-0.641	0.054	0.93	0.476

Panel B: Bubble									
	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.255	16.831	-0.072	-1.597	0.063	1.487	0.106	2.467	0.670
XRP	0.153	6.277	0.037	0.695	0.066	1.085	0.031	0.506	0.319
STRAT	0.142	8.856	-0.001	-0.035	-0.050	-1.059	0.000	-0.005	0.401
LTC	0.246	15.949	-0.005	-0.084	-0.029	-0.614	-0.059	-1.190	0.642
ETC	0.211	13.524	0.027	0.849	-0.046	-1.077	-0.167	-3.797	0.638
DASH	0.201	13.631	0.120	3.094	-0.148	-4.476	-0.067	-1.634	0.614
ZEC	0.205	11.441	0.071	1.942	0.007	0.228	-0.038	-1.027	0.600
LSK	0.204	15.792	-0.055	-1.558	0.034	0.619	0.144	2.293	0.571
XMR	0.194	11.643	0.017	0.284	-0.110	-2.786	-0.102	-1.503	0.569

Panel C: Post Bubble									
	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.232	22.335	0.021	0.493	-0.055	-1.285	0.007	0.134	0.759
XRP	0.213	14.561	0.103	2.570	-0.021	-0.481	-0.018	-0.270	0.660
STRAT	0.215	26.727	-0.016	-0.556	0.046	1.269	0.044	1.246	0.785
LTC	0.226	21.649	0.006	0.142	0.058	1.380	-0.101	-2.335	0.730
ETC	0.210	17.020	-0.002	-0.049	0.021	0.498	0.015	0.312	0.693
DASH	0.209	22.442	0.103	3.089	0.031	0.852	0.050	1.173	0.778
ZEC	0.182	10.803	0.088	2.141	0.039	0.933	-0.011	-0.232	0.613
LSK	0.171	12.400	0.090	2.254	-0.056	-1.122	-0.016	-0.301	0.591
XMR	0.216	24.821	0.078	1.434	0.024	0.493	0.033	0.705	0.747

3.5.3 The Shifting Relationship between Variation in Cryptocurrency and Bitcoin on Returns and RV

Next, we investigate whether the Bitcoin pricing bubble affects the abilities of PCs of returns or volatility to explain the time series variation of Bitcoin returns and volatility. Panel B of Table 3.8 reports results from the regression of Bitcoin return (or volatility) on PCs during the three sub-periods. Compared with the full sample results given in Panel A, the sub-period results suggest considerable instability in coefficients. In particular, coefficient signs on the first PC flip in the post-bubble period for both returns and volatility. Not surprisingly, therefore, sub-period R^2 figures are much higher than the apparently mis-specified full sample regression.

Table 3.12: Regression of Cryptocurrency Return on PCs and BTC during Pre-Bubble, Bubble, and Post Bubble Periods

The table shows parameter estimates of daily returns regressed on principal components of 9 cryptocurrencies and the Bitcoin daily return conditional on Bitcoin bubble problem during the October 2016 - November 2018. Noted that, for each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. Also, we take residuals from the equation (3.7) as principal components from the other 8 cryptocurrencies and orthogonalize it to data from Bitcoin. Robust t statistics for each principle component and Bitcoin are shown as t_{PC} amd t_{BTC}.

Panel A: Pre-Bubble

	Constant	$t_{Constant}$	*BTC*	t_{BTC}	*PC*1	t_{PC1}	*PC*2	t_{PC2}	*PC*3	t_{PC3}	*PC*4	t_{PC4}	R^2
ETH	1.073	2.818	-0.557	-3.495	1.893	5.705	-0.249	-0.721	0.591	1.196	-0.828	-1.818	0.268
XRP	1.787	2.386	-0.858	-4.671	-0.087	-0.128	0.446	0.513	1.540	1.963	1.956	1.597	0.093
STRAT	1.419	2.382	-0.095	-0.644	1.668	3.550	0.224	0.471	-0.236	-0.548	0.763	1.298	0.092
LTC	0.669	1.530	-0.438	-3.430	0.926	2.843	0.594	0.934	0.971	2.149	0.497	0.919	0.114
ETC	0.894	2.791	-0.397	-4.146	2.244	7.623	0.111	0.216	-0.855	-2.095	-1.494	-3.869	0.348
DASH	1.032	2.776	-0.438	-1.704	1.322	3.116	-0.679	-0.963	0.710	1.803	1.409	2.406	0.206
ZEC	-1.033	-1.060	-0.586	-2.929	1.878	6.307	-0.423	-0.776	0.906	1.163	-1.780	-2.301	0.137
LSK	0.596	1.589	-0.537	-3.565	1.638	5.536	-0.211	-0.542	-0.568	-1.379	-1.586	-3.113	0.252
XMR	0.826	2.226	-0.319	-2.285	1.769	4.304	-0.100	-0.251	0.614	1.528	0.359	0.493	0.225

Panel B: Bubble

	Constant	$t_{Constant}$	*BTC*	t_{BTC}	*PC*1	t_{PC1}	*PC*2	t_{PC2}	*PC*3	t_{PC3}	*PC*4	t_{PC4}	R^2
ETH	0.373	1.364	-0.392	-6.437	2.073	10.380	-0.691	-2.334	-0.110	-0.219	0.236	0.604	0.518
XRP	0.220	0.393	-0.491	-6.135	2.144	5.460	0.583	0.610	0.340	0.341	-0.997	-0.890	0.261
STRAT	0.096	0.165	-0.166	-1.448	2.593	5.341	-0.371	-0.411	-0.325	-0.330	0.011	0.013	0.255
LTC	0.216	0.549	-0.226	-2.710	1.854	5.766	-0.539	-1.003	0.345	0.665	-1.230	-2.365	0.337
ETC	-0.032	-0.088	-0.311	-4.304	2.492	11.685	1.169	1.677	1.978	2.549	-1.720	-1.222	0.466
DASH	0.306	1.061	-0.417	-4.558	2.056	10.964	0.811	1.563	0.430	1.095	-1.466	-2.269	0.437
ZEC	-0.096	-0.349	-0.429	-9.204	2.703	15.862	0.897	1.995	0.404	1.162	-0.426	-1.011	0.622
LSK	0.703	1.388	-0.187	-1.498	3.024	9.185	-0.457	-0.416	0.351	0.455	1.005	0.878	0.355
XMR	0.267	0.772	-0.339	-7.185	2.206	10.587	1.245	1.571	-1.137	-1.545	0.780	1.026	0.437

Panel C: Post Bubble

	Constant	$t_{Constant}$	*BTC*	t_{BTC}	*PC*1	t_{PC1}	*PC*2	t_{PC2}	*PC*3	t_{PC3}	*PC*4	t_{PC4}	R^2
ETH	-0.404	-3.002	0.119	3.013	1.026	12.530	-0.339	-2.567	-0.618	-3.810	-0.225	-1.558	0.530
XRP	-0.034	-0.141	0.149	2.485	1.320	9.671	-0.290	-1.612	0.657	2.107	-0.301	-1.097	0.366
STRAT	-0.472	-2.641	0.473	10.415	1.111	9.401	-0.001	-0.005	0.221	0.929	0.174	0.702	0.451
LTC	-0.204	-1.576	0.127	3.120	0.838	11.495	0.171	1.248	-0.350	-2.042	0.282	1.816	0.351
ETC	-0.236	-1.140	0.140	2.587	0.996	9.667	0.314	1.377	-0.574	-2.148	-0.489	-1.743	0.268
DASH	-0.318	-2.434	0.145	4.617	1.113	10.245	0.575	2.580	-0.080	-0.540	-0.161	-0.972	0.432
ZEC	-0.166	-0.836	0.181	3.678	1.034	10.657	-0.324	-1.640	0.046	0.251	-0.412	-2.060	0.292
LSK	-0.435	-1.699	0.372	5.916	1.058	8.423	0.172	0.683	0.505	1.844	0.512	2.042	0.303
XMR	-0.122	-0.823	0.272	9.241	0.646	7.125	0.025	0.138	-0.640	-3.403	-0.278	-1.331	0.316

The explanation for the shifting relationship between common factors of cryptocurrency and the variation in Bitcoin return and volatility is not clear. However, it does suggest that the shifting fundamental behavior of Bitcoin after the bubble burst is important. Bitcoin became considerably less volatile in the third quarter 2018 yet the other nine cryptocurrencies remained highly volatile.

Further analysis is required to determine whether the Bitcoin bubble had a significant impact on the relationship between Bitcoin returns and volatility and each cryptocurrency's return and volatility. Table 3.12 shows the regression results of the cryptocurrency returns on Bitcoin returns and the first four orthogonalized PCs during the pre-bubble, bubble and post-bubble periods. The results are broadly similar to the regressions without adding the Bitcoin return, as shown in Table 3.10, and the R^2 figures are barely changed.

More interestingly, we see that after the bubble bursts, the relationship between Bitcoin returns and the returns of each cryptocurrency has significantly changed. In the pre-bubble and bubble periods, the relationship is negative and significant for most cryptocurrencies except STRAT (not significant in pre-bubble and bubble periods), LSK (not significant in the bubble period). However, all cryptocurrency returns are positive and significant at the 1% level in the post-bubble period. The relatively weak relationship noted above for the full sample regression is in part due to this structural shift.

Table 3.13 reports the results of volatility regression considering $log(RV_t)$ of Bitcoin and the first four orthogonalized PCs. Not all cryptocurrency $log(RV_t)$ are significantly related to Bitcoin RV before the bubble. For example, STRAT is not significant at all and LSK is only positively significant at the 10% level. The nature of the positive relationship between Bitcoin volatility and other cryptocurrencies strengthens during the Bitcoin bubble period and all cryptocurrencies' volatilities are strongly positively significant at the 1% level. Again, though, we see that the relationship between Bitcoin volatility and that of other cryptocurrencies is reversed and less significant after the bubble burst.

We also see that the change in the relationships for return and volatility are inverted. Bitcoin return becomes positively related to those of the other cryptos, while Bitcoin volatility becomes negatively related with other crypto volatilities post-bubble. The reason for the relationship shifting between the Bitcoin and Cryptocurrency return and volatility as the Bitcoin pricing bubble burst is unclear. Nevertheless, we conclude that:

Fact 7: *There is heterogeneity in the relationship between Bitcoin and other cryptocurrencies for both returns and volatility after the Bitcoin pricing bubble burst.*

3.6 Realized Cryptocurrency Beta and Systematic Risk Ratio

In this section, we study realized covariance between the nine cryptocurrencies and Bitcoin. As we found in section 3.5, Bitcoin acts as a (weak) fundamental factor in addition to PCs from the cryptocurrencies. Furthermore, Bitcoin captures almost 55% of the market value in cryptocurrency. We seek to test whether the role of Bitcoin is that of a market index proxy. Therefore, we compute "market"-style betas in the cryptocurrency market using Bitcoin as the market proxy. Given the demonstrated impact of the Bitcoin pricing bubble in our sample, we compute the dynamic, model-free, realized betas with our high-frequency returns.

TABLE 3.13: Regression of Cryptocurrency Log RV_t on PCs and BTC during Pre-Bubble, Bubble, and Post Bubble Periods

The table shows parameter estimates of log realized volatility regressed on principal components of 9 cryptocurrencies and the Bitcoin log realized volatility conditional on Bitcoin bubble problem during the October 2016 - November 2018. Noted that, for each cryptocurrency, we reconduct a principal component analysis based only on the other 8 cryptocurrencies, to avoid endogeneity issues in the regressions. In addition, we take residuals from the equation (3.7) as principal components from the other 8 cryptocurrencies and orthogonalize it to data from Bitcoin. Robust t statistics for each principle component and Bitcoin are shown as t_{PC} amd t_{BTC}.

Panel A: Pre-Bubble	$Constant$	$t_{Constant}$	BTC	t_{BTC}	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	1.095	11.998	0.399	5.984	0.252	11.752	0.077	1.332	-0.117	-2.085	-0.029	-0.556	0.600
XRP	1.398	25.138	0.463	8.864	0.187	6.836	0.129	2.698	-0.118	-1.951	-0.011	-0.226	0.599
STRAT	2.086	20.478	-0.046	-0.458	0.224	7.832	0.023	0.480	0.081	0.940	-0.171	-2.036	0.358
LTC	1.212	19.086	0.426	8.255	0.221	10.107	-0.227	-5.760	0.170	4.027	0.080	1.843	0.658
ETC	1.671	25.788	0.192	4.064	0.180	9.558	-0.077	-2.530	-0.074	-1.560	0.047	1.191	0.464
DASH	1.445	19.300	0.362	6.421	0.272	14.202	-0.012	-0.484	-0.022	-0.480	-0.035	-0.615	0.643
ZEC	1.814	20.439	0.146	2.326	0.132	4.455	-0.271	-4.263	-0.041	-0.603	-0.139	-1.896	0.329
LSK	2.200	31.956	0.094	1.790	0.149	6.153	0.055	1.560	-0.039	-0.832	-0.017	-0.292	0.351
XMR	1.516	20.285	0.187	3.382	0.194	9.407	0.160	2.796	-0.035	-0.615	0.053	0.916	0.476

Panel B: Bubble	$Constant$	$t_{Constant}$	BTC	t_{BTC}	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.699	5.961	0.399	5.405	0.249	16.440	-0.060	-1.303	0.059	1.497	0.105	2.651	0.675
XRP	1.158	9.885	0.379	5.344	0.138	5.463	0.022	0.394	0.057	0.899	0.038	0.615	0.353
STRAT	1.708	19.771	0.220	4.251	0.138	8.250	-0.009	-0.235	-0.049	-1.087	-0.003	-0.055	0.404
LTC	1.015	13.025	0.244	4.633	0.251	14.193	-0.013	-0.250	-0.027	-0.559	-0.060	-1.270	0.644
ETC	1.238	16.503	0.290	6.565	0.209	11.949	0.023	0.734	-0.046	-1.047	-0.165	-3.855	0.638
DASH	1.190	12.343	0.221	4.007	0.198	13.962	0.124	3.122	-0.153	-4.692	-0.068	-1.680	0.616
ZEC	1.420	12.406	0.200	3.311	0.207	12.124	0.067	1.730	0.008	0.261	-0.038	-1.018	0.601
LSK	1.806	15.333	0.180	2.810	0.209	15.481	-0.047	-1.259	0.032	0.601	0.146	2.398	0.574
XMR	1.249	14.201	0.241	4.636	0.192	10.087	0.021	0.346	-0.111	-2.863	-0.103	-1.526	0.569

Panel C: Post Bubble	$Constant$	$t_{Constant}$	BTC	t_{BTC}	$PC1$	t_{PC1}	$PC2$	t_{PC2}	$PC3$	t_{PC3}	$PC4$	t_{PC4}	R^2
ETH	0.934	17.020	-0.118	-2.628	0.231	23.267	0.029	0.748	-0.051	-1.248	0.012	0.239	0.761
XRP	1.064	17.401	-0.028	-0.664	0.215	14.643	0.095	2.476	-0.016	-0.362	-0.009	-0.129	0.662
STRAT	1.434	36.552	-0.056	-2.047	0.216	27.929	-0.008	-0.252	0.052	1.431	0.051	1.457	0.787
LTC	1.054	27.154	-0.107	-2.801	0.226	21.755	0.005	0.119	0.058	1.378	-0.100	-2.280	0.731
ETC	1.230	31.106	-0.041	-0.941	0.211	17.279	-0.013	-0.279	0.029	0.660	0.014	0.276	0.696
DASH	1.283	39.734	-0.178	-6.421	0.206	23.075	0.083	2.746	0.047	1.355	0.063	1.716	0.793
ZEC	1.336	30.178	-0.067	-1.929	0.182	10.851	0.097	2.156	0.043	1.014	-0.017	-0.347	0.614
LSK	1.271	34.565	0.030	1.065	0.174	13.746	0.079	2.052	-0.037	-0.744	-0.035	-0.756	0.605
XMR	1.444	47.315	-0.168	-5.754	0.214	26.351	0.063	1.164	0.014	0.299	0.031	0.648	0.755

3.6.1 Realized Covariance Construction

We calculate 1-minute log returns each day based on log mid-prices. We then compute overlapping[13] 5-minute realized covariances between cryptocurrency i and Bitcoin as:

$$RCov_t^{oc} = \frac{n}{5(n-4)} \sum_{k=1}^{n-4} \widetilde{r}_{crypto,t_k} \widetilde{r}_{BTC,t_k} \tag{3.8}$$

After merging data for Bitcoin and cryptocurrency i, there are again long trading breaks that we solve using the Hansen and Lunde (2005) method. The close to open return for cryptocurrency i and Bitcoin are denoted by $r_{crypto,t}^{co}$ and $r_{BTC,t}^{co}$ respectively. Due to the variety of data breaks on different trading days, we use the simulated data from perfect days to calculate the optimal weights subject to different breaking timings

[13] For a more detailed method regarding overlapping trading spans, see Barndorff-Nielsen and Shephard (2004).

and duration. The final calculation is as follows:

$$RCov_{i,t}(w) = \hat{w}_1 \sum_{b=1}^{B} r^{co}_{b,crypto,t} r^{co}_{b,BTC,t} + \hat{w}_2 \sum_{b=1}^{B+1} RCov^{oc}_{b,t} \tag{3.9}$$

Recalling the bubble analysis in section 3.5, we calculate daily model-free realized betas for each cryptocurrency. We follow studies byAndersen et al. (2005) and Patton and Verardo (2012) and the realized beta is defined as:

$$R\beta_{i,t} = \frac{RCov_{i,t}}{RV_{BTC,t}} \tag{3.10}$$

The realized covariance $RCov_{i,t}$ is a cross-product of the intraday cryptocurrency return and the Bitcoin return estimated by either equation (3.4) or (3.5) based on whether a day has trading breaks.

