The Journal of Business Inquiry 2017, 17, Issue 1 (Special Issue), 41-54
http:www.uvu.edu/woodbury/jbi/articles/
ISSN 2155-4072

Social Data Predictive Power Comparison Across Information Channels and User Groups: Evidence from the Bitcoin Market

By PENG XIE, JIMING WU, AND CHONGQI WU*

In the context of Bitcoin, we examine the relationship between Bitcoin price movement and social data sentiment. Baseline findings reveal that social media provides value-relevant information in both short-term and long-term predictions. By comparing the predictive power across different information channels and different user groups, we found that (1) while speculative information predicts both long-term and short-term returns effectively, fundamental-related information only predicts long-term returns, and that (2) prediction accuracy is higher for less active users than for active users on social media, especially in long-term prediction.

Keywords: Social Media, Digital Currency, User-Generated Content, Text Mining

JEL Classification: M15, G17

I. Introduction

Early during the last decade, people started to realize that the Internet was playing an increasingly important role in the financial markets (Tumarkin and Whitelaw, 2001). Besides traditional information sources such as earnings releases, financial analyst recommendations, and news services, technology advancement makes other means of information sources available. Today, social media has become an important outlet of value-relevant information and a new way to assist investment decisions.

Many practitioners embraced this method and achieved huge success. For example, Kensho, a large-scale data processing platform similar to Google search, focusing on answering real-time investment related queries, posets threats to financial analyst professionals. Datasift, a US-based company offering a powerful cloud platform to extract value from social media and make predictions, is currently worth more than a billion dollars. Cayman Atlantic, a hedge fund that invests based on sentiment analysis of Twitter and other media, achieved a cumulative annual return of 25.10% during 2014[1] and 10.42% during 2015[2]. Many other companies from different industries such as Goldman Sachs, Thomson Reuter's Eikon, IBM, and Bloomberg have also started to offer services based on social media sentiment analysis.

This phenomenon is landscape-shifting in the finance industry and has attracted attention from researchers. There is already abundant literature on the impact of traditional news media on stock prices (Davis *et al.*, 2012; Loughran and McDonald, 2011; Tetlock, 2007; Tetlock *et al*,.

* Peng Xie, corresponding author, California State University, East Bay, Department of Management. Phone: (404) 369-6131. Email: peng.xie@scheller.gatech.edu. Jiming Wu, California State University, East Bay, Department of Management, Phone: (510) 885-3099. Email: jiming.wu@csueastbay.edu. Chongqi Wu, California State University, East Bay, Department of Management. Phone: (510) 885-3568. Email: chongqi.wu@csueastbay.edu.

[1] For detailed information about the cumulative returns, please refer to Cayman Atlantic (2014).

[2] For detailed information about the cumulative returns, please refer to Cayman Atlantic (2015).

2008), and researchers are catching up with this trend to study the informativeness of social media for the financial markets (Antweiler and Frank, 2004; Chen *et al.*, 2014; Das and Chen, 2007; Tumarkin and Whitelaw, 2001). Several major online communities have been investigated, such as the Yahoo! Finance message board, RagingBulls and Seeking Alpha, etc.

In this research, we follow this line of investigation to answer two related questions: (1) Does the social media wear-in time differ across social media information channels? and (2) Is the prediction accuracy related to social media users' level of activity. To preview our results, we found that speculation-related information predicts both long-term and short-term returns, while fundamental-related information only predicts long-term returns. By comparing different user groups, we found that more accurate information comes from inactive users rather than active users.

Our research context is an emerging digital currency, usually known as cryptocurrency. It is a decentralized peer-to-peer electronic payment network. Though our research background is limited to the Bitcoin market, the insights can be readily generalized to other markets satisfying the following two conditions: (1) There must be a market to enable free trading of the underlying assets, and (2) There must be a social media to enable communications between peer investors. In recent years, social media have become an important unofficial information outlet due to their rapid development. With this trend, we believe that our paper will potentially shed light on trading behavior in an increasing number of domains in the years to come.

The rest of the paper is organized as follows. Section II reviews the related literature and develops our hypotheses. Section III describes our data and empirical analyses. Section IV concludes.

II. Literature Review and Hypothesis Development

Why do loosely organized social media play a role in the financial markets where the trading involving millions of dollars is conducted every day? It has already been noticed earlier in the finance literature that stock market participation increases with social interaction due to the word-of-mouth effect or observational learning (Hong *et al.*, 2004). However, to predict market price movements with social media, we have to answer two key questions.

First, why are people willing to share quality information with others? There are several reasons. First, people derive utility from attention and recognition from posting quality information that is subsequently confirmed by price movements. Second, message board viewers' reading and trading can have a price impact and expedite the convergence of market prices to what the authors perceived to be fair. Therefore, informed actors have the incentive to publicize their investment ideas (Chen *et al.*, 2014). Third, it has been shown that people contribute their knowledge when they perceive an enhancement of their reputation, and they contribute without expectations of reciprocity from others or high commitment to the network (McLure-Wasko and Faraj, 2005). The latent benefit of a social exchange process can be emotional comforts or social rewards such as approval, status, and respect (Shi *et al.*, 2014).

Second, why do people trust information on social media, where there is no guarantee of the information's quality? Several mechanisms have been identified in the related literature. Social media user-generated content (UGC) can potentially affect stock prices in the following ways (Tumarkin and Whitelaw, 2001): (1) posting activities may help predict stock returns if the message contains new information; (2) even if messages do not contain new information, they may also provide a better indication of general market sentiment than is already contained in the trading records; (3) even without any value-relevant information, investors may follow the buy and sell recommendations of message board users; and (4) traders may recognize the momentum generated by investors who follow message boards, thus exaggerating this effect. The existence of persuasion bias in social media also serves as an explanation (DeMarzo *et al.,* 2001). People fail to adjust properly for possible repetition of the information they receive, so when an individual in a social network hears a piece of information over and over again from peers, he or she will be further influenced.

In the related literature, some researchers attempt to predict short-term stock returns (Antweiler and Frank, 2004; Das and Chen, 2007; Dougal *et al.,* 2012; Solomon, 2012; Tetlock, 2007; Tumarkin and Whitelaw, 2001; Wysocki, 1998), while others predict long-term stock returns (Chen *et al.,* 2014; Davis *et al.,* 2012; Dewally, 2003; Womack, 1996). However, when we use social media analytics to predict market price movements, it is crucial to know approximately how long it takes for the information to be factored into the price (the "wear-in" time). A recent related study compared the wear-in time and predictive value of different information outlets and found that social media have higher predictive value and shorter wear-in time (Luo *et al.,* 2013). Still little is known about whether or not the wear-in time depends on the type of the information. In this study, we are on a mission to answer this question. Specifically, we compare two types of information: fundamental-related information and speculative information.

Fundamental-related information unveils inherent value and predicts future trends. But for a volatile market such as the Bitcoin market, it is unlikely that fundamental-related information is value-relevant in the short term because there are many market surprises constantly affecting the short-term Bitcoin returns since the Bitcoin market is still in an early stage. Examples include unexpected technical advancements, shocks, and security concerns, among other issues. Under such circumstances, even if the long-term implications embedded in the social data are correct, people are reluctant to trust and take actions immediately due to unexpected shocks and hyper risks. And to make things worse, traditionally people place less emphasis on social media outlets compared to financial analysts (Chen *et al.,* 2014). So we expect fundamental-related information to have a very limited short-term impact. However, if the prediction of the long-term trend is accurate, no matter whether people trust it or not initially, the future price movement will ultimately confirm the original social media predictions. As a result, we expect that fundamental-related information predicts, if it can, only the long-term Bitcoin price changes.

Now take a look at the speculative information. Speculation is defined as a process for transferring price risks (Tirole, 1982). It is the practice of engaging in financial transactions in an attempt to profit from fluctuations in the market value of a tradable good, rather than trying to profit from the underlying financial attributes. There is no determinate result in the finance literature as to whether or not speculation occurs in the short or long term. Both cases exist. So we expect that speculative information affects both short-term and long-term price movements.

In light of these considerations, we propose our first hypothesis:

Hypothesis 1: Fundamental-related information only predicts long-term price movements, while speculative information predicts both long-term and short-term price movements.

Next, we compare the predictive power of different user groups on social media. In recent years, searching for efficient ways to locate influential social media participants and to take advantage of them in marketing and advertising has attracted attention from many practitioners and researchers. Social media users differ in their activity level and their informativeness.

Understanding who keeps the social network attractive and who influences the activity of others is vital (Trusov *et al.,* 2009). In the related literature, influential people are believed to have three attributes: (1) they are convincing, (2) they know a lot (i.e., are experts), and (3) they have a large number of social ties (Goldenberg *et al.,* 2009). Most researchers focused on the third point. They are interested in influential people who can create buzz. However, very little attention is directed to those people who actually have accurate insights. Those people do not necessarily overlap with those who have a lot of social ties or those who are active on social media. In this paper, we compare the prediction accuracy of active users with high levels of activity to the prediction accuracy of inactive users with less presence on social media.