Fact 8: *Cryptocurrency betas with Bitcoin were negative before the Bitcoin bubble burst but became positive after the bubble burst.*

Figure 3.12 plots the daily model-free realized betas for the nine cryptocurrencies. The red line in each plot is a 99% confidence interval. Figure 3.12 shows clearly that realized betas are negative until February 2018. As the Bitcoin bubble bursts, almost all realized betas rise towards zero before trending upwards from April 2018. Therefore, we state that:

The realized beta measures the systematic risk of a cryptocurrency in comparison to the benchmark Bitcoin as a proxy for the cryptocurrency market factor. However, the measurement of beta does not mean we can directly suggest the extent to which the variation in cryptocurrency returns is driven by the variation of Bitcoin as a fundamental factor. We follow the Christoffersen et al. (2019) study to calculate Systematic Risk Ratio (SRR) for cryptocurrency i as :

$$SSR_{i,t} = \frac{R\beta^2_{i,t}\ RV_{BTC,t}}{RV_{i,t}} \tag{3.11}$$

Based on the definition of SRR, this ratio gives the fraction of cryptocurrency i's variance explained by Bitcoin's variance. By using intraday high frequency data, we calculate the daily systematic risk ratio for each cryptocurrency throughout the sample period. Figure 3.13 plots the SSR for each cryptocurrency and the red line is the upper bound of the 99% confidence interval of SSR. There is clear evidence that the Bitcoin variance is a powerful way to explain the cryptocurrency variance during the bubble period. In fact, the SSR reaches its highest level near the peak of the Bitcoin bubble

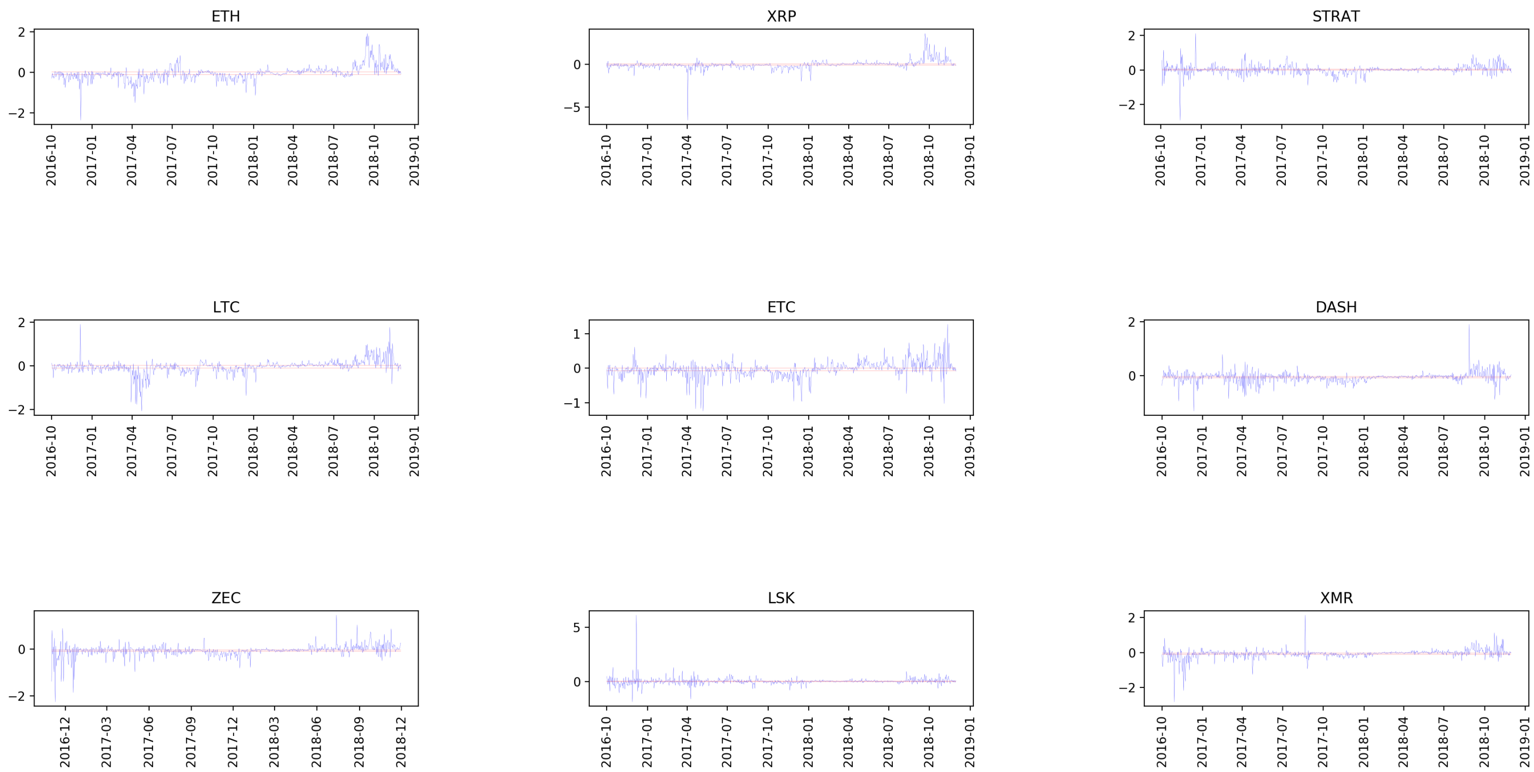


FIGURE 3.12: Cyrptocurreny Realized Betas

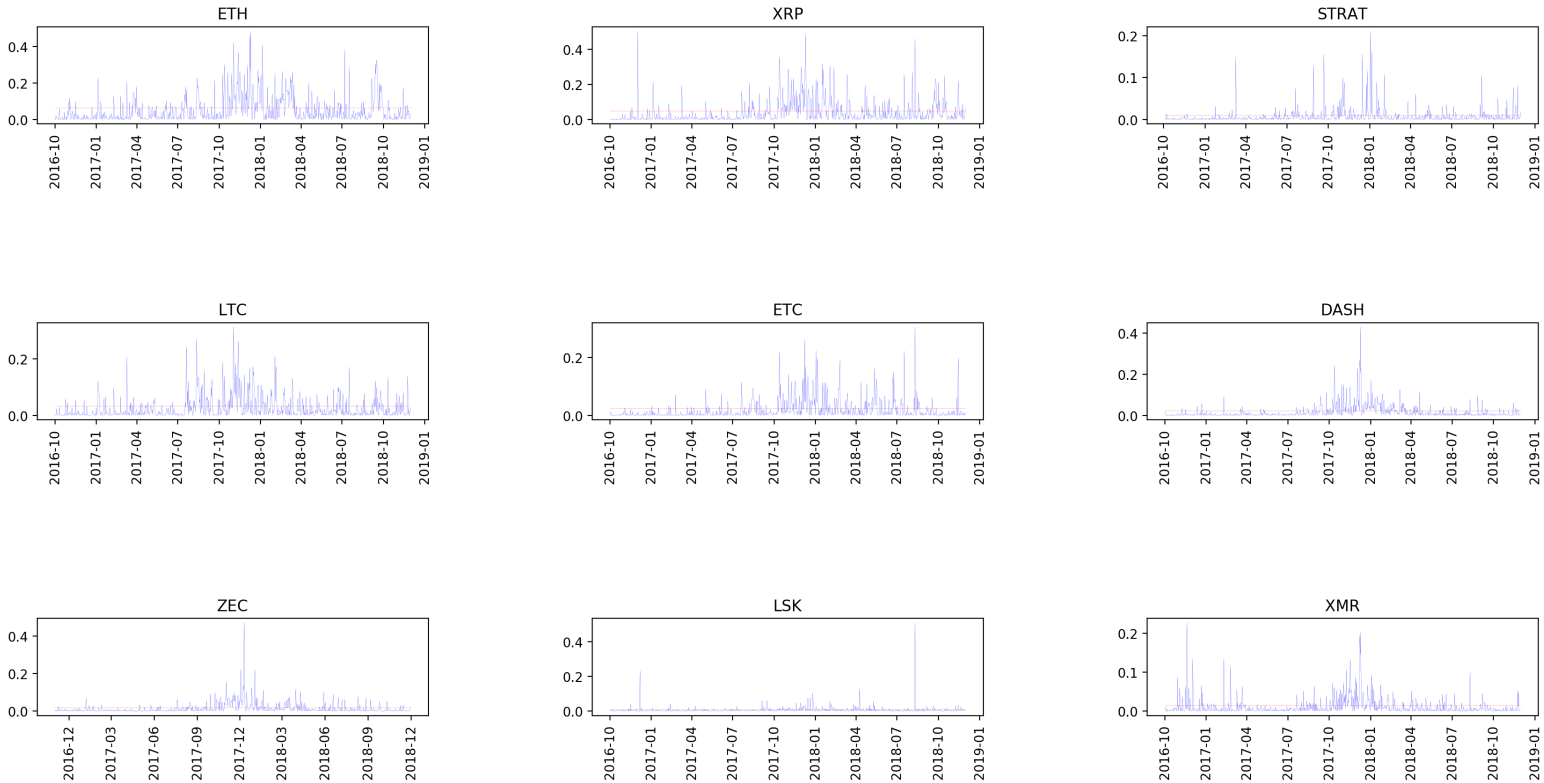


Figure 3.13: Systematic Risk Ratio (SRR) for Cryptocurrencies

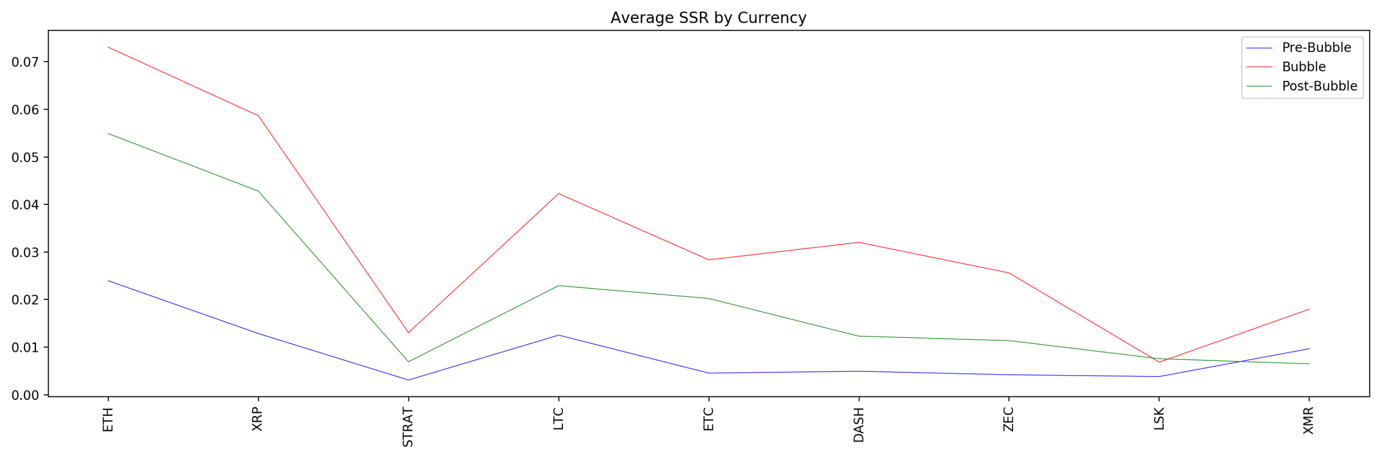


Figure 3.14: The Average SSR across the Cryptocurrencies

(December 2017) for all cryptos except LSK. This pattern matches the beta plots, which show more negative significant beta clustered during the Bitcoin bubble period. Figure 3.14 plots the average SSR across the cryptocurrencies. It is clear that, while the fraction of cryptocurrency variance explained by Bitcoin variance is greater during the bubble period and after the bubble burst, the explanatory power of the Bitcoin variance remains elevated compared to the pre-bubble period. Therefore, we assert that:

Fact 9: *The fraction of variance of cryptocurrency explained by the Bitcoin variance is high during the bubble period, and the explained fraction remains at an elevated level in the post-bubble period.*

3.7 Conclusions

In this study, we have presented a set of stylized facts on cryptocurrency returns and volatility. Specifically, from our analysis of high-frequency tick data on the most liquid nine cryptocurrencies from October 2016 to November 2018, we assert the following:

Fact 1: *Daily realized cryptocurrency volatility has high persistence.*
Fact 2: *The distribution of the logarithm of realized volatility of cryptocurrencies is close to normal.*
Fact 3: *The factor structure in daily cryptocurrency volatility is stronger than the factor structure in returns.*
Fact 4: *Economic and financial factors do not have strong explanatory power on the common factors of cryptocurrency return and volatility and there is a weak inverse relationship between cryptocurrency risk and macroeconomic indices.*
Fact 5: *Bitcoin can be considered for most cryptocurrencies as a fundamental factor able to explain a small proportion of the variations in return and volatility.*
Fact 6: *The Factor Structure model is more powerful in explaining variation in returns and volatilities during the Bitcoin bubble period and this explanatory power persists - and, for volatilities actually increases further - after the Bitcoin bubble burst.*
Fact 7: *There is heterogeneity in the relationship between Bitcoin and other cryptocurrencies for both returns and volatility after the Bitcoin pricing bubble burst.*
Fact 8: *Cryptocurrency betas with Bitcoin were negative before the Bitcoin bubble burst but became positive after the bubble burst.*

Fact 9: *The fraction of variance of cryptocurrency explained by the Bitcoin variance is high during the bubble period, and the explained fraction remains at an elevated level in the post-bubble period.*

Our study uncovers the properties of cryptocurrency and constructs a factor structure model. The cryptocurrencies are strongly explained by their own common factors but not by the fundamental economic factors used in most economics and finance studies. Taking into consideration Bitcoin as a fundamental factor, the nature of the relationship between Bitcoin and other cryptocurrencies shifted in terms of both return and volatility after the Bitcoin bubble burst. The strong common components of volatility across the major cryptocurrencies need to be considered as part of risk management when making investment decisions in cryptocurrency.

Chapter 4

General Conclusions

Information incorporation plays a central role in asset pricing for both the stock and fintech markets. On the one hand, my research seeks to develop an understanding of the psychological basis of economic decision-making by relaxing assumptions of rationality and adding behavioral variables that enrich the study of economics, without sacrificing the virtues of normative analysis. On the other hand, the rise in interest in cryptocurrencies and their potential to be widely adopted and used like fiat currencies has attracted the attention of academic research. However, the fundamental value and information of cryptocurrencies remains a mystery. I investigate a factor structure for the nine most liquid cryptocurrencies and argue that information contained in their own factor structure contributes to the explanation of the anomalous pricing behaviors in the cryptocurrency market.

First, in this thesis, I follow traditional behavioral economics and finance studies to relax the assumption of perfect rationality and argue that investors suffer from psychological biases, such as mood sensitivity, which ultimately lead to mispricing in the stock markets by causing decisions to acquire insufficient information. As investors become moody, they tend to acquire less earnings-related information before the announcements. This deficient information acquisition makes mood-sensitive stocks riskier than mood-insensitive stocks. The empirical results are consistent with theoretical predictions. Moody stocks earn higher expected excess returns than sober stocks. The higher expected returns generated by mood-sensitive stocks can be understood as risk premia required by the investors who hold these stocks in their portfolios. This study supports the view that investors' decision-making on information acquisition is subject to psychogenic irrationality in the financial markets. The extra risks added into stocks caused by this sub-optimal economic behavior should be compensated by the risk premium.

Second, I investigate whether the bias raised during the information acquisition process is induced using biased exogenous information and not necessarily solely from irrational agents. In essence, even perfectly rational investors with optimum behaviors are unable to compensate for the bias inherent in the information to which they are exposed, such as the tone of news. Introducing a new irrationality channel in the form of biased information transmission into a static information acquisition model, I am able to show bias in investors' perceptions of uncertainties concerning risky assets. This bias is generated exogenously, namely through the consumption of biased, publicly-available information from news media before an investment decision is made. In fact, this biased perception of uncertainties makes investors biased either with respect to the informativeness of price or to the value of the firm-specific information; thus, in this equilibrium with biased beliefs, investors' information acquisition deviates from the rational expectations. The model yields testable predictions that are verified by using a novel news dataset. The study shows that sentiment from news media, either concerning the stock market or particular firms, as a proxy for biased public information, has an inverse relationship with information acquisition when measured by information incorporated into the price before and after a firm's earnings announcement. In addition, information risk in the risky assets deviates from the rational expectations model as a result of the biased effect of news sentiment. The empirical results of predictability from firm-specific news sentiment on future cross-sectional stock returns support this theoretical proposition. In sum, these findings suggest that biased public information inherent in news sentiment serves to irrationalize investors' acquisition of firm-specific information through a biased perception of risky assets' uncertainties. Additionally, firm-specific news sentiment contributes to variations in information risk in risky assets and results in a variation of cross-sectional stock returns, which are compensation (risk premium) requested by investors to hold the risky assets.

Lastly, given the existing studies rare success in finding information from the traditional financial assets that can explain the pricing behaviors of cryptocurrency, I question whether a factor structure containing information serves as a "pricing model" to explore the variation of returns and volatilities in the cryptocurrency market. In view of this, I argue that the cryptocurrency market is distinct from traditional financial assets. I use high-frequency quote and tick data from the nine most liquid cryptocurrencies and Bitcoin to derive nine stylized facts from the analysis. The study finds that factor structure contributes to explain variations in cryptocurrency returns and volatility. My study also proposes that Bitcoin can be seen as a "market factor" that can explain variations in the returns and volatility of other cryptocurrencies. Additionally, I have made an original contribution to the literature with the first study to calculate the start and end dates of the Bitcoin bubble, and also investigated the shifting relationships between each of the nine cryptocurrencies and Bitcoin before and after the bubble burst.

Appendix A

Chapter 1 Appendix

A.1 Definitions of Financial Characteristics

Financial data comes from the CRSP/Compustat merged database. We merge stock data from CRSP and financial data by linking PERMNO (CRSP) and LPERMNO (CRSP/Compustat). If data from PERMNO or LPERMNO is missing or not matched correctly, we fetch data by checking tickers in two databases. The financial data from fiscal year-ends $t-1$ for stock i in month t returns from June of year y to May of year $y+1$. All variables are winsorized at 99.5 and 0.5%.

Market Capitalization = Total Market Value (MKTVALT) at the end of fiscal year. If market value data is not available, we take closing price × common shares outstanding at the end of the fiscal year.
Book value of equity = SEQ + TXDB + ITCB -BVPS
SEQ is the book value of shareholders' equity. TXDB is deferred taxes. ITCB is investment tax credit. BVPS is book value of preferred stock, taken from PSTKRV (redemption value), PSTKL (liquidating value) or PSTK (par value) depending availability in the database. If there is no available data for preferred stock, BVPS is set to zero. We delete data which is missing either SEQ or TXDB.
B/M = Book Value of Equity / Total Market Value
Dividend Yield is dividends per share on the end of fiscal year (DVPSX_F).
If data is missing, we fill the data in as zero. Dividend Paid is the probability of a company paying a dividend. We calculate the total dividend paid as equal to dividend paid per share × common share outstanding, and then set a dummy variable as equal to 1. Otherwise, total dividend is equal to 0 if there is no dividend paid.
Operating Cash Flow is the operating activity net cash flow (OANCF).
EPS is earnings per share (Basic), excluding extraordinary items (EPSPX).

ROA = Net Income/Loss (NI) / Total Asset (AT)
EBITDA/ASSET = Earnings Before Interest (EBITDA) / Total Asset (AT)
For realistic investment purposes, we delete data missing key profitability information such as net income, operating cash flow or EBITDA.
Book Leverage = Total Debt Including Current (DT) / Total Asset (AT)
PPE/ASSET = Property, Plant and Equipment (PPEGT) / Total Asset (AT)
R&D= (XRD)/Total Asset (AT)
We calculate the research and development Expense by scaling the total assets.
If there is data missing on total debt, R&D and PPE, we fill the data in as zero.
Revenue is total revenue (REVT).
Asset Growth = $TotalAsset_t$- $TotalAsset_t - 1/TotalAsset_{t-1}$
External Financing is the difference between the percentage change of asset from year $t-1$ to year t and the percentage change of retained earnings from $t-1$ to t. However, retained earnings can be 0 or negative, and we define the calculation regarding the change of retained earnings as follows: $Retained\ earnings_{t-1}$ (re_{t-1}) is not equal to 0: $re_t - re_{t-1}/re_{t-1}$. If re_{t-1} is equal to 0 and $re_t > 0$, we set the change of retained earnings as equal to 1. If re_{t-1} is equal to 0 and $re_t < 0$, we set the change of retained earnings as equal to -1. Otherwise, the change of retained earning equals 0.
Age is calculated as the first date from which the company's data is available in the database up to December 2016.
Idiosyncratic risk is measured by taking RSE of residuals from the Carhart pricing model as below:

$$R_{i,t} = \alpha_i + \beta_{MKT,i}MKT_t + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{CMA,i}CMA_t + \beta_{RMW,i}RMW_t + \beta_{MOM,i}MOM_t + \epsilon_{i,t}$$

A.2 Control Variable Definition

The earnings data used to test the mood biasing effect on information acquisition in section 1.2.2 is from Institutional Brokers Estimate System (I/B/E/S). The sample period is from 2008 to 2016, subject to Twitter data availability. We also conduct the test by extending the data to 2018, and the results are not changed a lot.
VIX: Daily closing value of VIX. Source: Wharton Research Data Services-CBOE Indexes.
EPU: Daily news-based Economic Policy Uncertainty Index. Source: BBM.
$Size$: Natural log of market value of equity. Source: Compustat.

RV: Return volatility is measured as standard deviation of daily return at each month. Source: CRSP.

$Institutional\ Ownership$ ($ITOW$): This is the institutional ownership percentage from Thomson Reuters Institutional (13f) Holdings data file.

$\overline{IVOL}$: Moving average stock idiosyncratic volatility is calculated based on the window between day $t-24$ and $t-4$. Sources: CRSP and Kenneth R. French Data Library.

$Price$: Average daily closing price from day $t-42$ to $t-21$ before a quarterly earnings announcement. Source: CRSP.

$NUMEST$: Number of analyst's earnings forecasts in the most recent month before a quarterly earnings announcement. Source: Institutional Brokers Estimate System (I/B/E/S).

$Turn$: Turnover is total number of shares traded over a period divided by total outstanding shares. Source: CRSP.

$\%Positive$: Counts of days with positive daily mood change divided the total days in the most recent month before the firm earnings announcements.

$\%Negative$: Counts of days with negative daily mood change divided the total days in the most recent month before the firm earnings announcements.