According to the literature review at the beginning of Section II, the motivations to share on social media are multifold, including latent emotional benefits (enjoyment of helping, reciprocal relationships), reputation enhancement, and expedited price convergence. Active social media users who regularly engage in communications are well rewarded with emotional benefits. However, for those who are very inactive, emotional benefits are not the main purpose; therefore, they must be motivated by other incentives such as reputation enhancement (McLure-Wasko and Faraj, 2005), and expedited price convergence. Intuitively, inactive social media users usually do not talk online for the sake of talk. If they share information with others, most probably they want to make a point.

In light of these considerations, we propose our second hypothesis:

Hypothesis 2: Inactive social media participants provide more insightful information than active social media participants do.

Almost all related studies are conducted using stock market data. We try to summarize the expected differences in the two settings. First, we expect better identification from the Bitcoin market. In the stock market, there are many confounding effects outside the social media such as periodic financial statements, firm announcements, and opinions from professional financial analysts. Some of the influences are difficult to control properly. Without the above-mentioned confounding factors, the Bitcoin market offers a much cleaner research background. Although there are editorial media outlets from major new services such as *The Wall Street Journal* news wires, we are able to control them in our paper using textual analysis. Second, we expect stronger effects from social media in the Bitcoin market because investors have to rely heavily on social media to obtain new information about Bitcoin in the absence of adequate official information sources. This prediction is supported by the comparison between coefficients estimates of our paper and those of a comparable stock market paper mentioned in the result section.

III. Data and Methods

A. Bitcoin and Bitcoin Return

Bitcoin, a type of digital currency (also known as cryptographic currency) launched early in 2009, has been increasingly recognized in recent years. The Bitcoin market capitalization shots up to over 10 billion US$ during 2016[3]. Though at a first look this technology resembles the credit card payment system, there are fundamental differences: (1) cryptocurrency platforms are running

[3] For detailed Bitcoin market capitalization data over time, please refer to the following link: https://blockchain.info/charts/market-cap.

on specialized currency, and the exchange rates with fiat currencies are decided at the exchanges, and (2) there is no central authority maintaining the operations, regulating the issuance of the currency, or keeping detailed records of every transaction.

Cryptocurrencies are also different from other types of virtual currencies (such as e-cash, DGC, prepaid card, etc.) in that its existence does not depend on any issuing institution, nor is it backed by precious metal. Its existence is based on cryptographic algorithms and a formula stipulating the growth of currency supplies outstanding. The motivations of building such a system are multifold, for example: (1) less dispute cost due to irreversible transactions; (2) no user identity theft and enhanced security; (3) global accessibility since transactions between payers and payees are not geographically limited; (4) money goes to payee's account almost instantly; (5) controlled inflation; (6) the amount of money transacted and the transaction frequencies are not limited by a third party, and (7) anonymity and untraceable transactions.

The cryptocurrency industry has had substantial impacts on both the global currency system and the electronic payment system. Since it is still in its infancy, research on this topic has just begun. Most researchers approach this topic from a technical aspect. Many such studies discuss issues in the Bitcoin mining process (Eyal and Sirer, 2014; Johnson *et al.*, 2014; Miller *et al.*, 2015; O'Dwyer and Malone, 2013), and some examine other technical issues such as anonymity in the Bitcoin system (Reid and Harrigan 2013). However, studies from the perspectives of economics and finance are limited. A recent study investigates whether users' interest in digital currency is based on its appeal as a currency or as an asset, and found that uninformed users adopt Bitcoin mainly as a speculation tool (Glaser *et al.*, 2014). An earlier 2013 study echoed the point of view that Bitcoin cannot be treated as currency due to its high volatility and hyper risks (Yermack, 2013). Though cryptocurrency has drawn some attention, there are many more issues to be addressed.

The baseline of this study is to predict Bitcoin price movements using related social data sentiments. We calculate the returns of Bitcoin using the exchange rate between Bitcoin and US$. The data period is from May 17, 2011 to October 28, 2014. To track the Bitcoin price movements, we collect Bitcoin price data from Bitstamp, a major "foreign exchange" between Bitcoin and many other fiat currencies. Similar to foreign exchange markets, the Bitcoin market is open 24 hours a day, and seven days a week. The Bitcoin prices used in the analyses are the 24:00 o'clock price on each day (the daily close price). All time stamps are based on GMT. The day t Bitcoin return is calculated as $(P_t - P_{t-1})/P_{t-1}$.

The Bitcoin market has been very volatile, especially during its earlier years. At the time of this study, the entire Bitcoin system is still immature: constant revolutions, disasters, and new government regulations frequently surprise the Bitcoin market. Over the entire data period, the highest daily return reached 41.38%, and most dramatic declines bottomed at -50.31%. But from the point view of model identification, this instability is advantageous since more variations are embedded in our data.

B. Social Media

We downloaded social media discussions from Bitcointalk.org, which is a very popular online message board about cryptocurrencies. There are many discussion sections on this website. However, most of them are either off-topic or only distantly related to Bitcoin valuations. Though there is a comprehensive discussion section called "Bitcoin Discussion," which contains Bitcoin general discussions, we cannot effectively separate information into different categories. Since one of our intentions in this paper is to compare the "wear-in" time between different types of information, we only collect social data from specialized discussion sections. In particular, we employed a python script to download message board discussions from three sections: Speculative (Speculation about the Bitcoin price), Economics (Bitcoin from economics point of view, inflation/deflation, exchanges, Bitcoin loans etc.), and Trading Discussion (discussions about doing business with Bitcoin, best trading practices, delivery methods etc.). Examples of discussion topics in the three sections are provided in Table 1 (the exact words from bitcointalk.org).

To measure the sentiment of the social media discussions, we follow the literature and use the percentage of negative words. The negative word list we use is constructed by Loughran and McDonald (Loughran and McDonald 2011), which is a word list modified from the Harvard Psychosociological Dictionary (2017) to fit into the financial contexts. The sentiment expressed by all discussions during a certain day is the average percentage of negative words for all postings and replies of that day (Chen *et al.* 2014). In this way, we calculate our three key sentiment variables $Speculation_t$, $Economics_t$, *and* $TradingDiscussion_t$ respectively.

Besides social media, traditional media are also important sources of information. To control their impact, we downloaded Bitcoin-related editorial news articles from FACTIVA. Specifically, we searched on FACTIVA with keyword "Bitcoin", and limited our attention to articles written in English and published in major newswires (*The Wall Street Journal*, Dow Jones news wire, and Reuters news wire). We ended up with 13,216 articles. The earliest article about Bitcoin on FACTIVA was published on May 17, 2011, the day on which our data collection starts.

Table 1：Discussion Topic Examples

Discussion Sections	Examples
Economics	Do you think Bitcoin will replace dollar soon? Will Bitcoin cause the end of public debt? Bitcoin or gold? What would you pick?
Speculation	Is this the next big run-up in price? 320$, what the hell is going on? Will BTC reach $350 during November?
Trading Discussions	Bitcoin arbitrage on Github: ~2% monthly return, market neutral long/short Selling Rate of BTC on Circle Higher Than Coinbase Best way for cashing in

C. Traditional Media Controls

To measure the information contents of the traditional media, we put together all articles published on the same day and applied textual analysis (calculating the percentage of negative words). Similarly, we used the negative word list constructed by Loughran. All time stamps in this paper are based on GMT. The descriptive statistics are shown in Table 2.

Table 2: Descriptive Statistics

	Economics	**Speculation**	**Trading Discussion**	**Traditional Media**
Total # Observations	1261	1261	1261	1261
Total # Articles	106,031	277,329	56,836	13,216
Avg. % Negative Words	1.52	1.30	1.16	1.61
StDev % Negative Words	0.62	0.58	0.72	0.77
Max % Negative Words	8.26	6.88	10.15	6.28
Min % Negative Words	0	0	0	0

D. Main Results

We organize our main analysis around the following baseline regression specification:

$$R_t=\alpha+\beta_1 Economics_t+\beta_2 Speculation_t+\beta_3 TradingDiscussion_t+X\delta+\varepsilon_t$$

This regression tests the baseline expectation in this paper. First, we examine the effects of social media discussions on the end-of-day price movements. Since Bitcoin is traded 24/7, the intraday return is calculated using the 0:00 price and 24:00 price of day *t*.

Our key independent variables are the average fractions of negative words in the three discussion sections: $Economics_t$, $Speculation_t$ and $TradingDiscussion_t$. If social media does help predict the end-of-day Bitcoin price movement, the coefficient estimates for the three sentiment measures should be negative. X includes our control variables: $TraditionalMedia_t$, $Volatility_t$, R_{t-1}, and R_{t-2}. Volatility_t is calculated as the sum of squared daily returns during the previous calendar month.