Appendix B

Chapter 2 Appendix

B.1 Theorems used to solve the model

Based on Bayes' rule of normal-normal updating (Back, 2010), X and Y are joint normally distributed. The expectation of X condition on Y can be projected:

$$\begin{aligned} E[X|Y] &= E[X] + \beta(Y - E[Y]) \\ \beta &= \frac{Cov(X,Y)}{Var(Y)} \\ Var(X|Y) &= Var(X) - \frac{[Cov(X,Y)]^2}{Var(Y)} \end{aligned} \tag{T.1}$$

Following Veldkamp (2011), the Wishart moment generating function of the exponential of a multi-variate quadratic form of a normal variable follows:

$$\begin{aligned} z &\sim N(0, \Sigma) \\ E[e^{z'Fz+G'z+H}] &= |I - 2\Sigma F|^{-1/2} exp[\frac{1}{2}G'(I - 2\Sigma F)^{-1}\Sigma G + H] \end{aligned} \tag{T.2}$$

B.2 Proof of Proposition 1

Investors who pay a cost c for acquiring firm-specific information, therefore, informed I investors' information set is:

$$\mathcal{F}_I = \{\overline{D}, M_1, e_1, \hat{p}_1\}$$

Based on T.1, informed investors' expected payoff and variance of the risky assets are:

$$\begin{aligned}
E^I_{b,1}[m_1] &= \frac{\sigma^2_{b,m}}{\sigma^2_{b,m}+\sigma^2_\eta}M_1 \\
Var^I_{b,1}[m_1] &= \frac{\sigma^2_{b,m}\sigma^2_\eta}{\sigma^2_{b,m}+\sigma^2_\eta} \\
E^I_{b,1}[D_2] &= \overline{D} + \frac{\sigma^2_{b,m}}{\sigma^2_{b,m}+\sigma^2_\eta}M_1 + e_1 \\
Var^I_{b,1}[D_2] &= Var^I_{b,1}[m_1] = \frac{\sigma^2_{b,m}\sigma^2_\eta}{\sigma^2_{b,m}+\sigma^2_\eta}
\end{aligned} \tag{B.21}$$

Uninformed investors don't observe the firm-specific information e_1, but they can partially learn about e_1 from the informative signal through price revealing $\hat{p}_1$. Therefore uninformed investors' U information set is :

$$\mathcal{F}_U = \{\overline{D}, M_1, \hat{p}_1\}$$

Based on T.1, uninformed investors learn about e_1 based on $\hat{p_1}$ is :

$$\begin{aligned}
E^U_{b,1}[e_1|\hat{p}_1] &= \frac{\sigma^2_{b,e}}{\sigma^2_{b,e}+\frac{K^2}{G^2}\sigma^2_x}\hat{p}_1 \\
Var^U_{b,1}[e_1|\hat{p}_1] &= \frac{K^2\sigma^2_{b,e}\sigma^2_x}{G^2\sigma^2_{b,e}+K^2\sigma^2_x}
\end{aligned} \tag{B.22}$$

Therefore, for uninformed investors, the expected payoff and variance of the risky asset are:

$$\begin{aligned}
E^U_{b,1}[D_2] &= \overline{D} + \frac{\sigma^2_{b,m}}{\sigma^2_{b,m}+\sigma^2_\eta}M_1 + \frac{\sigma^2_{b,e}}{\sigma^2_{b,e}+\frac{K^2}{G^2}\sigma^2_x}\hat{p}_1 \\
Var^U_{b,1}[D_2] &= Var^I_{b,1}[D_2] + Var^U_{b,1}[e_1] \\
&= \frac{\sigma^2_{b,m}\sigma^2_\eta}{\sigma^2_{b,m}+\sigma^2_\eta} + \frac{K^2\sigma^2_{b,e}\sigma^2_x}{G^2\sigma^2_{b,e}+K^2\sigma^2_x}
\end{aligned} \tag{B.23}$$

As defined in Grossman and Stiglitz (1980) and stated in Andrei et al. (2019), I defined price informativeness as:

$$
\begin{aligned}
Corr(\hat{p}_1, e_1) = \rho = \frac{Cov(\hat{p}_1, e_1)}{\sigma_{b,e}\sigma_{\hat{p}_1}} &= \frac{\sigma_{b,e}^2}{\sigma_{b,e}\sqrt{\sigma_{b,e}^2 + \frac{K^2\sigma_x^2}{G^2\sigma_{b,e}^2}}} \\
\rho^2 &= \frac{\sigma_{b,e}^2}{\sigma_{b,e}^2 + \frac{K^2\sigma_x^2}{G^2\sigma_{b,e}^2}} \\
\text{Define informativeness:}\ \ n = \frac{\rho^2}{1-\rho^2} &= \frac{\lambda_1\sigma_{b,e}^2}{\alpha Var_{b,1}^I[D_2]^2\sigma_x^2} \\
\text{Denote}\ \ \Phi = \frac{n}{1+n} &= \rho^2
\end{aligned}
\tag{B.24}
$$

Following Back (2010), the customary optimal portfolios for informed and uninformed investors with CARA utility are:

$$
\begin{aligned}
q_1^I &= \frac{E_{b,1}^I[D_2] - r_f P_1}{\alpha Var_{b,1}^I[D_2]} \\
q_1^U &= \frac{E_{b,1}^U[D_2] - r_f P_1}{\alpha Var_{b,1}^U[D_2]}
\end{aligned}
\tag{B.25}
$$

Therefore, B.21–25 yield equations (2.9) and (2.10) for informed and uniformed investors' optimal portfolios.

To find linear conjecture equilibrium price, the market clearing condition follows equation (2.6). Then, using terms A.22, A.23, and A.25 to replace terms in equation (2.6) yields:

$$
\begin{aligned}
\lambda_1\gamma\phi_I(E_{b,1}^I[D_2 - r_f P_1]) + (1-\lambda_1)\gamma\phi_U(E_{b,1}^U[D_2 - r_f P_1])) = x_1 \\
\gamma = \frac{1}{\alpha}, \quad \phi_I = \frac{1}{Var_{b,1}^I[D_2]}, \quad \phi_U = \frac{1}{Var_{b,1}^U[D_2]}
\end{aligned}
\tag{B.26}
$$

After taking tedious algebra, the unknown coefficients A, B, G, K, and H of the linear conjectured price P_1 in equation (2.7) can be easily solved and showed in equation (2.11) of Proposition 1.

B.3 Proof of Proposition 2

To find the fraction of investors who become informed about e_1 in equilibrium, I solve equation (2.12), the indifference condition proposed by Grossman and Stiglitz (1980) by applying T.2 as:

$$\begin{aligned}
F &= -\frac{1}{2}Var^{I}[D_2]^{-1} \\
G' &= -(E^{U}_{b,1}[D_2] - r_f P_1)Var^{I}_{b,1}[D_2]^{-1} \\
H &= -\frac{1}{2}(E^{U}_{b,1}[D_2] - r_f P_1)^2 Var^{I}_{b,1}[D_2]^{-1} \\
\Sigma &= Var^{U}_{b,1}[e_1|\hat{p}_1]
\end{aligned} \tag{B.27}$$

Applying B.27 yields:

$$E_b[U^I|P_1] = -|I - 2Var^{U}_{b,1}[e_1|\hat{p}_1](-\frac{1}{2})Var^{I}_{b,1}[D_2]^{-1}|^{-1/2}$$
$$e^{\frac{1}{2}E^{U}_{b,1}[D_2 - r_f P_1]^2 Var^{I}_{b,1}[D_2]^{-2}(I+Var^{U}_{b,1}[e_1|\hat{p}_1]Var^{I}_{b,1}[D_2]^{-1})^{-1}Var^{U}_{b,1}[e_1|\hat{p}_1] - \frac{1}{2}E^{U}_{b,1}[D_2 - r_f P_1]^2 Var^{I}_{b,1}[D_2]^{-1}} \tag{B.28}$$

Solving B.28 yields:

$$\begin{aligned}
E_b[U^I|P_1] &= -\Big(\frac{Var^{I}_{b,1}[D_2]}{Var^{U}_{b,1}[e_1|\hat{p}_1] + Var^{I}_{b,1}[D_2]}\Big)^{1/2} e^{\frac{1}{2}E^{U}_{b,1}[D_2 - r_f P_1]^2 Var^{I}_{b,1}[D_2]^{-1}\left[\frac{-Var^{I}_{b,1}[D_2]}{Var^{U}_{b,1}[e_1|\hat{p}_1]+Var^{I}_{b,1}[D_2]}\right]} \\
&= -\Big(\frac{Var^{I}_{b,1}[D_2]}{Var^{U}_{b,1}[e_1|\hat{p}_1] + Var^{I}_{b,1}[D_2]}\Big)^{1/2} e^{-\frac{1}{2}\frac{E^{U}[D_2 - r_f P_1]^2}{Var^{U}_{b,1}[D_2]}} \\
E_b[U^U|P_1] &= -e^{-\frac{1}{2}\frac{E^{U}[D_2 - r_f P_1]^2}{Var^{U}_{b,1}[D_2]}} \\
\frac{E_b[U^I]}{E_b[U^U]} &= e^{\alpha c}\sqrt{\frac{Var^{I}_{b,1}[D_2]}{Var^{U}_{b,1}[e_1|\hat{p}_1] + Var^{I}_{b,1}[D_2]}} = e^{\alpha c}\sqrt{\frac{Var^{I}_{b,1}[D_2]}{Var^{U}_{b,1}[D_2]}}
\end{aligned} \tag{B.29}$$

Therefore, applying B.22-24, it is straightforward to find the benefit and cost function $\Pi(*)$.

B.4 Proof of Corollary 1

To solve equilibrium λ_1 as a function of uncertainties ($Var^{I}_{b,1}[D_2]$ and $\sigma^2_{b,e}$), I set the cost and benefit function $\Pi(*) = 0$. Hence, I directly solve the numerator of $\Pi(*)$ equals to 0 as :

$$F(*) = \lambda_1^2\sigma^2_{b,e}\delta + \alpha^2 Var^{I}_{b,1}[D_2]^2\sigma^2_x\delta - \alpha^2 Var^{I}_{b,1}[D_2]\sigma^2_x\sigma^2_{b,e} = 0, \quad \text{where} \quad \delta = e^{2\alpha c} - 1 \tag{B.30}$$

By applying the implicit theorem in a region $F'(\lambda_1) \geq 0$, the $\frac{\partial \lambda_1}{\partial Var^I_{b,1}[D_2]}$ can be found as :

$$\begin{aligned}
\frac{\partial \lambda_1}{\partial Var^I_{b,1}[D_2]} &= -\frac{\partial F}{\partial Var^I_{b,1}[D_2]} \times \frac{\partial \lambda_1}{\partial F} \\
&= \frac{\alpha^2 \sigma_x^2 \sigma_{b,e}^2 - 2\alpha^2 \sigma_x^2 \delta Var^I_{b,1}[D_2]}{2\lambda_1 \sigma_{b,e}^2 \delta}
\end{aligned} \tag{B.31}$$

As long as $Var^I_{b,1}[D_2] \leq \frac{\sigma^2_{b,e}}{2\delta}$, which is the threshold of $Var^I_{b,1}[D_2]$ then $F(\lambda_1)$ increases as the λ_1 increases to reach the theoretical maximum fraction of informed investors. In that, $\frac{\partial \lambda_1}{\partial Var^I_{b,1}[D_2]} > 0$. On the one hand, as $\frac{\partial Var^I_{b,1}[D_2]}{\partial \sigma_{b,m}} > 0$ is known, by applying chain rule, it is easy to show that $\frac{\partial \lambda_1}{\sigma_{b,m}} > 0$. On the other hand, the bias function $\beta(S_m, \sigma_m^2)$ is inversely related to the biased perception of $\sigma^2_{b,m}$ as showed in equation (2.2). In other words, $\frac{\partial \sigma_{b,m}}{\partial S_m} < 0$ is monotonic decreasing. Noted that, without loss of generality , the bias function $\beta(*)$ is not assumed for particular function forms. By applying the chain rule, as a result, $\frac{\partial \lambda_1}{\partial S_m} < 0$.

In addition, the $\frac{\partial \lambda_1}{\partial \sigma^2_{b,e}}$ can be solved in the same steps:

$$\begin{aligned}
\frac{\partial \lambda_1}{\partial \sigma_{b,e}} &= -\frac{\partial F}{\partial \sigma_{b,e}} \times \frac{\partial \lambda_1}{\partial F} \\
&= \frac{\alpha^2 Var^I_{b,1}[D_2]\sigma_x^2 - \lambda_1^2 \delta}{2\lambda_1 \sigma_{b,e}^2 \delta} \\
\max(\lambda_1^2) &= \frac{\alpha^2 Var^I_{b,1}[D_2]\sigma_x^2}{2\delta} \text{ when } Var^I_{b,1}[D_2] = \frac{\sigma^2_{b,e}}{2\delta} \\
\lambda_1^2 &\leq \frac{\alpha^2 Var^I_{b,1}[D_2]\sigma_x^2}{2\delta} \\
\frac{\partial \lambda_1}{\partial \sigma_{b,e}} &= \frac{\alpha^2 Var^I_{b,1}[D_2]\sigma_x^2 - \frac{1}{2}\alpha^2 Var^I_{b,1}[D_2]\sigma_x^2}{2\lambda_1 \sigma_{b,e}^2 \delta} \\
&\text{this yields } \frac{\partial \lambda_1}{\partial \sigma_{b,e}} > 0 \text{ strictly.}
\end{aligned} \tag{B.32}$$

As the bias function $\beta(S_e, \sigma_e^2)$ in equation (2.2) indicates a monotonic decreasing relationship between biased perception of firm-specific uncertainty and firm-specific news sentiment as the proxy of biased pubic information received by investors, therefore, $\frac{\partial \sigma_{b,e}}{\partial S_e} < 0$ is implied by equation (2.2), by applying the chain rule with A.32, it is straightforward to show that $\frac{\partial \lambda_1}{\partial S_e} < 0$.

B.5 Proof of Proposition 3

Following O'Hara (2003), I assume the risky asset random supply $x_1 \sim N(\overline{x}, \sigma_x^2)$. Therefore, this non-zero expected random supply $\overline{x}$ implies a risky premium. Based on the market clearing condition, the expected return of the risky asset is :

$$\begin{aligned}
&\lambda\gamma\phi_I E^I[D_2] + (1-\lambda_1)\gamma\phi_U E^U[D_2] - x_1 = (\lambda\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)P_1 r_f \\
&\text{Expected Return:} \\
&E[R_2] = \frac{\lambda_1\gamma\phi_I E^I[D_2] + (1-\lambda_1)\gamma\phi_U E^U[D_2]}{r_f(\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)} - P_1 \\
&= \frac{E[x_1]}{r_f(\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)} \\
&= \frac{\alpha\overline{x}}{r_f(\lambda_1\phi_I + (1-\lambda_1)\phi_U)}
\end{aligned} \tag{B.33}$$

First, the expected return is a function of λ_1 the fraction of investors who are informed about e_1. The $\frac{\partial E[R_2]}{\partial \lambda_1}$ can be found as :

$$\frac{\partial E[R_2]}{\partial \lambda_1} = \frac{-\alpha\overline{x}[r_f\phi_I - r_f\phi_U + (1-\lambda_1)\frac{\partial \phi_U}{\partial \lambda_1}]}{(r_f\lambda_1\phi_I + r_f(1-\lambda_1)\phi_U)^2} \tag{B.34}$$

Clearly, $Var_{b,1}^U[D_2]$ the uninformed investors variance of the risky asset's payoff decreases as λ_1 increases because the price informativeness n increases. Therefore, it is easy to show that $\frac{\partial \phi_U}{\partial \lambda_1} > 0$, where $\phi_U = \frac{1}{Var_{b,1}^U[D_2]}$, $\lambda_1 \leq 1$ and $\phi_I \geq \phi_U$. As a result , $\frac{\partial E[R_2]}{\partial \lambda_1} < 0$ in B.34. Because $\frac{\partial \lambda_1}{\partial S_e} < 0$ argued in Appendix B.4, applying the chain rule yields $\frac{\partial E[R_2]}{\partial S_e} > 0$.

B.6 Proof of Corollary 2 and 3

The equilibrium fraction of informed investors in rational expectations is λ_1^e and expected return $E[R_2^e]$ reconciles to O'Hara (2003) study. As the model in this study indicates, firm-specific news sentiment S_e deviates λ_1^e to a biased belief equilibrium $\lambda_1^{b,e}$. On the one hand, as S_e increases and $\frac{\partial \lambda_1}{\partial S_e} < 0$, therefore $\lambda_1^{b,e} < \lambda_1^e$. As $\frac{\partial E[R_2]}{\partial \lambda_1} < 0$ proved in B.34, $\lambda_1^{b,e} < \lambda_1^e \rightarrow E_b[R_2] > E_e[R_2]$ which completes Corollary 2 proof. The Corollary 3 proof can be easily completed by the other way around.

B.7 A Toy Model of News Bias

The toy model of bias in media is motivated by Dyck and Zingales (2003). The suppliers of information such as journalists or news companies supply news as a function of bias:

$$N^s = \theta^s + \beta b \tag{B.35}$$

The bias can be either an optimistic or pessimistic tone used by the information suppliers to improve readership.

However, investors demand high accuracy in news, in that, too much bias deceases demand or readership of investors. Their demand function is negatively related to the bias imposed by the information suppliers:

$$N^d = \theta^d - \gamma b \tag{B.36}$$

As mentioned byDyck and Zingales (2003), I assume journalists or news suppliers choose to implement bias into their news, in a competitive market to equate demand and supply of news. Therefore, in equilibrium, the bias is:

$$b = \frac{\theta^d - \theta^s}{\beta + \gamma} \tag{B.37}$$

The B.37 indicates information from news is always subject to some bias. Dyck and Zingales (2003)) argue that a lower degree of bias induces excess demand for news and a higher degree of bias induces excess supply of news. Noted this, the B.35–B.37 only show the existence of bias in the news provided by the information suppliers. In other words, there is no need to specify the sign of bias.

In sum, this toy model assuredly motivates the idea that information from news contains bias, for which I argue in this study. More specifically, the bias is subject to news suppliers' choice of using either an optimistic (positive) or pessimistic (negative) tone (sentiment) in the news to potentially increase readership or fulfil readers' demand for news.

B.8 Variable Definition

Buzz: This measure is the sum of all references from the news about either the stock market or particular firms that are included in one of the TRMI indexes over 24 hours.

$Sentiment$: Overall positive references net of negative references in news about either the stock market or particular firms over 24 hours.
$EmotionVsFact$: The sum of the absolute value of all emotions and opinions (both positive, negative, surprise and uncertainty) minus the sum of the absolute value of all facts (topics and other subjects/themes/nouns) divided by the sum of all references in the news.
VIX: Daily closing value of VIX. Source: Wharton Research Data Services-CBOE Indexes.
EPU: Daily news-based Economic Policy Uncertainty Index. Source: BBM.
$SP500$ Realized volatility is downloaded from Risk Lab by Da and Xiu (2019).
ME: Market value of equity in fiscal year closing price times total share of equity. Source: Compustat.
$Size$: Natural log of market value of equity. Source: Compustat.
BM: Book to Market Ratio as defined in Fama and French (1992). Source: Compustat.
$Illiquity$: Monthly illiquidity measure as per Amihud (2002). Source: CRSP.
OP: Operating Profitability, as defined in Fama and French (2015). Source: Compustat.
INV: Investment measure is defined as in Fama and French (2015) study. Source: Compustat.
RV: Return volatility is measured as standard deviation of daily return at each month. Source: CRSP.
MOM: Momentum Return Measure is defined as the cumulative return from $t-11$ to the month $t-1$ before the last month t. Source: CRSP.
ST: Return from the last month to capture short-term reversal effect. Source: CRSP.
$AbRet$: Daily holding period return minus the value-weighted market return. Source: CRSP.
$AbRet_{t-5,t-1}$: Five days cumulative abnormal return from $t-5$ to $t-1$. Source: CRSP.
$AbTurn$: Natural log turnover at day t net of the average turnover in the last five days. Source: CRSP.
SUE: Unexpected earnings is calculated based on Compustat data. The calculation follows Livnat and Mendenhall (2006). Source: Compustat.
SUE^{IBES}: Unexpected earnings is calculated based on I/B/E/S data. The calculation follows Livnat and Mendenhall (2006) study. Source: Institutional Brokers Estimate System (I/B/E/S).

$ForecastDispersion$: The standard deviation of analysts' earnings forecasts in the most recent month before quarterly earnings announcement and scaled by the stock price. Source: Institutional Brokers Estimate System (I/B/E/S).
$ForecastRevision$: The median analysts' 3-month earnings forecast revision is based on Chan et al. (1996). Source: Institutional Brokers Estimate System (I/B/E/S).
$Idiosyncratic\ Volatility$ ($IDIOVOL$): The residual standard error from Fama and French (2015) five factor plus momentum factor pricing model on a daily rolling basis. I require each company to have at least 60 observations to run the time-series regression. Sources: CRSP and Kenneth R. French Data Library.
$Abs(FFCAR)$: Absolute value of cumulative abnormal return (CAR) is calculated from Fama and French (2015) five factors plus momentum factor pricing model. The factor betas used to calculate CAR are estimated 90 days before a quarterly earnings announcement. Sources: CRSP and Kenneth R. French Data Library.
$Institutional\ Ownership$ ($ITOW$): This is the institutional ownership percentage from Thomson Reuters Institutional (13f) Holdings data file.
$\overline{IVOL}$: Moving average stock idiosyncratic volatility is calculated based on the window between day $t-24$ and $t-4$. Sources: CRSP and Kenneth R. French Data Library.
$Price$: Average daily closing price from day $t-42$ to $t-21$ before a quarterly earnings announcement. Source: CRSP.
$NUMEST$: Number of analyst's earnings forecasts in the most recent month before a quarterly earnings announcement. Source: Institutional Brokers Estimate System (I/B/E/S).
$Turn$: Turnover is total number of shares traded over a period divided by total outstanding shares. Source: CRSP.