Table 3 presents the result for the end-of-day price movement predictions. In Column (3), the coefficient estimates of $Speculation_t$ and $TradingDiscussion_t$ are -1.292 and -1.651 respectively, implying that the end-of-day price will be 129% (165%) lower when there are 1% more negative words in the Speculation (Trading Discussion) sections during that day. However, there are no significant results for $Economics_t$.

Table 3: Predict End-of-Day Price Change with Social Media

	(1)	(2)	(3)	(4)
$Economics_t$	-0.271 (-0.77)	-0.218 (-0.62)	-0.249 (-0.71)	-0.531 (-0.93)
$Speculation_t$	-0.812** (-2.27)	-0.775** (-2.17)	-0.747** (-2.09)	-1.292** (-2.26)
$TradingDiscussion_t$	-0.950*** (-3.12)	-0.939*** (-3.09)	-0.951*** (-3.13)	-1.651*** (-3.83)
R_{t-1}		0.077*** (2.74)	0.075*** (2.67)	0.098*** (2.79)
R_{t-2}		-0.057** (-2.05)	-0.059** (-2.11)	-0.039 (-1.12)
$Volatility_t$			0.013 (1.28)	0.025* (1.95)
$TraditionalMedia_t$				0.341 (1.10)

***=P<0.01, **=P<0.0 5, *=P<0.1.

This response may appear to be unrealistically large at first, but since the average fraction of negative words is around 1%, a 1% increase is rather significant. Also, the Bitcoin market is characterized by huge price volatility. This method of payment is not yet widely accepted in transaction partly due to the volatility problem. The US$ equivalence of 1 Bitcoin was only $0.30 in January 2011 but this number skyrocketed to $1,300 during November 2013. The sharpest one-day drop occurred on April 11, 2013 when the price fell from over $260 to $77.56.

Also, due to limited information sources within the Bitcoin market and the absence of institutional investors, social media are a major information source, and have significantly amplified effects on the price. Considering all the factors above, we expected a significant difference in the scales of the results from similar studies in the stock market. The results in Chen *et al*'s (2014) paper show a 0.25% to 0.28% drop in returns when the fraction of negative words in Seeking Alpha articles increases by 1%, which is a much smaller impact.

Many related studies on the stock market are focused on the prediction of the next-day price movements and obtain significant results. We also tested the predictive power of social media for the next-day price change. The results are presented in Table 4. Notice that the coefficient estimates for variables $Speculation_t$ and $TradingDiscussion_t$ are no longer significant. Evidence shows that social data only predicts the end-of-day price movement within the context of Bitcoin, but not the next-day return. This is different from the cases in the stock market. The reason lies in the differences in the structure between the stock market and the Bitcoin market.

Like all other cryptocurrencies, Bitcoin is traded 24/7, therefore if the market is efficient, any information that is valuable in the short term will factor into the price by the end of that day. As a result, the Bitcoin returns calculated using 0:00 price and 24:00 price only reflect price changes within the current day but not the next day. However, in the stock market, there are market closures. Any relevant information released after the market closure will only possibly affect the next-day return (returns in the stock market are usually calculated using closing prices of two successive days).

Table 4: Predict Next-Day Return with Social Media

	(1)	(2)	(3)	(4)
$Economics_t$	0.079 (0.22)	0.069 (0.20)	0.037 (0.11)	-0.202 (-0.34)
$Speculation_t$	0.009 (0.02)	-0.040 (-0.11)	-0.012 (-0.03)	-0.272 (-0.46)
$TradingDiscussion_t$	-0.102 (-0.33)	-0.085 (-0.28)	-0.098 (-0.32)	-0.039 (-0.09)
R_{t-1}		-0.052* (-1.85)	-0.054* (-1.92)	-0.035 (-0.97)
R_{t-2}		-0.026 (-0.91)	-0.027 (-0.97)	-0.066 (-1.85)
$Volatility_t$			0.013 (1.28)	0.014 (1.09)
$TraditionalMedia_t$				0.403 (1.27)

***=P<0.01, **=P<0.0 5, *=P<0.1.

Following the baseline results, we test our Hypothesis 1 next. As mentioned in the introduction, some researchers predict short-term returns but others predict long-term returns, and lately, they have begun to investigate the "wear-in" time of different social media metrics (Tirunillai and Tellis, 2011) and the "wear-in" time of different information channels (Luo *et al.*, 2013). In this article, we follow this line of investigation to test if the wear-in time depends on the type of information.