B.9 Robustness Tests

B.9.1 Market News Sentiment and Market Uncertainty Regression Test

Table 2.1 shows the negative Pearson correlation coefficients between stock market news sentiment and market uncertainty measures. In this section, I conduct a fixed effect regression to further verify the assumption in equation (2.2) that an increase of market news sentiment biases investors to perceive a lower market uncertainty. Specifically, I use three measures of market uncertainty: S&P500 realized volatility (RV_{500}), EPU and VIX. For each firm earnings announcement day, I calculate the average monthly market uncertainty from the announcement day (t) to the next 21 trading days ($t+21$ for RV and VIX) or 31 calendar days ($t+31$ for EPU). The fixed effect regression is as follows:

$$Dep_{i,[t,t+21/31]} = \beta_0 + \beta_1 Sentiment_{m,[t-21,t-1]} + X\delta + \epsilon_{i,t} \tag{B.38}$$

where $Sentiment_{m,[t-21,t-1]}$ is the $Buzz$-weighted average stock market news sentiment from 21 trading days up to 1 day before the earnings announcement. X is a vector of control variables including size ($Size$), turnover ($Turn$), average price ($Price$), return volatility (RV) and institutional ownership ($ITOW$) (see detailed definitions in Appendix B.8) and δ is the coefficient vector. Since volatility is strongly persistent, I also control the lag variable, which is the average value one month before the earnings announcement for each market uncertainty measure.

Table B.1 displays the results from equation (B.38). I control month-, year- and firm-fixed effects and the standard errors are clustered by firm- and year- fixed effects. Columns (1)-(3) shows that stock market news sentiment negatively predict market uncertainty across all three measures of economic uncertainty. In sum, the results are consistent with the negative correlation shown in Table 2.1. More importantly, this test confirms the assumption in equation (2.2) to serve the biasing channel in the model that investors' perception of market uncertainty is irrationalized by reading news characterized by a non-neutral tone.

B.9.2 Alternative Measure of Firm-Specific Information Acquisition

Specifically, I calculate the average total count of search volume for the files in the most recent month before the announcement. I then take the natural logarithm of the average of total SEC files searching volumes ($LogSEC_{i,t}$). To some extent, the count of SEC EDGAR file searching volume is a more straightforward way to understand investors' acquisition of firm-specific information. The fixed effect reggression is as follows:

$$LogSEC_{i,t} = \beta_0 + \beta_{j,1} Sentiment_{j,t-21,t-1} + X\delta + \epsilon_{i,t}, \text{ where } j \in \{m, i\} \tag{B.39}$$

where X is the vector of control variables which are same as the test in Table 2.3 and δ is the vector of coefficient. Not surprisingly, the results are consistent with those of Table 2.3 (with price the jump measure). As news tones tend to be more positive, investors are less willing to download the company's SEC files, showing a decrease in firm-information acquisition.

TABLE B.1: Market News Sentiment and Market Uncertainty

This table presents the results of regressions of market uncertainty measures on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (B.38): $Dep_{i,[t,t+21/31]} = \beta_0 + \beta_1 Sentiment_{m,[t-21,t-1]} + X\delta + \epsilon_{i,t}$ and $Dep_{i,[t,t+21/31]}$ is the average monthly market uncertainty from the announcement day (t) to the next 21 trading days (t+21 for RV and VIX) or 31 calendar days (t+31 for EPU). The news sentiment variable $Sentiment_{m,[t-21,t-1]}$ and $Buzz_{m,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{m,t}$-weighted average in the study window. Control variables include: $Buzz_{m,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, lagged dependent variables are calculated as 21 (31) days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.8. Standard errors are clustered by both firm- and time- fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	(1)	(2)	(3)
Dependent Variable	$RV_{500,[t,t+21]}$	$EPU_{t,t+31}$	$VIX_{t,t+21}$
$Sentiment_{m,[t-21,t-1]}$	-0.166^{***}	-121.736^{***}	-15.111^{***}
	(0.002)	(1.346)	(0.243)
$RV_{500,[t-21,t-1]}$	0.254^{***}		
	(0.005)		
$EPU_{t-21,t-1}$	0.0003^{***}	0.522^{***}	0.040^{***}
	(0.000)	(0.003)	(0.001)
$Buzz_{m,[t-21,t-1]}$	0.381^{***}	-58.285^{***}	38.643^{***}
	(0.005)	(2.115)	(0.479)
$VIX_{t-21,t-1}$		0.174^{***}	0.326^{***}
		(0.021)	(0.005)
Controls	Yes	Yes	Yes
FE Firms	Yes	Yes	Yes
FE Month	Yes	Yes	Yes
FE Year	Yes	Yes	Yes
Observations	91,873	91,873	91,873
R-squared	0.620	0.728	0.720
Number of Firms	10,241	10,241	10,241

Cluster standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

B.9.3 Fama-Macbeth Regression Excluding Earnings Announcement Days

I re-conduct analysis to confirm the robustness of the impact of the firm-specific news sentiment on cross-sectional stock returns, by excluding earnings announcement days and sorting regression data by different financial characteristics, which may potentially affect the predictability of the deviation of information risk resulted by news sentiment.

Table B.3 shows the results from running daily cross-sectional Fama–Macbeth (1973) regressions (2.28) with daily cross-sectional data, excluding all earnings announcement days. Because, as argued by Tetlock et al. (2008), firm-specific news is most likely to be clustered near the time of a company's earnings announcement, there is a concern is raised that the inclusion of these days may amplify the impact from firm-specific news sentiment and other effects from news related to company earnings.[1]

[1]For example, Tetlock et al. (2008) did find that earnings-related news has incremental benefit to uncover firms' value-relevant information. Thus, Tetlock (2010) thoroughly considers that information

Table B.2: News Sentiment Impact on Information Acquisition Measured by Counts of SEC Files Clicks

This table presents the results of regressions of the count of SEC EDGAR file searching volume as the proxy for firm-specific information acquisition on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (B.39): $LogSEC_{i,t} = \beta_0 + \beta_{j,1} Sentiment_{j,t-21,t-1} + X\delta + \epsilon_{i,t}$, where $j \in \{m, i\}$ and $LogSEC_{i,t}$ is the average of total SEC files searching volumes in the most recent month before the earnings announcement. The news sentiment variable $Sentiment_{j,[t-21,t-1]}$ and $Buzz_{j,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{j,t}$-weighted average in the study window. Control variables include: $Buzz_{j,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, economic uncertainty proxies (VIX and EPU) and the numbers of analyst coverage is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.8. Standard errors are clustered by both firm- and time- fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	Panel A Stock Market News Sentiment			*Panel B Firm-Specific News Sentiment*		
	(1)	(2)	(3)	(1)	(2)	(3)
Dependent Variable	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$
$Sentiment_{i,t-21,t-1}$				-0.052^{***}	-0.042^{**}	-0.042^{**}
				(0.017)	(0.017)	(0.017)
$Sentiment_{m,t-21,t-1}$	-0.231^{***}	-0.247^{***}	-0.227^{**}		-0.42**	-0.394**
	(0.085)	(0.091)	(0.09)		(0.187)	(0.189)
$Buzz_{m,t-21,t-1}$		-0.016	-0.034			
		(0.025)	(0.028)			
$Buzz_{i,t-21,t-1}$					0.023***	0.022***
					(0.004)	(0.004)
$VIX_{t-21,t-1}$		0.004*			0.001	
		(0.002)			(0.004)	
$EPU_{t-21,t-1}$			0.0004			-0.0004
			(0.000)			(0.0006)
LagDep	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Day of Week	Yes	Yes	Yes	Yes	Yes	Yes
FE Year-Month	Yes	Yes	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,412	39,971	39,971	9,183	9,121	9,121
R-squared		0.845	0.845		0.861	0.861
Number of Firms	3,660	3,641	3,641	2,586	2,568	2,568

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Columns (1)–(3) in Table B.3 show that excluding news of earnings announcements slightly reduces all of the coefficients, showing the small incremental benefit from news on earnings announcement days. Nonetheless, all results remain both statistically and economically significant. Indeed, news released while an earnings announcement is being made is more likely to attract investors' attention. In addition, information from news or online media reported close to a firm's earnings announcement plays an important role in transmitting firm-fundamental information to investors and traders (Tetlock et al., 2008; Chen et al., 2014). However, as Table B.3 demonstrates, excluding news reported near the time of earnings announcement does not comprise the significance of news sentiment's positive predictability on cross-sectional stock returns. In an unreported table, I also test by excluding firm-specific news on the earning announcement day, on the day that precedes it and on the day that follows it; the results

asymmetry dissolution from public news may be led by earnings news. To accommodate for this, Tetlock excludes earnings-related news in the main regression analysis as a robustness concern.

Table B.3: Cross-Sectional Return Predictability from Firm-Specific News Sentiment without Earnings Announcement Days

This table presents results excluding data on firm earnings announcement days and results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$. Variables measured by news content and all other control variables are known by day t. Columns (1)–(3) report the time-series average of the coefficients based on the model in equation (2.28): $DepVar_{i,t+1} = \beta_0 + \beta_1 Sentiment_{i,t} + \delta X + \epsilon_{i,t}$ for each trading day t, where $DepVar_{i,t+1}$ is R^e_{t+1}, $R^e_{t+2,t+5}$, and $R^e_{t+2,t+10}$, respectively. The variable $Sentiment_{i,t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i,t} * Sentiment_{i,t}$, $EmotionVsFact_{i,t} * AbRet_{i,t}$, and $Buzz_{i,t} * AbRet_{i,t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals from $Sentiment_{i,t}$. Additionally, abnormal return $AbRet_{i,t}$ at day t and its related interactions such as $AbRet_{i,t} * Size_{i,t}$ and $AbRet_{i,t} * AbTurn_{i,t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the Last Five Days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.8. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

	(1)	(2)	(3)
Dependent Variable	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	2.146	3.086	3.605
	(8.074)	(5.969)	(4.383)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-1.078	-0.878	-0.968
	(-4.805)	(-2.093)	(-1.562)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-1.311	0.852	-0.172
	(-2.649)	(1.139)	(-0.150)
$Buzz_{i,t} * AbRet_{i,t}$	3.537	4.065	6.245
	(6.753)	(5.337)	(5.377)
$Buzz_{i,t} * ME_{i,t}$	-0.112	0.771	0.597
	(-0.377)	(1.154)	(0.507)
$Buzz_{i,t}$	-0.133	-0.288	1.034
	(-0.457)	(-0.458)	(0.985)
$EmotionVsFact_{i,t}$	-0.207	-0.780	-0.304
	(-0.781)	(-1.529)	(-0.382)
$AbRet_{i,t}$	-4.396	-5.768	-7.039
	(-6.555)	(-4.876)	(-4.462)
$ME_{i,t}$	-1.660	-5.741	-13.150
	(-3.265)	(-3.742)	(-4.397)
$BM_{i,t}$	-0.443	-2.092	-2.244
	(-0.840)	(-1.360)	(-0.742)
$OP_{i,t}$	0.079	0.433	0.929
	(0.210)	(0.429)	(0.500)
$IVN_{i,t}$	0.013	-2.317	-4.918
	(0.039)	(-2.264)	(-2.467)
$RV_{i,t}$	-0.157	-0.220	-0.966
	(-0.206)	(-0.086)	(-0.186)
$MOM_{i,t}$	-0.313	1.365	3.566
	(-0.504)	(0.718)	(0.937)
$ST_{i,t}$	-0.833	-1.792	-3.663
	(-1.478)	(-1.118)	(-1.206)
$AbRet_{i,t} * Size_{i,t}$	-2.694	-6.420	-8.607
	(-5.458)	(-7.562)	(-6.886)
$AbTurn_{i,t}$	-5.432	-1.171	-4.822
	(-4.105)	(-0.517)	(-1.300)
$AbRet_{i,t-5,t-1}$	-2.827	-4.091	-4.665
	(-4.793)	(-3.047)	(-2.088)
$AbRet_{i,t} * AbTurn_{i,t}$	0.326	-0.728	-1.433
	(1.080)	(-1.374)	(-1.949)
$Constant$	3.399	16.207	36.189
	(1.891)	(2.406)	(2.616)
Daily Average Firms	512	511	511
Adjusted R-squared	0.147	0.138	0.135
Observations	2,538,963	2,537,599	2,536,117

are similar to Table B.3.

B.9.4 Sub-sample Fama-Macbeth Regression Analysis

I divide data into samples based on characteristics of firm size, illiquidity, analyst coverage, analyst forecast dispersion and institutional ownership. For each day, I divide stocks into two sub-samples, high and low, based on the daily cross-sectional median of each characteristic. Each sub-sample must have at least 50 firms to run the Fama–Macbeth (1973) regression model.

Panels A through E in Table B.4 are regression results based on the sub-samples for size, analyst coverage, analyst forecast dispersion, illiquidity and institutional ownership respectively. The high and low size sub-sample regression shows similar results to Table 2.5. Unsurprisingly, news sentiment predictability in the small firm sub-sample has a relatively stronger effect than the big firm sub-sample. In the small size sub-sample, news sentiment is statistically significant in its prediction of all future returns for R^e_{t+1} and cumulative returns $R^e_{t+2,t+5/10}$. However, it is well addressed empirically in the existing literature that large firms generally make more information available to investors and have less information asymmetry than small firms (Banz, 1981; Barry and Brown, 1984; Atiase, 1985; Freeman, 1987), thus showing a relatively weak effect on cumulative returns $R^e_{t+2,t+5/10}$. By the same token, Panel D shows very similar results as Panel A, because small and illiquid stocks are commonly known to share similar issues, especially in respect of information asymmetry. However, results in Panel B and C (for analyst coverage and analyst forecast dispersion, respectively) do not change much compared to the results from the full sample shown in Table 2.5. Both the number of analysts following a company and how analysts hold different beliefs about companies' earnings performance are unable to explain the cross-sectional variation of stock returns raised by variation of information asymmetry risk implied by news sentiment.

Finally, it is intriguing that variation in institutional ownership does not explain the positive predictive effect of news sentiment on cross-sectional stock returns at $t+1$, however, the significance of prediction on cumulative returns up to $t+5$ and $t+10$ are reduced in both the high and low institutional ownership sub-sample regressions. The reduction in significance is more pronounced in the low institutional ownership sub-sample. A potential reason is that institutional investors are relatively easier or cost-efficient to be informed (Hendershott et al., 2015). In other words, when it comes to making an investment into an asset, institutional investors are more likely to be biased by news sentiment in their perception of uncertainties in the risky asset. Hence,

Table B.4: Cross-Sectional Return Predictability from Firm-Specific News Sentiment Sorted by Firm Characteristics

This table presents results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$ with different sub-samples sorted into two portfolios based on financial characteristics. For each day, I divide stocks into two sub-samples: high and low, based on the daily cross-sectional median of each characteristic. From panel A to E, samples are sorted based on firm size, analyst coverage, analyst forecast dispersion, illiquidity measure and institutional ownership. The low and high sub-panels report the time-series average of the coefficients from each characteristic sorted sub-samples and is based on the model in equation (2.28): $DepVar_{i,t+1} = \beta_0 + \beta_1 Sentiment_{i,t} + \delta X + \epsilon_{i,t}$ for each trading day t, where $DepVar_{i,t+1}$ is R^e_{t+1}, $R^e_{t+2,t+5}$, and $R^e_{t+2,t+10}$, respectively. The variable $Sentiment_{i,t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i,t} * Sentiment_{i,t}$, $EmotionVsFact_{i,t} * AbRet_{i,t}$, and $Buzz_{i,t} * AbRet_{i,t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals from $Sentiment_{i,t}$. Additionally, abnormal return $AbRet_{i,t}$ at day t and its related interactions such as $AbRet_{i,t} * Size_{i,t}$ and $AbRet_{i,t} * AbTurn_{i,t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the Last Five Days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.8. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Size Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression

	Low			High		
	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	2.702	3.386	5.098	1.222	1.142	0.060
	(6.638)	(4.383)	(4.303)	(3.963)	(1.911)	(0.063)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-1.193	-0.549	-0.994	-0.838	-0.106	0.337
	(-3.408)	(-0.762)	(-1.003)	(-3.066)	(-0.219)	(0.441)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-2.139	1.650	-2.487	-0.971	0.422	1.756
	(-2.841)	(1.300)	(-1.306)	(-1.883)	(0.460)	(1.303)
$Buzz_{i,t} * AbRet_{i,t}$	3.539	4.108	6.853	2.374	3.161	4.010
	(4.486)	(3.270)	(3.679)	(4.786)	(3.669)	(3.069)
$BUZZ_{i,t}$	-0.163	-0.421	1.289	-0.111	-1.004	-1.108
	(-0.377)	(-0.520)	(1.016)	(-0.303)	(-1.152)	(-0.724)
$Buzz_{i,t} * Size_{i,t}$	-0.147	0.268	0.204	0.069	1.087	1.887
	(-0.335)	(0.287)	(0.136)	(0.250)	(1.496)	(1.435)
$AbRet_{i,t}$	-4.769	-2.323	-2.186	-4.760	-8.288	-11.453
	(-5.456)	(-1.558)	(-1.018)	(-7.186)	(-7.197)	(-7.179)
$ABbTurn_{i,t}$	-6.553	0.484	-2.101	-3.397	2.310	1.470
	(-3.128)	(0.147)	(-0.408)	(-2.902)	(0.963)	(0.333)
$AbRet_{i,t-5,t-1}$	-1.759	-2.554	-1.352	-4.187	-6.747	-10.216
	(-2.389)	(-1.643)	(-0.547)	(-6.638)	(-4.727)	(-4.458)
$AbRet_{i,t} * AbtTurn_{i,t}$	0.666	-1.962	-2.907	0.409	2.197	1.716
	(1.173)	(-2.182)	(-2.263)	(0.853)	(2.530)	(1.551)
$AbRet_{i,t} * Size_{i,t}$	-2.653	-4.104	-5.024	-0.769	-2.246	-2.810
	(-3.455)	(-3.217)	(-2.939)	(-1.491)	(-2.381)	(-2.059)
$ME_{i,t}$	-1.430	-4.669	-7.941	-0.709	-2.343	-5.472
	(-2.563)	(-3.095)	(-2.838)	(-1.739)	(-1.867)	(-2.264)
$BM_{i,t}$	-1.041	-3.457	-4.322	-0.706	0.188	0.869
	(-1.637)	(-1.953)	(-1.277)	(-1.222)	(0.113)	(0.275)
$OP_{i,t}$	-0.331	0.420	0.918	-0.176	0.988	1.919
	(-0.595)	(0.292)	(0.356)	(-0.383)	(0.829)	(0.856)
$IVN_{i,t}$	0.243	-4.634	-6.096	-0.794	-1.971	-4.708
	(0.459)	(-3.995)	(-3.081)	(-1.394)	(-1.366)	(-1.758)
$RV_{i,t}$	-0.489	0.029	0.271	-0.068	-2.272	-5.803
	(-0.603)	(0.011)	(0.056)	(-0.089)	(-0.919)	(-1.190)
$MOM_{i,t}$	-0.677	1.821	2.703	-0.234	2.434	5.805
	(-0.861)	(0.868)	(0.688)	(-0.360)	(1.186)	(1.420)
$ST_{i,t}$	-1.395	-1.731	-2.354	-0.761	-2.185	-6.247
	(-1.646)	(-0.985)	(-0.746)	(-1.482)	(-1.291)	(-1.982)
Constant	3.843	18.437	41.149	2.631	12.610	25.236
	(2.019)	(2.624)	(2.908)	(1.609)	(2.080)	(2.071)
Daily Average Firms	271.625	271.607	271.607	271.169	271.151	271.151
Adjusted R-squared	0.138	0.125	0.121	0.184	0.173	0.170

Panel B Analyst Coverage Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	2.777	3.869	5.093	1.36	1.653	0.621
	(7.515)	(5.215)	(4.575)	(3.936)	(2.527)	(0.572)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-1.217	-0.063	-0.976	-0.904	-1.385	-0.993
	(-3.855)	(-0.099)	(-0.984)	(-3.104)	(-2.394)	(-1.159)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-1.706	1.319	1.792	-1.044	-0.129	-1.284
	(-2.401)	(1.068)	(1.027)	(-1.951)	(-0.126)	(-0.883)
$Buzz_{i,t} * AbRet_{i,t}$	3.338	4.645	7.462	2.576	2.901	4.379
	(4.128)	(3.883)	(4.352)	(4.658)	-3.132	(3.228)
$BUZZ_{i,t}$	-0.063	-0.268	1.213	0.044	-0.594	0.649
	(-0.153)	(-0.357)	(1.0)	(0.113)	(-0.605)	(0.388)
$Buzz_{i,t} * Size_{i,t}$	-0.383	-0.428	-2.229	0.313	1.904	2.89
	(-0.935)	(-0.458)	(-1.451)	(0.918)	-2.165	(1.891)
$AbRet_{i,t}$	-4.637	-2.66	-1.841	-4.513	-7.167	-9.789
	(-5.598)	(-1.858)	(-0.878)	(-6.228)	(-5.214)	(-5.202)
$ABbTurn_{i,t}$	-7.013	-0.89	-4.739	-2.686	-5.112	-8.206
	(-3.577)	(-0.273)	(-0.926)	(-2.203)	(-2.004)	(-1.826)
$AbRet_{i,t-5,t-1}$	-1.437	-3.376	-4.173	-3.988	-6.29	-8.579
	(-1.985)	(-2.266)	(-1.747)	(-5.843)	(-4.048)	(-3.447)
$AbRet_{i,t} * AbtTurn_{i,t}$	0.789	-0.43	-0.656	0.228	1.644	1.041
	(1.412)	(-0.489)	(-0.491)	(0.479)	-1.727	(0.79)
$AbRet_{i,t} * Size_{i,t}$	-3.23	-3.719	-5.496	-1.178	-2.223	-2.679
	(-4.551)	(-2.957)	(-3.017)	(-1.869)	(-2.01)	(-1.6)
$ME_{i,t}$	-2.169	-5.6	-11.767	-1.148	-4.694	-10.855
	(-3.963)	(-3.495)	(-3.862)	(-2.118)	(-2.888)	(-3.626)
$BM_{i,t}$	-0.936	-1.154	-1.084	-0.494	-2.765	-4.264
	(-1.557)	(-0.711)	(-0.342)	(-0.766)	(-1.495)	(-1.204)
$OP_{i,t}$	0.147	2.098	4.879	-0.184	-0.663	-1.322
	(0.27)	(1.576)	(2.024)	(-0.353)	(-0.527)	(-0.536)
$IVN_{i,t}$	0.088	-3.701	-5.153	-0.501	-1.263	-4.753
	(0.187)	(-3.254)	(-2.504)	(-1.043)	(-0.85)	(-1.764)
$RV_{i,t}$	0.147	-0.157	-1.094	-0.36	-0.669	-1.077
	(0.18)	(-0.063)	(-0.225)	(-0.41)	(-0.24)	(-0.194)
$MOM_{i,t}$	-1.176	0.363	0.783	-0.134	2.936	5.576
	(-1.51)	(0.175)	(0.204)	(-0.185)	-1.371	(1.368)
$ST_{i,t}$	-0.841	-2.799	-3.098	-0.812	-1.057	-4.913
	(-0.96)	(-1.606)	(-0.931)	(-1.445)	(-0.601)	(-1.507)
$Constant$	3.552	17.06	37.934	2.62	12.709	27.871
	(1.953)	(2.518)	(2.78)	(1.533)	-2.028	(2.212)
Daily Average Firms	277	277	277	257.409	257.396	257.396
Adjusted R-squared	0.147	0.133	0.130	0.191	0.181	0.177

Panel C Analyst Forecast Dispersion Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	1.76	2.371	1.997	1.905	3.116	4.043
	(5.251)	(3.585)	(1.953)	(4.826)	(3.743)	(3.295)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-1.269	-0.301	0.2	-0.521	-0.956	-1.714
	(-3.99)	(-0.511)	(0.25)	(-1.595)	(-1.403)	(-1.731)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-0.877	-1.648	-1.203	-1.463	1.37	-2.239
	(-1.346)	(-1.49)	(-0.677)	(-1.869)	(1.101)	(-1.221)
$Buzz_{i,t} * AbRet_{i,t}$	3.699	3.319	4.078	3.86	2.824	4.505
	(5.416)	(3.074)	(2.718)	(5.136)	(2.307)	(2.455)
$BUZZ_{i,t}$	0.368	-0.37	0.1	-0.423	-0.025	1.123
	(1.067)	(-0.505)	(0.09)	(-0.938)	(-0.027)	(0.714)
$Buzz_{i,t} * Size_{i,t}$	-0.105	1.001	1.254	0.103	0.293	-0.592
	(-0.295)	(1.241)	(0.901)	(0.243)	(0.31)	(-0.366)
$AbRet_{i,t}$	-4.8	-9.15	-11.04	-5.06	-2.029	-3.391
	(-6.44)	(-6.092)	(-5.928)	(-6.063)	(-1.283)	(-1.498)
$ABbTurn_{i,t}$	-2.435	-0.358	-5.606	-4.91	-1.188	-4.997
	(-1.817)	(-0.088)	(-0.727)	(-3.052)	(-0.41)	(-1.036)
$AbRet_{i,t-5,t-1}$	-4.345	-6.648	-9.987	-2.543	-4.293	-4.958
	(-6.426)	(-4.49)	(-4.81)	(-3.641)	(-2.589)	(-1.873)
$AbRet_{i,t} * AbtTurn_{i,t}$	-0.163	0.723	1.431	0.439	-0.208	-1.571
	(-0.339)	(0.728)	(0.856)	(0.707)	(-0.233)	(-1.159)
$AbRet_{i,t} * Size_{i,t}$	-2.868	-3.536	-5.058	-3.176	-4.184	-5.272
	(-4.278)	(-3.348)	(-2.922)	(-4.213)	(-3.343)	(-2.638)
$ME_{i,t}$	-1.896	-4.015	-10.012	-0.996	-4.411	-9.577
	(-3.288)	(-2.611)	(-3.448)	(-1.58)	(-2.575)	(-2.935)
$BM_{i,t}$	-0.415	-0.674	0.351	-0.646	-0.626	0.256
	(-0.608)	(-0.374)	(0.1)	(-0.97)	(-0.332)	(0.071)
$OP_{i,t}$	-0.379	0.108	1.435	0.362	1.486	3.445
	(-0.649)	(0.085)	(0.595)	(0.667)	(1.107)	(1.474)
$IVN_{i,t}$	0.166	-1.322	-1.578	-0.804	-4.393	-7.929
	(0.355)	(-1.106)	(-0.741)	(-1.477)	(-3.192)	(-3.177)
$RV_{i,t}$	0.658	3.278	5.933	0.166	-1.217	-3.516
	(0.869)	(1.419)	(1.301)	(0.189)	(-0.452)	(-0.677)
$MOM_{i,t}$	-0.886	0.732	1.796	-1.87	-1.5	-3.279
	(-1.229)	(0.369)	(0.471)	(-2.515)	(-0.7)	(-0.797)
$ST_{i,t}$	-0.892	-3.584	-6.866	-0.983	-1.435	-2.614
	(-1.455)	(-2.198)	(-2.377)	(-1.372)	(-0.658)	(-0.697)
$Constant$	4.449	17.67	37.831	1.841	11.891	25.376
	(2.951)	(3.147)	(3.37)	(0.882)	(1.566)	(1.656)
Daily Average Firms	260	260	260	259	259	259
Adjusted R-squared	0.161	0.142	0.141	0.169	0.154	0.148

Panel D Illiquidity Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	1.142	1.501	0.088	2.86	3.088	4.998
	(3.638)	(2.365)	(0.087)	(6.971)	(4.025)	(4.205)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-0.883	-0.368	0.428	-1.252	-0.23	-0.925
	(-3.16)	(-0.733)	(0.588)	(-3.613)	(-0.34)	(-0.929)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-0.533	0.635	1.103	-1.934	1.67	-1.73
	(-1.025)	(0.673)	(0.795)	(-2.674)	(1.259)	(-0.927)
$Buzz_{i,t} * AbRet_{i,t}$	2.339	3.522	4.117	4.453	4.742	8.101
	(4.464)	(3.941)	(2.948)	(5.388)	(3.604)	(4.243)
$BUZZ_{i,t}$	0.112	-0.653	0.926	-0.132	-0.396	1.011
	(0.308)	(-0.719)	(0.584)	(-0.308)	(-0.488)	(0.783)
$Buzz_{i,t} * Size_{i,t}$	0.19	1.156	1.182	-0.206	0.538	-1.462
	(0.604)	(1.429)	(0.825)	(-0.49)	(0.618)	(-1.043)
$AbRet_{i,t}$	-4.759	-8.53	-11.667	-5.213	-1.545	-1.018
	(-6.807)	(-6.772)	(-6.784)	(-6.079)	(-1.029)	(-0.478)
$ABbTurn_{i,t}$	-5.283	1.509	0.798	-6.076	0.316	-1.511
	(-3.171)	(0.589)	(0.171)	(-3.05)	(0.088)	(-0.276)
$AbRet_{i,t-5,t-1}$	-3.895	-7.317	-10.087	-1.543	-1.96	-1.696
	(-5.843)	(-5.048)	(-4.183)	(-2.033)	(-1.21)	(-0.704)
$AbRet_{i,t} * AbtTurn_{i,t}$	0.838	2.026	1.407	0.749	-2.025	-2.42
	(1.503)	(2.166)	(1.165)	(1.273)	(-2.075)	(-1.772)
$AbRet_{i,t} * Size_{i,t}$	0.318	-1.192	-1.979	-2.679	-3.745	-4.574
	(0.561)	(-1.133)	(-1.271)	(-3.668)	(-2.899)	(-2.596)
$ME_{i,t}$	-1.114	-2.809	-7.411	-1.73	-4.933	-9.84
	(-2.333)	(-2.01)	(-2.738)	(-3.1)	(-3.166)	(-3.419)
$BM_{i,t}$	-0.723	-1.413	-2.445	-0.785	-2.286	-1.666
	(-1.254)	(-0.836)	(-0.751)	(-1.269)	(-1.302)	(-0.514)
$OP_{i,t}$	0.175	0.089	0.282	-0.386	1.065	3.374
	(0.385)	(0.077)	(0.127)	(-0.689)	(0.706)	(1.329)
$IVN_{i,t}$	-0.514	-1.273	-4.151	0.276	-4.791	-6.807
	(-0.98)	(-0.92)	(-1.581)	(0.535)	(-4.155)	(-3.519)
$RV_{i,t}$	0.754	-2.117	-4.153	-0.784	0.932	1.288
	(0.893)	(-0.791)	(-0.792)	(-0.987)	(0.384)	(0.272)
$MOM_{i,t}$	-0.236	2.36	6.442	-1.302	0.631	0.369
	(-0.34)	(1.136)	(1.585)	(-1.698)	(0.305)	(0.095)
$ST_{i,t}$	-0.096	-0.127	-1.676	-1.715	-4.569	-6.553
	(-0.18)	(-0.072)	(-0.506)	(-1.865)	(-2.604)	(-1.972)
$Constant$	2.211	12.088	25.489	3.84	18.393	40.803
	(1.328)	(1.962)	(2.055)	(2.029)	(2.647)	(2.914)
Daily Average Firms	272	272	272	271	271	271
Adjusted R-squared	0.190	0.180	0.177	0.141	0.125	0.122

Panel E Institutional Ownership Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	2.778	1.942	2.29	2.406	3.074	3.799
	(3.365)	(1.149)	(0.905)	(3.15)	(1.899)	(1.594)
$EmotionVsFact_{i,t} * Sentiment_{i,t}$	-1.005	-0.62	-1.005	-1.106	-1.406	-3.344
	(-1.571)	(-0.436)	(-0.518)	(-1.721)	(-1.045)	(-1.858)
$EmotionVsFact_{i,t} * AbRet_{i,t}$	-2.264	-3.645	-7.842	-0.899	3.594	2.011
	(-1.794)	(-1.185)	(-1.888)	(-0.701)	(1.858)	(0.718)
$Buzz_{i,t} * AbRet_{i,t}$	-0.084	4.79	5.854	2.754	4.163	3.573
	(-0.051)	(1.769)	(1.634)	(2.057)	(1.664)	(1.013)
$BUZZ_{i,t}$	-1.33	-1.757	-0.713	-0.044	2.773	1.975
	(-1.824)	(-1.105)	(-0.25)	(-0.053)	(1.553)	(0.81)
$Buzz_{i,t} * Size_{i,t}$	0.784	0.445	-2.177	-0.847	-3.855	-7.218
	(0.841)	(0.225)	(-0.695)	(-1.096)	(-2.451)	(-2.907)
$AbRet_{i,t}$	-3.938	-6.846	0.934	-5.32	-4.632	-5.251
	(-2.399)	(-2.136)	(0.212)	(-3.195)	(-1.706)	(-1.463)
$ABbTurn_{i,t}$	-5.884	-6.452	-16.193	-0.212	-2.444	-2.599
	(-1.76)	(-0.967)	(-1.45)	(-0.076)	(-0.516)	(-0.407)
$AbRet_{i,t-5,t-1}$	-1	0.107	1.084	-3.141	-3.813	-3.996
	(-0.687)	(0.03)	(0.183)	(-2.298)	(-1.433)	(-0.918)
$AbRet_{i,t} * AbtTurn_{i,t}$	1.076	0.392	-1.48	0.005	-0.764	0.343
	(0.933)	(0.189)	(-0.419)	(0.005)	(-0.532)	(0.181)
$AbRet_{i,t} * Size_{i,t}$	-2.633	-8.505	-11.968	-1.936	-4.178	-3.111
	(-1.77)	(-2.835)	(-2.807)	(-1.488)	(-2.047)	(-1.026)
$ME_{i,t}$	0.232	-3.913	-14.418	-0.797	-3.789	-6.178
	(0.195)	(-1.079)	(-2.169)	(-0.691)	(-1.257)	(-1.138)
$BM_{i,t}$	2.019	2.113	3.055	-0.145	0.663	8.899
	(1.514)	(0.516)	(0.403)	(-0.11)	(0.185)	(1.238)
$OP_{i,t}$	1.296	1.004	4.84	1.377	4.925	13.532
	(1.162)	(0.359)	(1.023)	(1.112)	(1.579)	(2.551)
$IVN_{i,t}$	0.349	-2.145	1.307	0.102	-2.791	-7.858
	(0.308)	(-0.775)	(0.248)	(0.104)	(-1.103)	(-1.616)
$RV_{i,t}$	1.144	2.346	3.261	-0.74	-1.33	-3.605
	(0.613)	(0.415)	(0.308)	(-0.543)	(-0.297)	(-0.413)
$MOM_{i,t}$	-1.171	1.842	2.742	-0.843	4.457	12.514
	(-0.847)	(0.502)	(0.37)	(-0.55)	(0.984)	(1.538)
$ST_{i,t}$	-0.354	-5.154	-11.613	-1.496	-3.999	-8.967
	(-0.26)	(-1.458)	(-1.645)	(-1.163)	(-1.143)	(-1.45)
$Constant$	1.538	15.635	36.591	5.221	21.748	51.6
	(0.456)	(1.274)	(1.512)	(1.478)	(1.682)	(2.031)
Daily Average Firms	262	262	262	262	262	262
Adjusted R-squared	0.188	0.171	0.172	0.148	0.134	0.136

their information acquisition decision, in equilibrium, is subject to biased beliefs rather than to rational expectations. To conclude, stocks with high institutional ownership show relatively stronger empirical results from firm-specific news sentiment than stocks with low institutional ownership.

In sum, even though the above robustness test shows that news sentiment has a stronger impact of return predictability on small and illiquid stocks, on average, the cross-sectional variation of stock returns resulting in information risk implied by firm-specific news sentiment is robust in all sub-samples, which may imply potential problems in stocks such as information asymmetry (size and illiquidity), investors' alternative beliefs (analyst forecast coverage and dispersion) and better-informed investors (institutional ownership).

B.9.5 q-factor model testing

Hou et al. (2015) develop an empirical asset pricing model known as the q-factor. The q-factor model indicates that expected excess returns can be explained by the sensitivities of the market factor, a size factor, an investment factor and a return on equity factor. More importantly, Hou, Xue and Zhang conduct comprehensive empirical testing on existing anomalies in cross-sectional stock returns and demonstrate the strong explanatory power of the q-factor model. The authors argue that the q-factor is a very competitive alternative to the Fama–French five factors model.

Therefore, I re-conduct all the tests in section 2.5.3 with the q-factor model. Panel A in Table B.5 reports Pearson correlations between the news sentiment factor and factors from the q-factor model. Clearly, there are almost no economically significant correlations between the firm-specific news sentiment factor and other factors. Panel B in Table B.5 shows risk-adjusted alphas across different models by running time-series regressions of the zero-cost news sentiment portfolio on the base line q-factor model and adding additional factors as customary controls. Essentially, the results are as expected and are in line with the findings of the Fama–French factor models. The news sentiment zero-cost portfolio maintains positive significant daily abnormal returns in the range of 6.1 to 6.5 basis point. None of the factors from the baseline q-factor model have strong explanatory power on the cross-sectional variation of stock returns, where returns are the result of a deviation in information asymmetry attributed to the biased tone in firm-specific news sentiment. Overall, traditional asset pricing factors developed based on firm fundamentals, with either the Fama–French five factors or the novel q-factor, lack the capability to capture the pricing effect caused by firm-specific news sentiment.

Table B.5: Firm-Specific News Sentiment Factor-q-factor Model Testing

This table shows daily risk-adjusted returns (α) from a firm-specific news sentiment zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment: $N/S; N/B; M/S; M/B; P/S; P/B$ sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% ($P/S; P/B$) and the average of short position in the two negative sentiment portfolios 30% ($N/S; N/B$). Each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows the Pearson correlation between the news sentiment portfolio return and pricing factors from the q-factor model. Panel B presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of q-factor with Pastor and Stambaugh liquidity factor, momentum factor and short- and long- term reversal factors. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	$R_{MKT,t}$	$R_{ME,t}$	$R_{IA,t}$	$R_{ROE,t}$	$R_{EG,t}$
$Sentiment_t$	-0.108	0.019	0.069	0.095	0.092
$R_{MKT,t}$		0.148	-0.308	-0.385	-0.36
$R_{ME,t}$			0.019	-0.172	-0.259
$R_{IA,t}$				0.187	0.14
$R_{ROE,t}$					0.577

Panel B Risk-Adjusted Firm-Specific News Sentiment Zero-Cost Portfolio Returns by q-factor Model

	$Sentiment_t$	q-factor	q-factor + PLS	q-factor+UMD	q-factor+Full
α	0.066	0.062	0.062	0.060	0.065
t_α	(6.397)	(6.009)	(6.005)	(5.921)	(6.453)
$R_{MKT,t}$		-0.045	-0.046	-0.030	-0.014
$t_{R_{MKT}}$		(-3.115)	(-3.198)	(-2.274)	(-0.942)
$R_{ME,t}$		0.059	0.059	0.027	0.036
$t_{R_{ME}}$		(2.131)	(2.104)	(1.076)	(1.318)
$R_{IA,t}$		0.053	0.050	0.083	0.077
$t_{R_{IA}}$		(1.437)	(1.385)	(2.412)	(1.974)
$R_{ROE,t}$		0.059	0.058	-0.073	-0.067
$t_{R_{ROE}}$		(1.544)	(1.538)	(-1.895)	(-1.712)
$R_{EG,t}$		0.090	0.093	0.081	0.045
$t_{R_{EG}}$		(2.173)	(2.264)	(2.055)	(1.109)
$PSLIQ_t$			0.014	0.019	0.019
t_{PSLIQ}			(0.849)	(1.132)	(1.187)
UMD_t				0.155	0.154
t_{UMD}				(8.518)	(8.514)
ST_t					-0.087
t_{ST}					(-4.509)
LT_t					-0.043
t_{LT}					(-1.412)
$\overline{R}^2$	0.007	0.018	0.019	0.048	0.057
$Days$	5241	5241	5241	5241	5241

B.9.6 News Sentiment Factors vs. Other News Factors ?

Section 2.5.3 demonstrates that the zero-cost portfolio formed by daily firm-specific news sentiment generates a considerable amount of daily abnormal returns (α) which cannot be fully explained by customary factors from empirical asset pricing. I argue that this abnormal return from the theoretical implication in section 2.2 regarding firm-specific news sentiment causes a deviation in information risk, for which investors require compensation. However, as mentioned above in section 2.5.2, extant studies argue that firm-specific news sentiment contains firm value-relevant information. In fact, the Fama–Macbeth (1973) regression results in Table 2.5 confirm this finding from the significant coefficients of the controlled variable $EmotionVsFact_{i,t} * Sentiment_{i,t}$. Owing to the lack of explanatory power of traditional asset pricing factors, there may be a concern that the daily abnormal returns sorted by firm-specific news sentiment could be captured by other novel factors from quantified news measures.

Hence, I consider constructing an additional factor based on the empirical evidence of the news-related variables from Table 2.5. More specifically, $EmotionVsFact_{i,t} * Sentiment_{i,t}$ verifies the significance of news sentiment, including both genuine information and the biasing effect on investors' valuation of firm fundamentals. In fact, this interacted variable $EmotionVsFact_{i,t} * Sentiment_{i,t}$ presents an intriguing finding: news sentiment has a segmented effect between 'soft' information (for example, emotional or opinion references) and 'hard' information (for instance, firm-fundamental or factual references) in the news. On the one hand, the 'soft' information that is more focused on emotional references is more likely to bias investors' rational decisions. On the other hand, the 'hard' information – specifically, factual or fundamental information such as accounting details or earnings – is more helpful for investors to understand a company's business condition and will potentially lead investors to dissolve value-relevant information as they may be uninformed without reading the news.