Examples in Table 1 suggest that people are more interested in the inherent value, and the future trend of Bitcoin in Economics related topics, but are more concerned about the price change and predictions in Speculation and Trading related topics. We thus compare the predictive power of two different types of information: fundamental-related information and speculative information. Fundamental-related information is measured by *$Economics_t$*, and speculative information is measured by *$Speculation_t$* and *$TradingDiscussion_t$*. Our Hypothesis 1 posits that the fundamental-related information predicts long-term price changes, while in contrast, speculative information predicts both long-term and short-term price changes.

A similar model is used in this section except that the dependent variable is now the cumulative returns ($R_{t,t+a}$). We use the social media discussions observed at time t to predict the cumulative returns from t to $t+a$. We empirically examine the predictive power for one-week, one-month, and three-month cumulative returns respectively. The result is shown in Table 5.

The first column in Table 5 is the same as the previous result in Table 3. We include it here just for comparison. Column 1 of Table 5 shows that Economics related discussions do not predict short-term returns, and this is also true for one-week and one-month cumulative return predictions. However, in the last column in Table 5, the results demonstrate a strong predictive power for the three-month cumulative returns. The coefficient estimate jumps from below 1 to a very high value. For the speculative information, represented by *$Speculation_t$* and *$TradingDiscussion_t$*, we only detect predictive powers for short-term price movements, basically within one week (columns 1 and 2 of Table 5).

Table 5: Fundamental-Related Information vs. Speculative Information

	R_t	$R_{t,t+7}$	$R_{t,t+30}$	$R_{t,t+90}$
Economics$_t$	-0.531 (-0.93)	-0.932 (-0.58)	-0.242 (-0.04)	-50.636*** (-3.11)
Speculation$_t$	-1.292** (-2.26)	-4.096** (-2.56)	-2.535 (-0.41)	-25.133 (-1.54)
TradingDiscussions$_t$	-1.651*** (-3.83)	-2.361* (-1.95)	-0.696 (-0.15)	-5.672 (-0.46)
R_{t-1}	0.098*** (2.79)	1.304*** (13.19)	0.719* (1.86)	0.814 (0.81)
R_{t-1}	-0.039 (-1.12)	1.152*** (11.88)	0.643* (1.69)	1.247 (1.26)
Volatility$_t$	0.025* (1.95)	0.151*** (4.25)	-0.499 (-3.59)	-2.882*** (-7.98)
TraditionalMedia$_t$	0.341 (1.10)	-0.848 (-0.98)	-7.042** (-2.08)	-10.244 (-1.16)

***=P<0.01, **=P<0.0 5, *=P<0.1.

Next, we investigate if the information provided by different user groups on social media platforms differs in informativeness for future price movements. As mentioned before, influential people on social media usually possess three attributes: (1) they are convincing; (2) they are experts, and (3) they have a lot of social ties. In this paper, we focus on the second point. What kind of social media users provide accurate information? This is the question we try to answer to test our Hypothesis 2.

On Bitcointalk.org, there are several user badges. From high to low in terms of activity level, they are: Legendary, Hero Member, Senior Member, Full Member, Member, Junior Member Newbie and Brand New. The activity score is calculated based on activity levels on Bitcointalk.org and the time since registration, specifically, Activity = min (time×14, total # posts), which means that high-level users are those who are active on this message board for a long enough time. We define two user groups: the active user group (Legendary, Hero Member, Senior Member, and Full Member level users) and the inactive user group (Member, Junior Member, Newbie and Brand New level users). We calculate the social media sentiments for topics initiated by active users and inactive users respectively and then redo Table 5 for each user group. The results are shown in Table 6.

We observe dramatic differences between active and inactive users in terms of predictive power in Table 6. Panel A reports the results for active users, and Panel B reports the results for inactive users. For discussions under the Economics category, active users do not predict returns for any of the four holding periods (first row of Panel A). However, in Table 5, Economics related discussions by all social media participants combined demonstrate predictive power for the long-term returns (three-month cumulative return). Therefore, this predictive power has to come from the inactive users. And this is indeed the case as shown in the Panel B of Table 6. The first row in Panel B evidences the predictive power of the Economics related topics for the inactive users. As we expected, *Economics$_t$*, the sentiment of the fundamental-related information, does not provide much valuable information for short-term price movements, but successfully predicts the long-term returns, and the coefficient estimates are almost three times larger than they are in Table 5.