Essentially, it should be noted that both of the terms from $EmotionVsFact_{i,t} * Sentiment_{i,t}$ range from [-1, +1]. Hence, the value of $EmotionVsFact_{i,t} * Sentiment_{i,t}$ has two implications depending on whether or not sentiment conditioned by the type of information (soft versus hard) generates genuine information or leads to a biased evaluation of the firm by investors.

First, taking the behavioral perspective, higher values of $EmotionVsFact_{i,t}$ aligned with firm-specific news sentiment ($Sentiment_{i,t}$) imply that sentiment is more likely to drive from emotional references, to make investors more biased about the firm valuation. In this case, for example, a positive sentiment or optimistic tone from more emotional references as the value of $EmotionVsFact_{i,t}$ increases in firm-specific news is more likely to cause investors to overprice the value of a firm. Once the value of a firm moves back to its fundamental value, firm-specific news sentiment predicting a reversal appears as the mispricing is corrected.

Second, taking the genuine information of instructing the firm fundamentals argument concerning firm-specific news, a higher value of $EmotionVsFact_{i,t} * Sentiment_{i,t}$ implies more negative fundamental information in the news. In this instance, both $Sentiment_{i,t}$ and $EmotionVsFact_{i,t}$ decline, causing their interaction value to become higher. For instance, a value of lower $EmotionVsFact_{i,t}$ means that there is more fundamental information in the firm-specific news, necessitating a lower value of $Sentiment_{i,t}$. Therefore, the interaction indicates negative fundamental information about the company, which can be acquired through investors' reading. As a result, investors correctly lower the valuation of the firm based on news containing more negative value-relevant information about the firm. The higher value of the interaction predicts lower stock future returns. Following the implication of $EmotionVsFact_{i,t} * Sentiment_{i,t}$, I construct an additional news factor to capture the effect of either biased valuation or genuine information from news sentiment about particular companies. The empirical results in Table 2.5 demonstrate that $EmotionVsFact_{i,t} * Sentiment_{i,t}$ predicts negative cross-sectional stock future returns (R^e_{t+1}). Therefore, portfolio sorting is based on the standardized value of $EmotionVsFact_{i,t} * Sentiment_{i,t}$ at day t. As mentioned by Tetlock (2011), sorting on this standardized interacting variable produces a similar result to sorting both of the variables. Next, the construction of the mimicking portfolio is the same as the firm news sentiment portfolio in section 2.5.2, but based on the value of $EmotionVsFact_{i,t} * Sentiment_{i,t}$. I construct the zero-cost portfolio to take the average of long positions, with news either containing the most emotional (E) negative sentiment (N) or the most fundamental (F) positive sentiment (P) 30% ($EN(FP)/S; EN(FP)/B$)[2] and the average of short positions in the stocks with either the most emotional positive sentiment or the most fundamental negative 30% ($EP(FN)/S; EP(FN)/B$).[3] In other words, the profit of this zero-cost portfolio drives from buying stocks with good news or which are undervalued due to bias ($EmotionVsFact_{i,t} * Sentiment_{i,t} \downarrow$) and selling stocks with bad news or which are overvalued due to bias ($EmotionVsFact_{i,t} * Sentiment_{i,t} \uparrow$). Lastly, I calculate the next day $t + 1$ value-weighted portfolio returns from this zero-cost trading strategy.

On the one hand, the zero-cost portfolio based on $EmotionVsFact_{i,t}*Sentiment_{i,t}$ (hereafter $EFSENT_{i,t}$) earns positive significant abnormal returns by about 2.7 basis

[2] Both these portfolios predict the positive stock future returns shown in Table 2.5.

[3] These two portfolios negatively predict the next day's stock returns, and the empirical result in Table 2.5 confirms that.

Table B.6: Latent Information of Firm-Specific News Sentiment Factor Risk Premium Testing

This table shows daily risk-adjusted returns (α) from a zero-cost portfolio constructed based on $EmotionVsFact_{i,t} * Sentiment_{i,t}$ ($EFSENT_{i,t}$) as a proxy of latent information from news sentiment factors, to capture the potential effects of genuine information or biased valuation about firm fundamentals for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on $EFSENT_{i,t}$ at day t into three sentiment portfolios: emotional (factual) and optimistic (pessimistic) ($EP(FN)$) 30%, neutral (M) 40%, emotional (factual) and pessimistic (optimistic) ($EN(FP)$) 30%. The six interacted value-weighted portfolios respecting size and latent information from news sentiment: $EP(FN)/S, EP(FN)/B, M/S, M/B, EN(FP)/S, EN(FP)/B$ sorting on the size and $EFSENT_{i,t}$ independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the stocks with news either containing the most emotional negative sentiment or the most fundamental positive sentiment 30% ($EN(FP)/S, (EN)FP/B$) and the average of short position in the stocks with either the most emotional positive sentiment or the most fundamental negative sentiment 30% ($EP(FN)/S, EP(FN)/B$). Each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows the Pearson correlation between the latent information of news sentiment portfolio return and conventional factors. Panel B presents the risk-adjusted return of the latent information of news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	MKT_t	SMB_t	HML_t	RMW_t	CMA_t	UMD_t	ST_t	LT_t	$PSLIQ_t$
$EFSENT_t$	0.0164	-0.0023	-0.0004	-0.0228	-0.0124	0.0297	0.0386	0.0029	0.0013
MKT_t		0.0702	-0.0119	-0.4246	-0.3331	-0.2572	0.3547	-0.0837	0.0850
SMB_t			0.0523	-0.2985	0.0553	0.0286	0.0138	0.2834	0.0422
HML_t				0.0876	0.4831	-0.3438	-0.0973	0.4771	0.0976
RMW_t					0.2797	0.1508	-0.2453	-0.1613	0.0402
CMA_t						0.0651	-0.2835	0.5202	0.0252
UMD_t							-0.1260	0.0301	-0.0656
ST_t								-0.1380	0.0607
LT_t									-0.0286

Panel B Risk-Adjusted Latent Information of Firm-Specific News Sentiment Zero-Cost Portfolio Returns

	$EFSENT_t$	$CAPM$	$FF3$	$FF5$	$FF5+UMD$	$FF5+Full$
α	0.027	0.026	0.026	0.027	0.026	0.025
t_α	(2.859)	(2.831)	(2.840)	(2.941)	(2.867)	(2.721)
MKT_t		0.009	0.010	0.003	0.008	0.002
t_{MKT}		(0.831)	(0.816)	(0.275)	(0.648)	(0.182)
SMB_t			-0.004	-0.011	-0.015	-0.014
t_{SMB}			(-0.199)	(-0.548)	(-0.789)	(-0.726)
HML_t			0.000	0.005	0.028	0.030
t_{HML}			(-0.001)	(0.259)	(1.218)	(1.306)
RMW_t				-0.030	-0.036	-0.032
t_{RMW}				(-0.969)	(-1.162)	(-0.980)
CMA_t				-0.010	-0.025	-0.014
t_{CMA}				(-0.223)	(-0.548)	(-0.288)
$PSLIQ_t$			0.000	0.001	0.002	0.000
t_{PSLIQ}			(-0.001)	(0.084)	(0.130)	(0.035)
UMD_t					0.034	0.036
t_{UMD}					(2.193)	(2.391)
ST_t						0.029
t_{ST}						(1.749)
LT_t						-0.003
t_{LT}						(-0.100)
$\overline{R}^2$	0.001	0.000	0.000	0.000	0.001	0.002
Days	5241	5241	5241	5241	5241	5241

Table B.7: Risk-Adjusted Firm-Specific News Sentiment Zero-Cost Portfolio Returns Controlling $EFSENT_{i,t}$.

This table shows daily risk-adjusted returns (α) from firm-specific news zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalization from the last month to split stocks into two portfolios sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment:$N/S, N/B, M/S, M/B, , P/S, P/B$ sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% ($P/S, P/B$) and the average of short position in the two negative sentiment portfolios 30% ($N/S, N/B$) each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. By adding an additional pricing factor $EFSENT_{i,t}$ – an invented news factor capturing latent information in news sentiment such as genuine information or biased valuation about firm fundamentals – the table presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

	$Sentiment_t$	$CAPM$	$FF3$	$FF5$	$FF5+UMD$	$FF5+Full$
α	0.067	0.069	0.069	0.065	0.063	0.067
t_α	(6.567)	(6.758)	(6.929)	(6.554)	(6.331)	(6.808)
$EFSENT_t$	-0.063	-0.062	-0.061	-0.060	-0.066	-0.062
t_{EFSENT}	(-1.649)	(-1.611)	(-1.616)	(-1.601)	(-1.752)	(-1.663)
MKT_t		-0.065	-0.069	-0.031	-0.016	0.001
t_{MKT}		(-4.734)	(-5.272)	(-2.555)	(-1.324)	(0.048)
SMB_t			0.051	0.055	0.040	0.035
t_{SMB}			(1.953)	(2.252)	(1.665)	(1.375)
HML_t			-0.123	-0.202	-0.124	-0.133
t_{HML}			(-4.260)	(-7.380)	(-4.750)	(-4.989)
RMW_t			0.021	0.056	0.036	0.027
t_{RMW}			(1.156)	(1.619)	(1.056)	(0.732)
CMA_t				0.234	0.183	0.146
t_{CMA}				(5.351)	(4.327)	(3.143)
$PSLIQ_t$			0.021	0.018	0.020	0.024
t_{PSLIQ}			(1.151)	(1.035)	(1.181)	(1.436)
UMD_t					0.117	0.112
t_{UMD}					(6.826)	(6.848)
ST_t						-0.086
t_{ST}						(-4.595)
LT_t						0.018
t_{LT}						(0.562)
R^2	0.011	0.014	0.027	0.038	0.055	0.064
$Days$	5241	5241	5241	5241	5241	5241

point per day (6.8% annualized abnormal return). Panel A in Table B.6 presents Pearson correlations between the $EFSENT_{i,t}$ and other fundamental factors. Clearly, there is hardly any correlation between $EFSENT_{i,t}$ and extant classical factors. Panel B shows risk-adjusted alphas based on different models. In fact, there is little reduction of abnormal return earned by sorting $EFSENT_{i,t}$ when controlling for other pricing factors across column (2)–(6). The full specification in column (6) is only reduced by 0.2 basis points. Therefore, the information implied by this novel news pricing factor $EFSENT_{i,t}$ cannot be explained by the existing fundamental pricing factors.

On the other hand, I add $EFSENT_{i,t}$ as an additional novel pricing factor to adjust for doubted latent effects such as genuine information contained in the news sentiment about the firm fundamentals or mis-valuation of a company caused by news sentiment. First the Pearson correlation between the news sentiment factor and the $EFSENT_{i,t}$ factor is about -0.058. Second, as shown in Table B.7, on average, the $EFSENT_{i,t}$ factor is only significantly negative around the 10% level to explain the abnormal return from the news sentiment factor. Even though $EFSENT_{i,t}$ may capture some effects in the firm-specific news sentiment to decrease the abnormal return from news sentiment portfolio-for example, good fundamental information about the firm or investors' undervaluation-the total amount of explanatory power from the $EFSENT_{i,t}$ factor is not economically significant. In fact, the abnormal return owing to the news sentiment factors remains almost at the same degree seen in Table 2.6 without controlling for the additional news effect factor $EFSENT_{i,t}$.

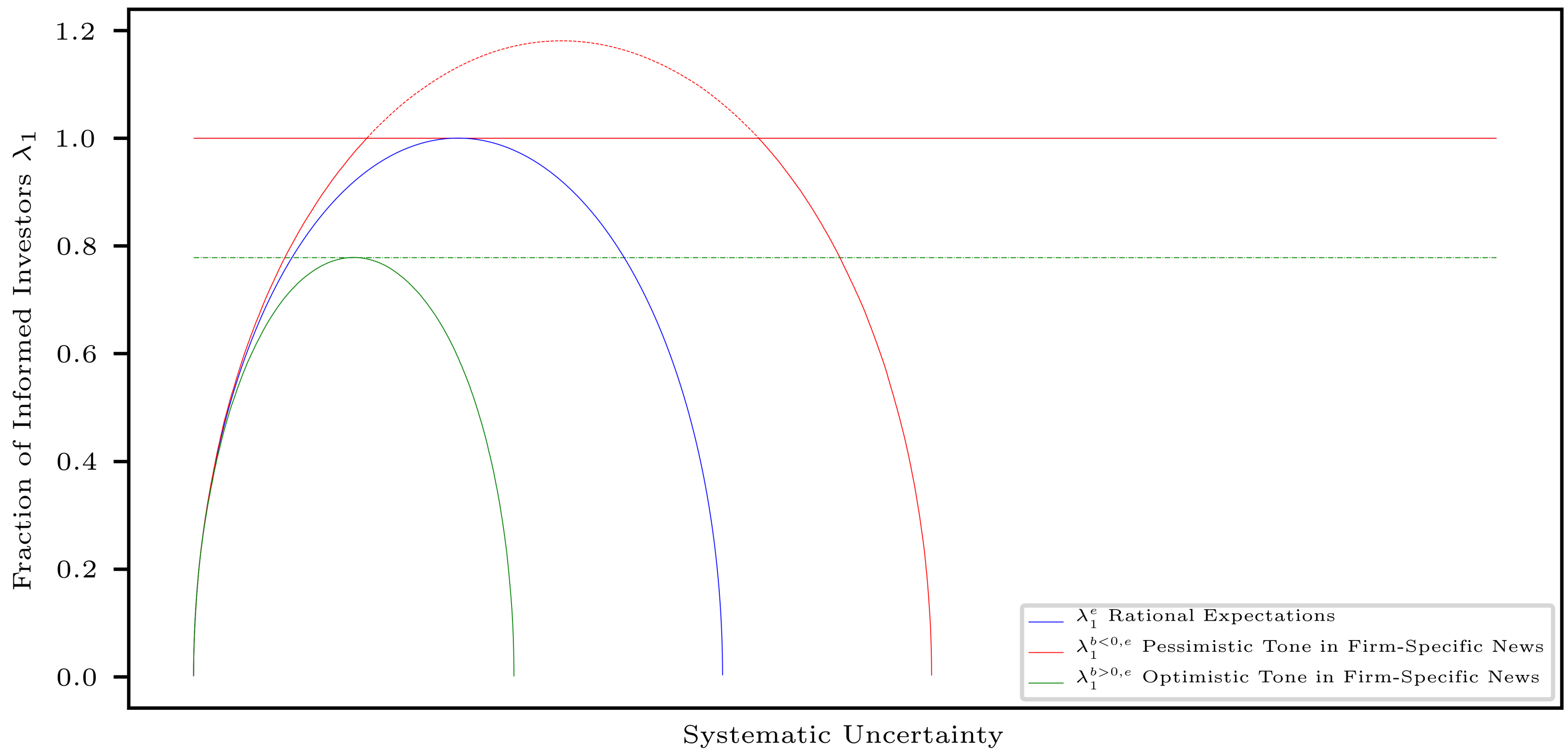
In sum, the abnormal excess return generated by firm-specific news sentiment that causes the deviation for information risk in assets is robust for both traditional fundamental factors in empirical asset pricing and the novel news factor which I propose in this study to capture either potential value-relevant information or mis-valuation effects from the firm news sentiment.

B.10 Additional Figures

In Figure B.1, I plot the equilibrium fraction of informed investors as a function of rational perception of market uncertainty respecting biased beliefs of firm-specific uncertainty.

FIGURE B.1: Firm-Specific News Sentiment Impact on Information Acquisition as a Function of $Var_1^I[D_2]$

This figure plots the equilibrium fraction of investors who want to acquire firm-specific e_1 as a function of unbiased systematic uncertainty $Var_1^I[D_2]$ in a biased perception of firm-specific uncertainty $\sigma_{b,e}^2$ by investors. I calibrate cost of information about e_1, $c = 0.002$; the variance of supply $\sigma_x^2 = 0.2$; the variance of the public signal noise σ_η^2 is randomly drawn from a uniform distribution; and investors' coefficient of risk aversion $\alpha = 3$. I set the rational perception of the firm-specific uncertainty $\sigma_e^2 = 0.0623$ as the median value of the sum of squared residuals from AR(1) of firms' $EPS_{i,t}$. The rational perception of systematic component uncertainty is randomly drawn from the distribution of S&P 500 realized volatility in the sample period. Without loss of generality, I assume the bias function $\beta(S_e, \sigma_e^2)$ in equation (2.2) is linear as $\sigma_{b,e}^2 = (1 - S_e)\sigma_e^2$ to generate biased perception of firm-specific uncertainty. The blue curve is the equilibrium fraction of informed investors under the rational expectations without the biased impact from firm-specific news sentiment ($S_e = 0$) on σ_e^2. The red curve is the equilibrium fraction of informed investors as $\sigma_{b,e}^2 \uparrow$ is upward (positively) biased when the tone in the news about particular firms' performance is decreased, or made more pessimistic, by one standard deviation $S_e \downarrow$ from $S_e = 0$ under the rational expectations. The green curve is the equilibrium fraction of informed investors as $\sigma_{b,e}^2 \downarrow$ is downward (negatively) biased when the tone in the news about firm-specific information is increased, or made more optimistic, by one standard $S_{b,e} \uparrow$ from $S_e = 0$ under the rational expectations. The standard deviation of firm-specific news sentiment σ_{S_e} is from the TRMI news sentiment indices of U.S.-listed firms and the value is 0.394.



Appendix C

Chapter 3 Appendix

C.1 Bitcoin Bubble Dating Calculation

We mainly follow the study by Phillips and Yu (2011) and Phillips et al. (2011) to dating the timeline of Bitcoin bubble during the irrational fanaticism in cryptocurrency market from April 2007 to February 2018. We first run recursive least square regression and estimate the autoregressive specification for Bitcoin price:

$$P_t = \mu + \delta P_{t-1} + \epsilon_t \quad \epsilon_t \sim i.i.d.\ (0, \sigma^2) \tag{C.1}$$

The independent and identically distributed ($i.i.d.$) assumption can also be relaxed to serially dependent errors. The null hypothesis is $H_0 : \delta = 1$ and the right-tailed alternative hypothesis is $H_1 : \delta > 1$ which indicates mildly explosive behavior in the process of Bitcoin price. We initialize our first recursion with 140 observations ($\tau_0 = nr_0$, which $r \in (0, 1]$ is a ratio of partitions to entire sample size n). The corresponding coefficient test statistics and Dickey-Fuller t statistics by DF_r^t, namely

$$DF_r^t := \left(\frac{\sum_{j-1}^{\tau} \tilde{X}_{j-1}^2}{\hat{\sigma}_\tau^2} \right) (\hat{\delta}_\tau - 1) \tag{C.2}$$

The successive observations in the subsequent regressions after the first initialization is $\tau = \lfloor nr \rfloor$. $\hat{\sigma}_\tau^2$ is the corresponding estimate of σ^2. $\tilde{X}_{j-1} = X_{j-1} - \frac{\sum_{j=1}^{\tau} X_{j-1}}{\tau}$. The critical value we use to compare the statistical value of Dickey-Fuller test is $cv_{\beta_n}^{df} = -0.08 + \ln(\lfloor nr \rfloor)/C$. Without loss of generality, we choose $C = 5$ to give a conservative test as Phillips and Yu (2011) suggests.

In Phillips and Yu (2011) study, they define the origination of the bubble by estimate $\hat{\tau}_e = \lfloor n\hat{r}_e \rfloor$ as flowing :

$$\hat{r}_e = \inf_{s \geq r_0} \{s : DF_s^t > cv_{\beta_n}^{df}\} \tag{C.3}$$

and the collapse of bubble by $\hat{\tau}_f = \lfloor n\hat{r}_f \rfloor$ as following :

$$\hat{r}_f = \inf_{s \geq \hat{r}_e + \gamma \ln(n)/n} \{s : DF_s^t < cv_{\beta_n}^{df}\} \tag{C.4}$$

However, the test statistics often dropped below or rose above the relevant critical values between these dates (see Figure 3.11). In real time, dating the bubble would have been difficult wholly based on this approach. But our interest is in historically dating the bubble solely to split our sample into the pre-bubble, bubble and post-bubble periods. We define to find the bubble collapse date as days after the collapse date is consistently showing no mildly explosive behavior in Bitcoin price process.

Finally, we conduct initialization to improve the dating process based on the procedure by Phillips et al. (2011) study. The Bitcoin bubble we find is from May 24, 2017, and the bubble collapse on January 28, 2018.

For full detailed information, we refer to review the comprehensive studies from Phillips and Yu (2011) and Phillips et al. (2011).