Table 6: Comparison Between Active and Inactive Users

	R_t	$R_{t,t+7}$	$R_{t,t+30}$	$R_{t,t+90}$
Panel A: Active User				
Economics$_t$	-0.816 (-1.41)	-0.665 (-0.41)	5.503 (0.88)	-22.514 (-1.43)
Speculation$_t$	-1.086 (-1.89)	-3.105* (-1.93)	3.649 (0.58)	-2.63 (-0.17)
TradingDiscussion$_t$	-1.795*** (-3.37)	-3.600** (-2.42)	4.457 (0.77)	5.666 (0.39)
Panel B: Inactive User				
Economics$_t$	-1.253 (-1.48)	-4.616* (-1.95)	-17.171** (-1.92)	-153.180*** (-6.62)
Speculation$_t$	-2.465** (-2.32)	-9.281*** (-3.12)	-33.282*** (-2.87)	-109.933*** (-3.78)
TradingDiscussion$_t$	-1.624*** (-2.84)	-0.618 (-0.39)	-11.046* (-1.78)	-32.639** (-2.09)

***=P<0.01, **=P<0.0 5, *=P<0.1.

For speculative information, the active user group shows predictive power only in the short-term (the second and third rows of Panel A). However, for the inactive users, the coefficient estimates are also significant for the one-month cumulative return prediction and the three-month cumulative return prediction (the second and third rows of Panel B). These results reveal that active participants are not necessarily informative on social media. The valuable information more likely comes from less active users because they share information with other not for emotional benefits (social comfort, maintaining reciprocal relationships, etc.), but to make a valid point. Even if those inactive social media participants do not post frequently, as long as the information or judgment is accurate, the price change in the future will ultimately confirm the value of the information.

To summarize, our analysis presented in Table 6 provides evidence that inactive users offer better predictions for future Bitcoin price movements, while the active users do not. The intuition behind this observation is that active users usually talk on social media for the sake of talk, while inactive users usually talk on social media to make a point. They have different motivations to share on social media which leads to differences in informativeness. However, active users still provide valuable information for short-term price movement, and the predictive power difference mainly shows up in the long run.

IV. Conclusion

The development of information technology has made available new sources of information to assist investments for retail investors. In this paper, we examined whether unregulated social media provide valuable information for short-term and long-term predictions of Bitcoin valuation. We found that it is possible for retail investors to identify value-relevant information via communications over social media. The main results in our research that add to the related literature are that fundamental-related information predicts only long-term returns, while speculative information predicts both long-term and short-term returns. Also, we found that active users on social media do not overlap with inactive users with accurate information. Information provided by inactive user exhibits stronger predictive power than that of active users, especially in long-term prediction.

With the rapid development and usage of social media, there is a huge amount of social data generated each day. Knowing who provides more accurate information is crucial. Our research provides guidelines for identifying useful information on social media. Our research also suggests ways to estimate the "wear-in" time of different types of information (speculative information and fundamental-related information). This is another important factor to consider when predicting future price movements with social data.

Lastly, we point out some limitations of the paper and propose future research opportunities to extend this paper. The dataset used in this paper is a time series dataset; though we can eliminate the time-invariant effects by controlling for the lagged price movements, it is hard to control the general trend over time. For future research, a panel data collected for multiple cryptocurrencies may solve the problem by adding time-fixed effects and cryptocurrency-fixed effects to the model.

A more challenging, but arguably more important question is the effect of real and fake news on the Bitcoin prices. Is it possible to distinguish the real news from the fake news? If so, will it affect the prices in a different way? Investors have limited capability to tell fake news items from real ones, especially in a market with limited access to official news outlets. As a result, the fake news may also have a significant impact on trading. It will be a breakthrough if we can potentially identify fake news in social data, and compare its effects to that of real news.

References

Antweiler, Werner, and Murray Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance,* 59(3): 1259-94.

Cayman Atlantic. 2014. "2014 Performance", http://www.caymanatlantic.com/performance/year-2014-performance.html (accessed March 5, 2017).

Cayman Atlantic. 2015. "2015 Performance", http://www.caymanatlantic.com/performance/year-2015-performance.html (accessed March 5, 2017).

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media." *The Review of Financial Studies,* 27(5): 1367-1403.

Das, Sanjiv R. and Mike Y. Chen. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web." *Management Science,* 53(9): 1375-88.

Davis, Angela K., Jeremy M. Piger and Lisa M. Sedor. 2012. "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language." *Contemporary Accounting Research,* 29(3): 845-68.

DeMarzo, Peter M., Jeffrey Zwiebel, and Dimitri Vayanos. 2001. "Persuasion Bias, Social Influence, and Unidimensional Opinions." Working Paper 1719.

Dewally, Michaël. 2003. "Internet Investment Advice: Investing with a Rock of Salt." *Financial Analysts Journal,* 59(4): 65-77.

Dougal, Casey, Joseph Engelberg, Diego García, and Christopher A. Parsons. 2012. "Journalists and the Stock Market." *The Review of Financial Studies,* 25(3): 639-79.

Eyal, Ittay, and Emin Gün Sirer. 2014. "Majority Is Not Enough: Bitcoin Mining Is Vulnerable." in *Financial Cryptography and Data Security,* ed. Nicolas Christin and Reihaneh Safavi-Naini, 436-54. Ithaca, NY: Springer-Verlag Berlin Heidelberg.

Glaser, Florian, Kai Zimmermann, Martin Haferkorn, Moritz Christian Weber, and Michael Siering. 2014. "Bitcoin - Asset or Currency? Revealing Users' Hidden Intentions." Paper presented at the Twenty Second European Conference on Information Systems, Tel Aviv.

Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong. 2009. "The Role of Hubs in the Adoption Process." *Journal of Marketing,* 73(2): 1-13.

Harvard Psychosociological Dictionary. 2017. "Descriptions of Inquirer Categories and Use of Inquirer Dictionaries." http://www.wjh.harvard.edu/~inquirer/homecat.htm (accessed July 24, 2017).

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. 2004. "Social Interaction and Stock-Market Participation." *The Journal of Finance,* 59(1): 137-63.

Johnson, Benjamin, Aron Laszka, Jens Grossklags, Marie Vasek, and Tyler Moore. 2014. "Game-Theoretic Analysis of DdoS Attacks Against Bitcoin Mining Pools," in *Financial Cryptography and Data Security,* ed. Rainer Böhme, Michael Brenner, Tyler Moore, and Matthew Smith, 72-86. Berkeley, CA: Springer-Verlag Berlin Heidelberg.

Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance,* 66(1): 35-65.

Luo, Xueming, Jie Zhang, and Wenjing Duan. 2013. "Social Media and Firm Equity Value." *Information Systems Research,* 24(1): 146-63.

McLure-Wasko, Molly, and Samer Faraj. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice." *MIS Quarterly*, 29(1): 35-57.

Miller, Andrew, Ahmed Kosba, Jonathan Katz, and Elaine Shi. 2015. "Nonoutsourceable Scratch-Off Puzzles to Discourage Bitcoin Mining Coalitions." *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*: 680-691.

O'Dwyer, Karl J., and David Malone. 2013. "Bitcoin Mining and Its Energy Footprint." Paper presented at the Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies, Ireland, June 26-27.

Reid, Fergal, and Martin Harrigan. 2013. "An Analysis of Anonymity in the Bitcoin System." in *Security and Privacy in Social Networks,* ed. Yaniv Altshuler, Yuval Elovici, Armin B. Cremers, Nadav Aharony, and Alex Pentland, 197-223. Dublin: Springer New York.

Shi, Zhan, Huaxia Rui, and Andrew B. Whinston. 2014. "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter." *MIS Quarterly,* 38(1): 123-42.

Solomon, David H. 2012. "Selective Publicity and Stock Prices." *The Journal of Finance,* 67(2): 599-638.

Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance,* 62(3): 1139-68.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *The Journal of Finance,* 63(3): 1437-67.

Tirole, Jean. 1982. "On the Possibility of Speculation Under Rational Expectations." *Econometrica,* 50(5): 1163-81.

Tirunillai, Seshadri, and Gerard J. Tellis. 2011. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance." *Marketing Science,* 31(2): 198-215.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin. 2009. "Determining Influential Users in Internet Social Networks." Ssrn elibrary.

Tumarkin, Robert, and Robert F. Whitelaw. 2001. "News or Noise? Internet Postings and Stock Prices." *Financial Analysts Journal,* 57(3): 41-51.

Womack, Kent L. 1996. "Do Brokerage Analysts' Recommendations Have Investment Value?" *The Journal of Finance,* 51(1): 137-67.

Wysocki, Peter D. 1998. "Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards." *University of Michigan Business School Working Paper No. 98025*.

Yermack, David. 2013. "Is Bitcoin a Real Currency? An Economic Appraisal." *NBER Working Paper No. 19747.*

.