C.2 Additional Figures

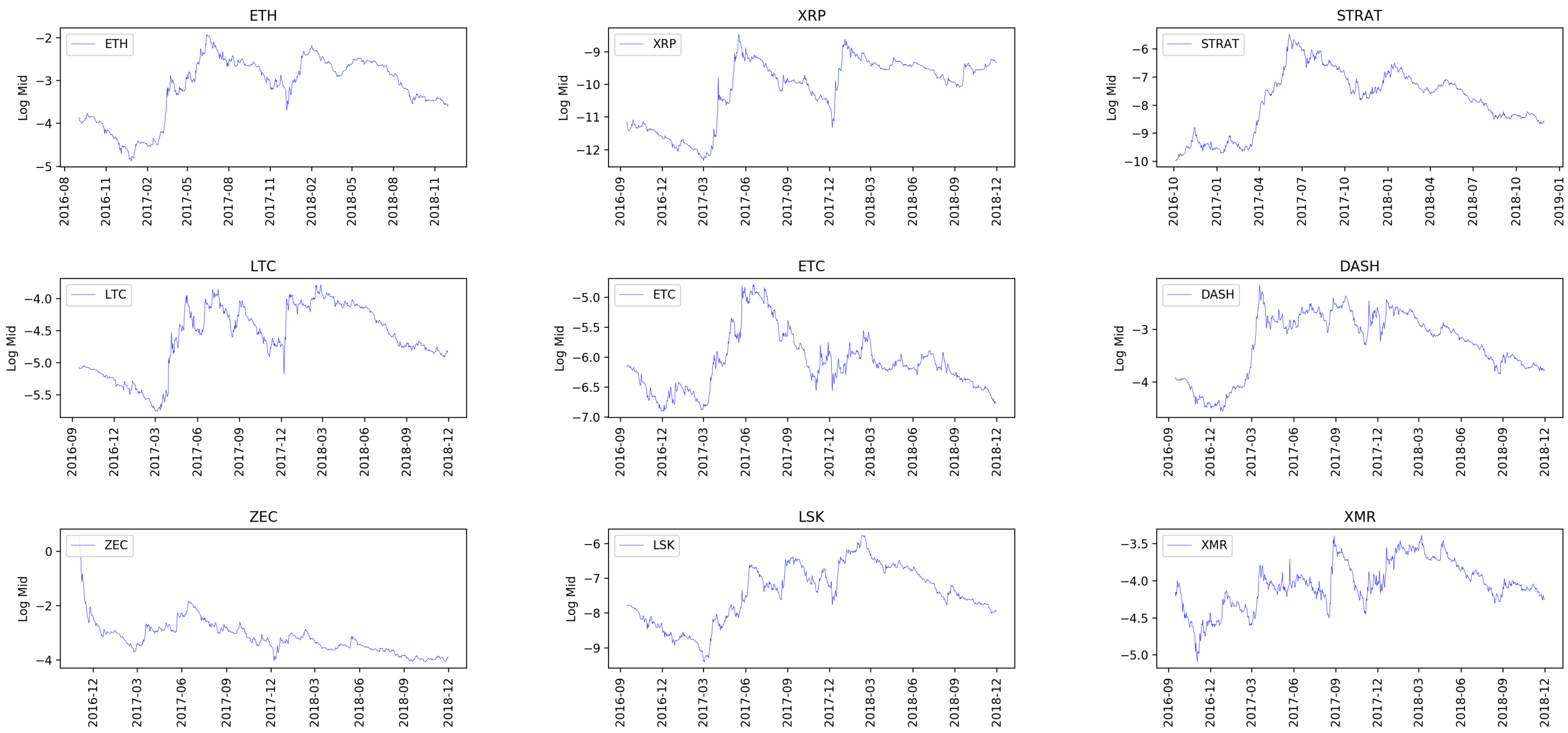


Figure C.1: Log Mid Price of Cryptocurrency

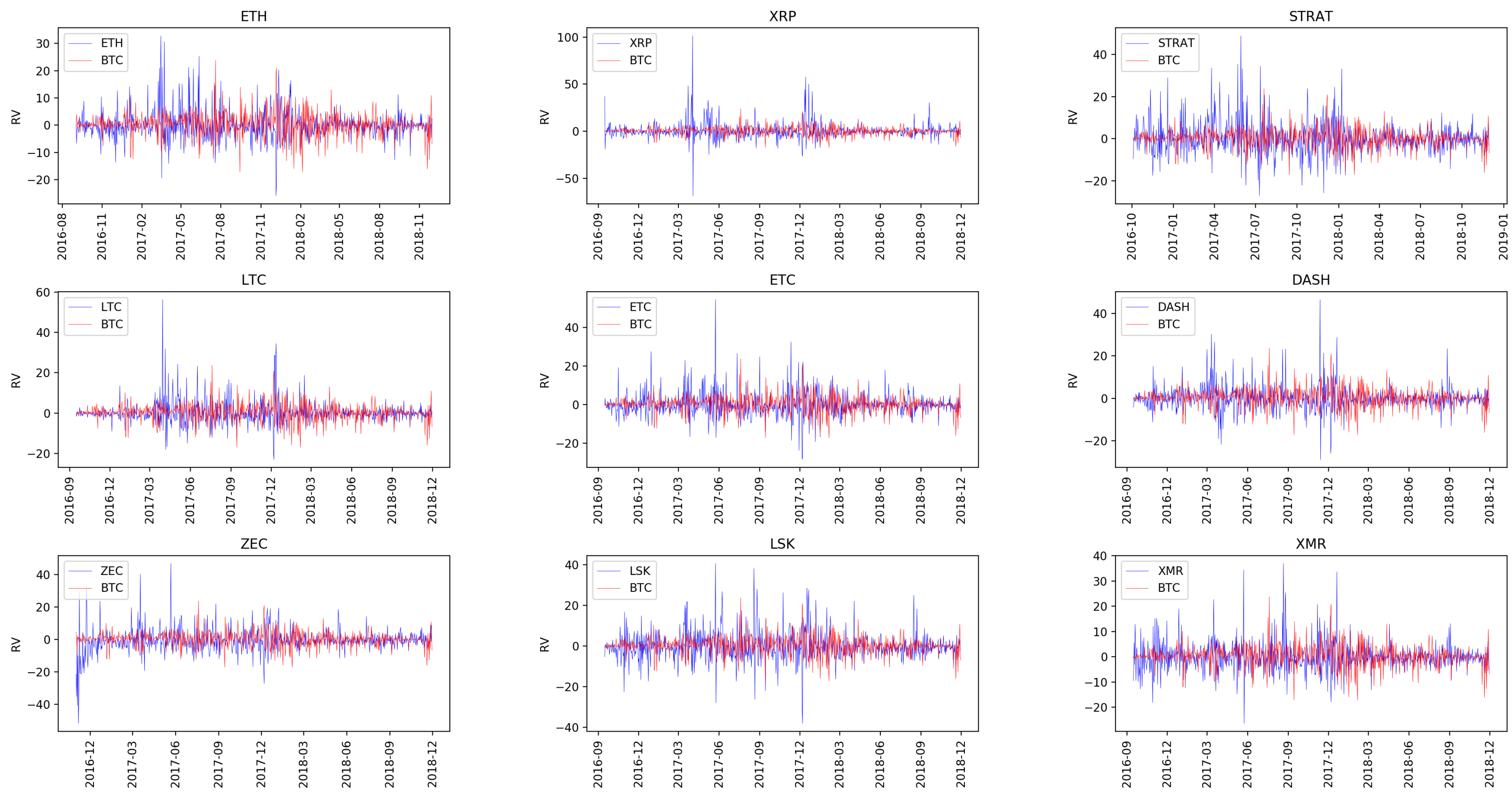


Figure C.2: Cryptocurrency Return

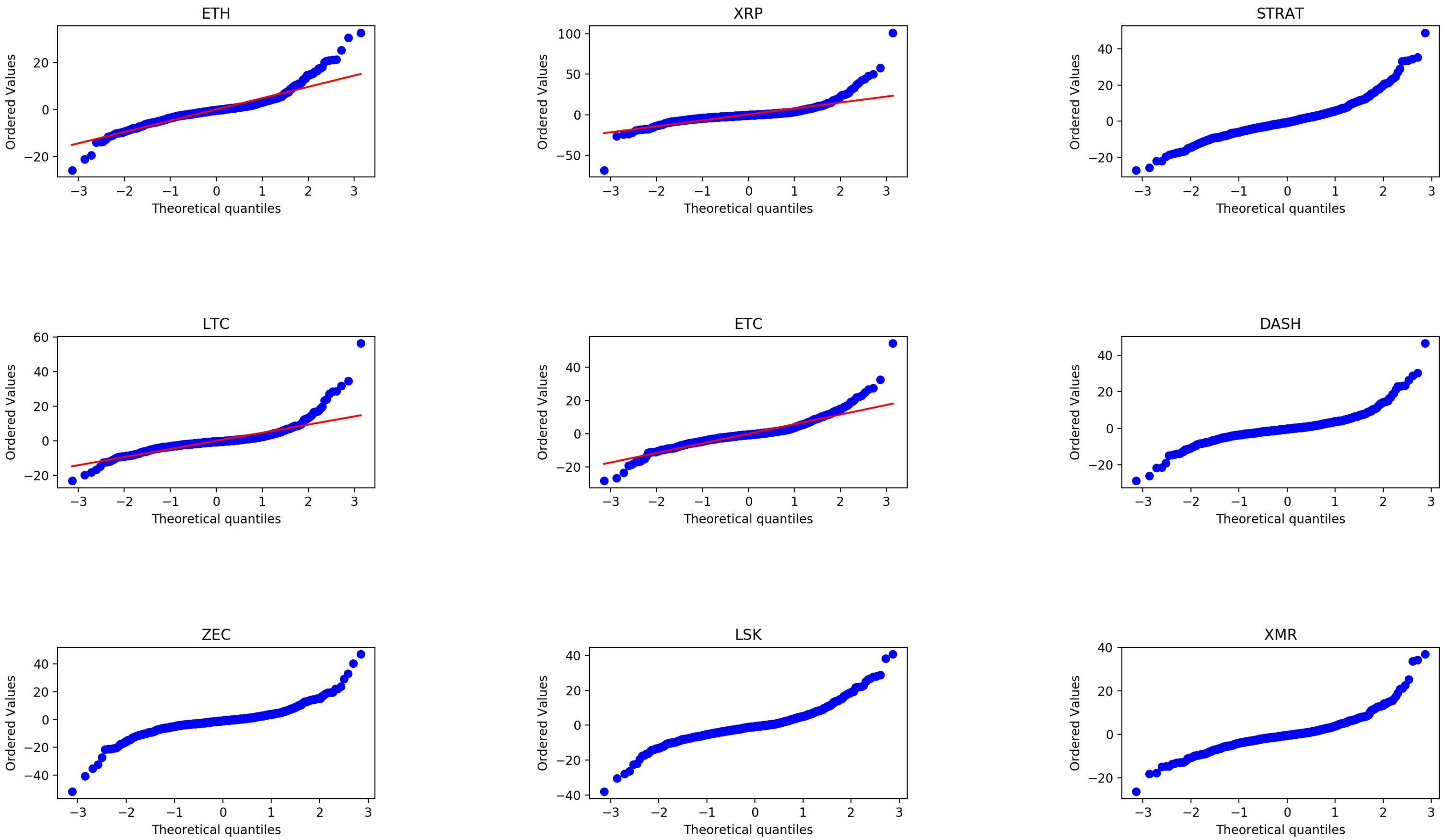


FIGURE C.3: Cryptocurrency Return QQ Plots

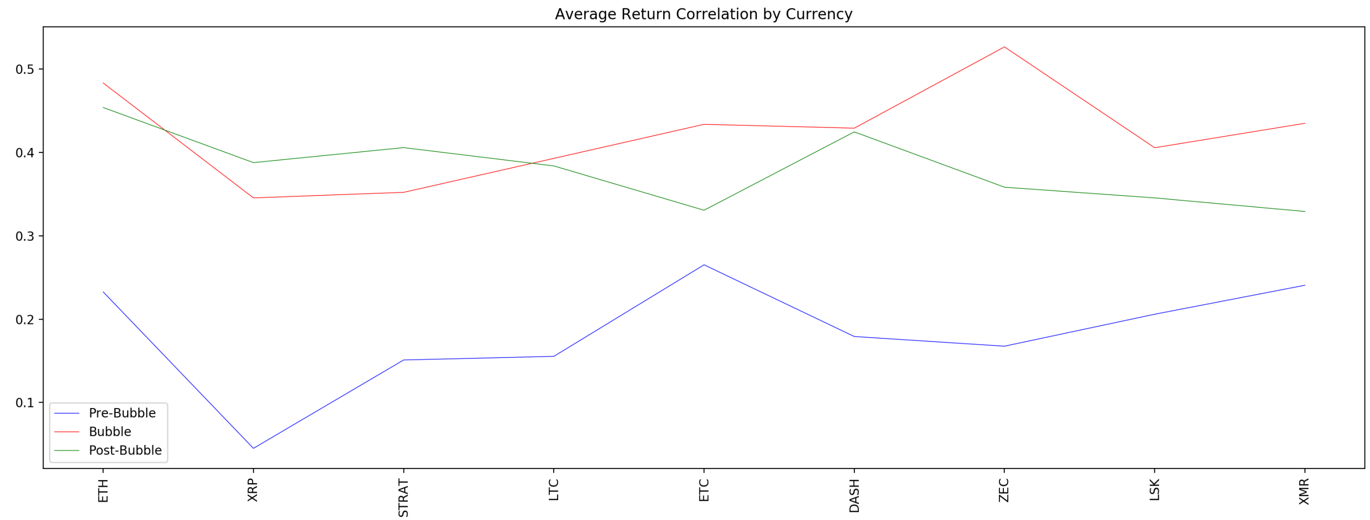


FIGURE C.4: Average Cryptocurrency Return Correlation in Pre-Bubble, Bubble and Post-Bubble

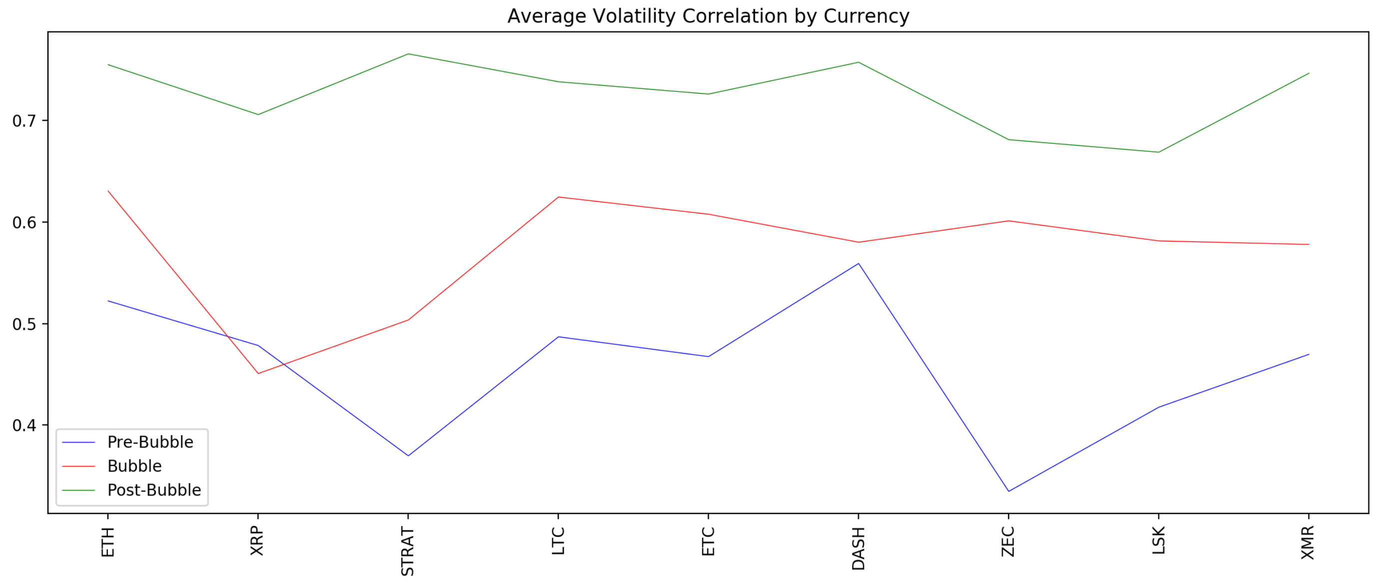


Figure C.5: Average Cryptocurrency Volatility Correlation in Pre-Bubble, Bubble and Post-Bubble

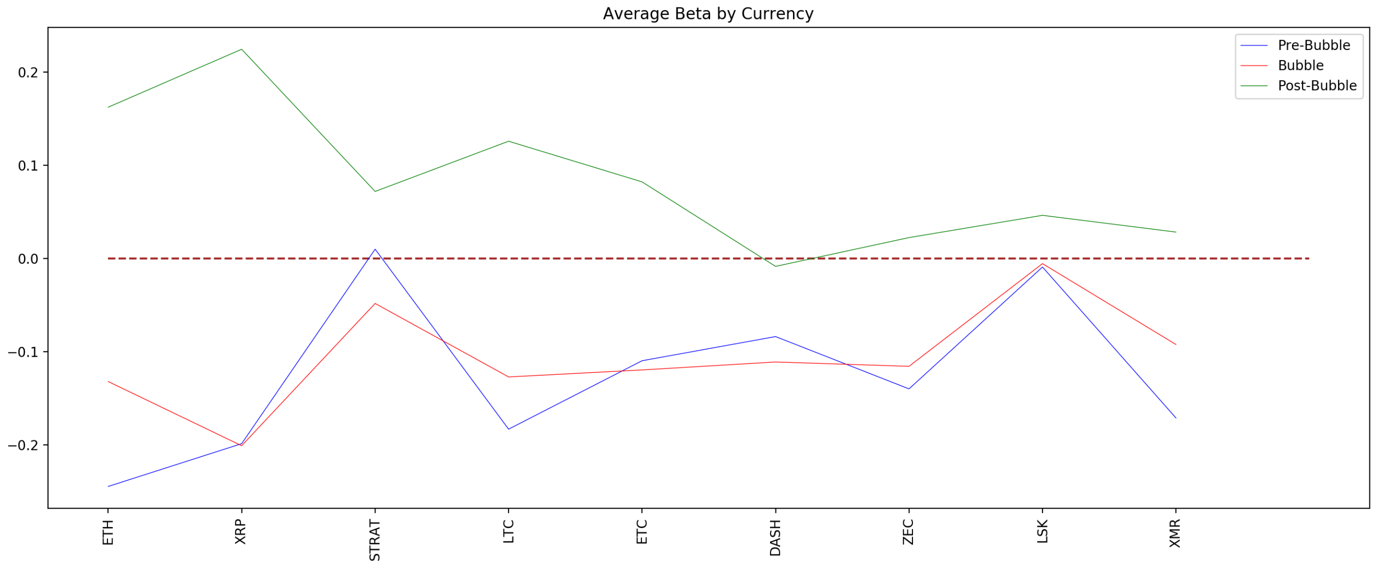


FIGURE C.6: Average Realized Betas in Pre-Bubble, Bubble and Post-Bubble

Bibliography

Abadi, J. and Brunnermeier, M. (2018). Blockchain economics. Technical report, National Bureau of Economic Research.

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

Aboody, D., Even-Tov, O., Lehavy, R., and Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, 53(2):485–505.

Abraham, A. and Ikenberry, D. L. (1994). The individual investor and the weekend effect. *Journal of Financial and Quantitative Analysis*, 29(2):263–277.

Ahn, M., Bonsall, S. B., and Van Buskirk, A. (2019). Do managers withhold bad news from credit rating agencies? *Review of Accounting Studies*, 24(3):972–1021.

Ait-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial studies*, 18(2):351–416.

Akhtar, S., Faff, R., Oliver, B., and Subrahmanyam, A. (2011). The power of bad: The negativity bias in australian consumer sentiment announcements on stock returns. *Journal of Banking & Finance*, 35(5):1239–1249.

Alti, A. and Tetlock, P. C. (2014). Biased beliefs, asset prices, and investment: A structural approach. *The Journal of Finance*, 69(1):325–361.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Amir, E., Levi, S., and Livne, T. (2018). Do firms underreport information on cyber-attacks? evidence from capital markets. *Review of Accounting Studies*, 23(3):1177–1206.

An, Z., Chen, C., Naiker, V., and Wang, J. (2020). Does media coverage deter firms from withholding bad news? evidence from stock price crash risk. *Journal of Corporate Finance*, 64:101664.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Anderson, S. P. and McLaren, J. (2012). Media mergers and media bias with rational consumers. *Journal of the European Economic Association*, 10(4):831–859.

Andrei, D., Friedman, H. L., and Ozel, N. B. (2019). Economic uncertainty and investor attention. *Available at SSRN 3128673*.

Andrei, D., Friedman, H. L., and Ozel, N. B. (2020). Economic uncertainty and investor attention. *Available at SSRN 3128673*.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294.

Arango-Kure, M., Garz, M., and Rott, A. (2014). Bad news sells: The demand for news magazines and the tone of their covers. *Journal of Media Economics*, 27(4):199–214.

Areni, C. S. and Burger, M. (2008). Memories of “bad” days are more biased than memories of “good” days: past saturdays vary, but past mondays are always blue. *Journal of Applied Social Psychology*, 38(6):1395–1415.

Atiase, R. K. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, 23(1):21–36.

Averill, J. R. (1980). A constructivist view of emotion. In *Theories of Emotion*, pages 305–339. Elsevier.

Back, K. (2010). *Asset pricing and portfolio choice theory.* Oxford University Press.

Baginski, S. P., Campbell, J. L., Hinson, L. A., and Koo, D. S. (2018). Do career concerns affect the delay of bad news disclosure? *The Accounting Review*, 93(2):61–95.

Bajo, E. and Raimondo, C. (2017). Media sentiment and ipo underpricing. *Journal of Corporate Finance*, 46:139–153.

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

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152.

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

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Bakshi, G., Gao, X., and Rossi, A. G. (2019). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, 65(2):619–641.

Bali, T. G., Bodnaruk, A., Scherbina, A., and Tang, Y. (2018). Unusual news flow and the cross section of stock returns. *Management Science*, 64(9):4137–4155.

Bali, T. G., Brown, S. J., and Tang, Y. (2017). Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics*, 126(3):471–489.

Bali, T. G., Engle, R. F., and Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns.* John Wiley & Sons.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Bao, D., Kim, Y., Mian, G. M., and Su, L. (2019). Do managers disclose or withhold bad news? evidence from short interest. *The Accounting Review*, 94(3):1–26.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Barberis, N. and Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1:1053–1128.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1-2):1–36.

Barry, C. B. and Brown, S. J. (1984). Differential information and the small firm effect. *Journal of Financial Economics*, 13(2):283–294.

Bassi, A., Colacito, R., and Fulghieri, P. (2013). 'o sole mio: An experimental analysis of weather and risk attitudes in financial decisions. *The Review of Financial Studies*, 26(7):1824–1852.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4):323–370.

Behrendt, S. and Schmidt, A. (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. *Journal of Banking and Finance*, 96:355–367.

Benamar, H., Foucault, T., and Vega, C. (2019). Demand for information, uncertainty, and the response of us treasury securities to news. *The Review of Financial Studies.*

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, 24(4):586–607.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Bernhardt, D., Krasa, S., and Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5-6):1092–1104.

Bhattacharya, U., Galpin, N., Ray, R., and Yu, X. (2009). The role of the media in the internet ipo bubble. *Journal of Financial and Quantitative Analysis*, 44(3):657–682.

Biais, B., Bisiere, C., Bouvard, M., and Casamatta, C. (2019). The blockchain folk theorem. *The Review of Financial Studies*, 32(5):1662–1715.

Bianchi, D. (2020). Cryptocurrencies as an asset class? an empirical assessment. *The Journal of Alternative Investments*.

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

Birru, J. (2018). Day of the week and the cross-section of returns. *Journal of Financial Economics*, 130(1):182–214.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Bonsall IV, S. B., Green, J., and Muller III, K. A. (2020). Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics*, 69(1):101264.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic expectations and stock returns. *The Journal of Finance*, 74(6):2839–2874.

Borri, N. (2019). Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, 50:1–19.

Borri, N. and Shakhnov, K. (2018). Cryptomarket discounts. *Available at SSRN 3124394*.

Bouchaud, J.-P., Krueger, P., Landier, A., and Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance*, 74(2):639–674.

Brown, G. W. and Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1):1–27.

Bushee, B. J. and Friedman, H. L. (2016). Disclosure standards and the sensitivity of returns to mood. *The Review of Financial Studies*, 29(3):787–822.

Busse, J. A. and Green, T. C. (2002). Market efficiency in real time. *Journal of Financial Economics*, 65(3):415–437.

Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University press.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4):905–939.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

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

Chan, L. K., Jegadeesh, N., and Lakonishok, J. (1996). Momentum strategies. *The Journal of Finance*, 51(5):1681–1713.

Chang, S.-C., Chen, S.-S., Chou, R. K., and Lin, Y.-H. (2008). Weather and intraday patterns in stock returns and trading activity. *Journal of Banking & Finance*, 32(9):1754–1766.

Chen, H., De, P., Hu, Y. J., and 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, W., Hribar, P., and Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, 56(3):751–796.

Chen, Z. and Petkova, R. (2012). Does idiosyncratic volatility proxy for risk exposure? *The Review of Financial Studies*, 25(9):2745–2787.

Chernenko, S., Hanson, S. G., and Sunderam, A. (2016). Who neglects risk? investor experience and the credit boom. *Journal of Financial Economics*, 122(2):248–269.

Chiu, J. and Koeppl, T. V. (2017). The economics of cryptocurrencies–bitcoin and beyond. *Available at SSRN 3048124.*

Choi, J. J. and Mertens, T. M. (2019). Extrapolative expectations and the equity premium. *Available at SSRN 3462056.*

Christoffersen, P., Lunde, A., and Olesen, K. V. (2019). Factor structure in commodity futures return and volatility. *Journal of Financial and Quantitative Analysis*, 54(3):1083–1115.

Chuen, D. L. K., Guo, L., and Wang, Y. (2017). Cryptocurrency: A new investment opportunity? *The Journal of Alternative Investments*, 20(3):16–40.

Chung, S.-L., Hung, C.-H., and Yeh, C.-Y. (2012). When does investor sentiment predict stock returns? *Journal of Empirical Finance*, 19(2):217–240.

Clore, G. and Schwartz, M. (1988). How do i feel about it? the informative function of affective states. *Affect Cognition and Social Behaviour, Toronto, CJ Hogrefe.*

Cody, E. M., Reagan, A. J., Dodds, P. S., and Danforth, C. M. (2016). Public opinion polling with twitter. *arXiv preprint arXiv:1608.02024.*

Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1):116–159.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., He, Z., and Li, J. (2019). Decentralized mining in centralized pools. *The Review of Financial Studies.*

Conway, M. and Giannopoulos, C. (1993). Dysphoria and decision making: Limited information use for evaluations of multiattribute targets. *Journal of Personality and Social Psychology*, 64(4):613.

Corbet, S., Larkin, C. J., Lucey, B. M., Meegan, A., and Yarovaya, L. (2018). The volatility generating effects of macroeconomic news on cryptocurrency returns. *Available at SSRN 3141986.*

Croft, G. P. and Walker, A. E. (2001). Are the monday blues ail in the mind? the role of expectancy in the subjective experience of mood 1. *Journal of Applied Social Psychology*, 31(6):1133–1145.

Da, R. and Xiu, D. (2019). When moving-average models meet high-frequency data: Uniform inference on volatility. *Chicago Booth Research Paper*, (17-27).

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

Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B., and Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12(4):413–438.

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

Daniel, K. and Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643.

De Bondt, W. F. and Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, 80(2):52–57.

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

Dey, S. (2014). Stock market prediction using twitter mood. *International Journal of Scientific & Engineering Research*, 5(5):44–47.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *the Journal of Finance*, 53(3):1131–1147.

Dobson, D. J. and Dobson, K. S. (1981). Problem-solving strategies in depressed and nondepressed college students. *Cognitive Therapy and Research*, 5(3):237–249.

Dodds, P. S. and Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752.

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. (2012). Journalists and the stock market. *The Review of Financial Studies*, 25(3):639–679.

Drobetz, W., Momtaz, P. P., and Schröder, H. (2019). Investor sentiment and initial coin offerings. *The Journal of Alternative Investments*, 21(4):41–55.

Duarte, J., Hu, E., and Young, L. (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics*, 135(3):795–815.

Duarte, J., Kamara, A., Siegel, S., and Sun, C. (2014). The systematic risk of idiosyncratic volatility. *Available at SSRN 1905731*.

Dyck, A. and Zingales, L. (2003). The media and asset prices. Technical report, Working Paper, Harvard Business School.

Easley, D., Hvidkjaer, S., and O'hara, M. (2002). Is information risk a determinant of asset returns? *The Journal of Finance*, 57(5):2185–2221.

Easley, D., Hvidkjaer, S., and O'hara, M. (2010). Factoring information into returns. *Journal of Financial and Quantitative Analysis*, 45(2):293–309.

Easley, D., Kiefer, N. M., O'hara, M., and Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finance*, 51(4):1405–1436.

Easley, D. and O'hara, M. (2004). Information and the cost of capital. *The Journal of Finance*, 59(4):1553–1583.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66(3):183.

Easterwood, J. C. and Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance*, 54(5):1777–1797.

Edmans, A., Fernandez-Perez, A., Garel, A., and Indriawan, I. (2021). Music sentiment and stock returns around the world.

Edmans, A., Garcia, D., and Norli, Ø. (2007). Sports sentiment and stock returns. *The Journal of Finance*, 62(4):1967–1998.

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

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

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

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

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

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

Freeman, R. N. (1987). The association between accounting earnings and security returns for large and small firms. *Journal of Accounting and Economics*, 9(2):195–228.

Frijda, N. H. et al. (1986). *The Emotions*. Cambridge University Press.

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

García, D., Sangiorgi, F., and Urošević, B. (2007). Overconfidence and market efficiency with heterogeneous agents. *Economic Theory*, 30(2):313–336.

Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2015). Media bias in the marketplace: Theory. In *Handbook of Media Economics*, volume 1, pages 623–645. Elsevier.

Glushkov, D. (2006). Sentiment beta. *Available at SSRN 862444*.

Goetzmann, W. N., Kim, D., Kumar, A., and Wang, Q. (2015). Weather-induced mood, institutional investors, and stock returns. *The Review of Financial Studies*, 28(1):73–111.

Goetzmann, W. N. and Zhu, N. (2005). Rain or shine: where is the weather effect? *European Financial Management*, 11(5):559–578.

Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746.

Griffin, J. M. and Lemmon, M. L. (2002). Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, 57(5):2317–2336.

Griffin, P. A. (1977). The time-series behavior of quarterly earnings: preliminary evidence. *Journal of Accounting Research*, 15(1):71–83.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.

Gu, C. and Kurov, A. (2020). Informational role of social media: Evidence from twitter sentiment. *Journal of Banking and Finance, Forthcoming.*

Guo, Hung, and Kontonikas (2020). The fed and the stock market: A tale of sentiment states. *Available at SSRN 3184974.*

Gurun, U. G. and Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance*, 67(2):561–598.

Hansen, P. R. and Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554.

Hasher, L. and Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108(3):356.

Heller, Y. and Winter, E. (2020). Biased-belief equilibrium. *American Economic Journal: Microeconomics*, 12(2):1–40.

Hendershott, T., Livdan, D., and Schürhoff, N. (2015). Are institutions informed about news? *Journal of Financial Economics*, 117(2):249–287.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283.

Heston, S. L. and Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445.

Hillert, A., Jacobs, H., and Müller, S. (2014). Media makes momentum. *The Review of Financial Studies*, 27(12):3467–3501.

Hillert, A., Jacobs, H., and Müller, S. (2018). Journalist disagreement. *Journal of Financial Markets*, 41:57–76.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, LXXV(4):1779–1832.

Hirshleifer, D., Jiang, D., and DiGiovanni, Y. M. (2020). Mood beta and seasonalities in stock returns. *Journal of Financial Economics*, 137(1):272–295.

Hirshleifer, D., Li, J., and Yu, J. (2015). Asset pricing in production economies with extrapolative expectations. *Journal of Monetary Economics*, 76:87–106.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032.

Hirshleifer, D. A., Myers, J. N., Myers, L. A., and Teoh, S. H. (2008). Do individual investors cause post-earnings announcement drift? direct evidence from personal trades. *The Accounting Review*, 83(6):1521–1550.

Holden, C. W. and Subrahmanyam, A. (2002). News events, information acquisition, and serial correlation. *The Journal of Business*, 75(1):1–32.

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

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Hu, A. S., Parlour, C. A., and Rajan, U. (2019). Cryptocurrencies: Stylized facts on a new investible instrument. *Financial Management*, 48(4):1049–1068.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, 28(3):791–837.

Huberman, G., Leshno, J., and Moallemi, C. C. (2019). An economic analysis of the bitcoin payment system. *Columbia Business School Research Paper*, (17-92).

Ifcher, J. and Zarghamee, H. (2011). Happiness and time preference: The effect of positive affect in a random-assignment experiment. *American Economic Review*, 101(7):3109–29.

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

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898.

Jiang, D., Norris, D., and Sun, L. (2019a). Weather, institutional investors, and earnings news. *Available at SSRN 3142635.*

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019b). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1):126–149.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2000). Losing sleep at the market: The daylight saving anomaly. *American Economic Review*, 90(4):1005–1011.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2003). Winter blues: A sad stock market cycle. *American Economic Review*, 93(1):324–343.

Kaplanski, G., Levy, H., Veld, C., and Veld-Merkoulova, Y. (2015). Do happy people make optimistic investors? *Journal of Financial and Quantitative Analysis*, 50(1/2):145–168.

Karampatsas, N., Malekpour, S., and Mason, A. (2018). Beyond market mood: Firm sentiment and the response to corporate earnings announcements. *Available at SSRN 2926368*.

Kaustia, M. and Rantapuska, E. (2016). Does mood affect trading behavior? *Journal of Financial Markets*, 29:1–26.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting returns with text data. Technical report, National Bureau of Economic Research.

Keller, A. S., Leikauf, J. E., Holt-Gosselin, B., Staveland, B. R., and Williams, L. M. (2019). Paying attention to attention in depression. *Translational psychiatry*, 9(1):1–12.

Kelly, B. and Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, 25(5):1366–1413.

Keloharju, M., Linnainmaa, J. T., and Nyberg, P. (2016). Return seasonalities. *The Journal of Finance*, 71(4):1557–1590.

Kliger, D. and Levy, O. (2003). Mood-induced variation in risk preferences. *Journal of Economic Behavior & Organization*, 52(4):573–584.

Ko, K. J. and Huang, Z. J. (2007). Arrogance can be a virtue: Overconfidence, information acquisition, and market efficiency. *Journal of Financial Economics*, 84(2):529–560.

Kothari, S. P., Shu, S., and Wysocki, P. D. (2009). Do managers withhold bad news? *Journal of Accounting Research*, 47(1):241–276.

Kramer, L. A. and Weber, J. M. (2012). This is your portfolio on winter: Seasonal affective disorder and risk aversion in financial decision making. *Social Psychological and Personality Science*, 3(2):193–199.

Kumar, A. and Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5):2451–2486.

Le Mens, G. and Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review*, 118(2):379.

Lee, C. M., Shleifer, A., and Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 46(1):75–109.

Lee, C. M. and Swaminathan, B. (2000). Price momentum and trading volume. *The Journal of Finance*, 55(5):2017–2069.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28.

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

Liu, Y. and Tsyvinski, A. (2018). Risks and returns of cryptocurrency. Technical report, National Bureau of Economic Research.

Livnat, J. and Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, 44(1):177–205.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of Financial Studies*, 15(4):1005–1047.

Lo, A. W., Repin, D. V., and Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, 95(2):352–359.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

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

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319.

Meeten, F. and Davey, G. C. (2011). Mood-as-input hypothesis and perseverative psychopathologies. *Clinical Psychology Review*, 31(8):1259–1275.

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

Mian, G. M. and Sankaraguruswamy, S. (2012). Investor sentiment and stock market response to earnings news. *The Accounting Review*, 87(4):1357–1384.

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053.

Müller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M., and Morgenegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance*, 14(6):1189–1208.

Nagel, S. (2005). Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics*, 78(2):277–309.

Nakamoto, S. (2019). Bitcoin: A peer-to-peer electronic cash system. Technical report, Manubot.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance*, 53(6):1887–1934.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance*, 58(4):1335–1354.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Park, M., Cha, C., and Cha, M. (2012). Depressive moods of users portrayed in twitter.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Patton, A. J. and Verardo, M. (2012). Does beta move with news? firm-specific information flows and learning about profitability. *The Review of Financial Studies*, 25(9):2789–2839.

Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance*, 69(5):2007–2043.

Peterson, R. L. (2016). *Trading on sentiment: The power of minds over markets.* John Wiley & Sons.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.

Phillips, P. C. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491.

Quercia, D., Ellis, J., Capra, L., and Crowcroft, J. (2011). In the mood for being influential on twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 307–314. IEEE.

Raghunathan, R. and Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes*, 79(1):56–77.

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., and Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1):1–12.

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., and Langer, E. J. (2017). Forecasting the onset and course of mental illness with twitter data. *Scientific Reports*, 7(1):1–11.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4):296–320.

Ryan, R. M., Bernstein, J. H., and Brown, K. W. (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. *Journal of Social and Clinical Psychology*, 29(1):95–122.

Saunders, E. M. (1993). Stock prices and wall street weather. *The American Economic Review*, 83(5):1337–1345.

Schilling, L. and Uhlig, H. (2019). Some simple bitcoin economics. *Journal of Monetary Economics*, 106:16–26.

Schwarz, N. and Clore, G. L. (2007). Feelings and phenomenal experiences. *Social psychology: Handbook of basic principles*, pages 385–407.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790.

Shen, D., Liu, L., and Zhang, Y. (2018). Quantifying the cross-sectional relationship between online sentiment and the skewness of stock returns. *Physica A: Statistical Mechanics and its Applications*, 490:928–934.

Shu, H.-C. (2010). Investor mood and financial markets. *Journal of Economic Behavior & Organization*, 76(2):267–282.

Silberman, E. K., Weingartner, H., and Post, R. M. (1983). Thinking disorder in depression: Logic and strategy in an abstract reasoning task. *Archives of General Psychiatry*, 40(7):775–780.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sletten, E. (2012). The effect of stock price on discretionary disclosure. *Review of Accounting Studies*, 17(1):96–133.

Smith, C. A. and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4):813.

Sockin, M. and Xiong, W. (2020). A model of cryptocurrencies. Technical report, National Bureau of Economic Research.

Spiess, D. K. and Affleck-Graves, J. (1995). Underperformance in long-run stock returns following seasoned equity offerings. *Journal of Financial Economics*, 38(3):243–267.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Stone, A. A., Schneider, S., and Harter, J. K. (2012). Day-of-week mood patterns in the united states: On the existence of 'blue monday','thank god it's friday'and weekend effects. *The Journal of Positive Psychology*, 7(4):306–314.

Storbeck, J. and Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass*, 2(5):1824–1843.

Swaminathan, B. (1996). Time-varying expected small firm returns and closed-end fund discounts. *The Review of Financial Studies*, 9(3):845–887.

Szymanowska, M., De Roon, F., Nijman, T., and Van Den Goorbergh, R. (2014). An anatomy of commodity futures risk premia. *The Journal of Finance*, 69(1):453–482.

Tesser, A. and Rosen, S. (1975). The reluctance to transmit bad news. *Advances in Experimental Social Psychology*, 8:193–232.

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

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *The Review of Financial Studies*, 23(9):3520–3557.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, 24(5):1481–1512.

Tetlock, P. C. (2014a). Information transmission in finance. *Annual Review Financial Economics*, 6(1):365–384.

Tetlock, P. C. (2014b). Information transmission in finance. *Annual Review of Financial Economics*, 6(1):365–384.

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

Tirole, J. (2002). Rational irrationality: Some economics of self-management. *European Economic Review*, 46(4-5):633–655.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Van Nieuwerburgh, S. and Veldkamp, L. (2009). Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3):1187–1215.

Veldkamp, L. L. (2006). Media frenzies in markets for financial information. *American Economic Review*, 96(3):577–601.

Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.

Weber, W. E. (2016). A bitcoin standard: Lessons from the gold standard. Technical report, Bank of Canada Staff Working Paper.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies*, 31(6):2184–2226.

Wojcik, S. and Hughes, A. (2019). Sizing up twitter users. *PEW Research Center*, 24.

Wyer Jr, R. S. and Srull, T. K. (2014). *Handbook of Social Cognition: Volume 1: Basic Processes*. Psychology Press.

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

Yermack, D. (2015). Is bitcoin a real currency? an economic appraisal. In *Handbook of Digital Currency*, pages 31–43. Elsevier.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

Zhang, W., Li, X., Shen, D., and Teglio, A. (2016). Daily happiness and stock returns: Some international evidence. *Physica A: Statistical Mechanics and its Applications*, 460:201–209.

Zhao, R. (2020). Quantifying the cross sectional relation of daily happiness sentiment and stock return: Evidence from us. *Physica A: Statistical Mechanics and its Applications*, 538:122629.