From Bitcoin to Big Coin:

The Impacts of Social Media on Bitcoin Performance

Feng Mai
Department of Operations, Business Analytics, and Information Systems
Carl H. Lindner College of Business
University of Cincinnati
Cincinnati, Ohio 45221-0130, USA
maifg@mail.uc.edu

Qing Bai
Department of Accounting and Finance
College of Business
University of Wisconsin-Eau Claire
Eau Claire, WI 54702-4004, USA
baiq@uwec.edu

Zhe Shan
Department of Operations, Business Analytics, and Information Systems
Carl H. Lindner College of Business
University of Cincinnati
Cincinnati, Ohio 45221-0130, USA
zhe.shan@uc.edu

Xin (Shane) Wang
Marketing Discipline
Ivey Business School
Western University
London, Ontario, Canada
xwang@ivey.uwo.ca

Roger H. L. Chiang
Department of Operations, Business Analytics, and Information Systems
Carl H. Lindner College of Business
University of Cincinnati
Cincinnati, Ohio 45221-0130, USA
roger.chiang@uc.edu

Abstract

As the world's first completely decentralized digital payment system, Bitcoin represents a revolutionary phenomenon in financial markets. This study examines predictive relationships between social media and bitcoin returns by considering the relative effect of different social media platforms (Internet forum vs. microblogging) and the dynamics of the resulting relationships using vector autoregressive and vector error correction models. The results suggest that more bullish forums have a positive, statistically significant relationship with future bitcoin returns at a daily level. Internet forum predictive metrics outperform microblogging ones at a daily frequency, but their effects are opposite at an hourly frequency. The user-generated content contributed by the vocal minority and the silent majority exhibit distinct relationships with bitcoin performance, in terms of both transaction volume and returns. The implications of these results for research and practice are notable with regard to the transformative power of social media analytics in networked business environments subject to the dynamics of bitcoin performance.

Keywords: Bitcoin, cryptocurrency, Internet forum, prediction value, social media, Twitter, user-generated content

Since the 2008 invention of Bitcoin by an unidentified programmer known as Satoshi

Nakamoto, the virtual currency has achieved great success, such that by 2013, the value of all

bitcoins in the world surpassed 1 billion USD.[1] Remarkable for a cryptocurrency that exists

solely in digital form and is backed by no central bank or other authority, a bitcoin in September

2014 was worth US$400–500 and traded by approximately 1.5 million Bitcoin users, who

exchange more than 100,000 bitcoins daily. Well-known companies such as Dell and

Newegg.com accept bitcoin; several Bitcoin ATMs now operate in five cities on four continents.

According to CoinDesk's (an online publication that tracks digital currencies) recent estimate,

there will be eight million bitcoin trading accounts and 100,000 companies that accept bitcoin by

the end of 2014. Thus the rise of Bitcoin seems unstoppable.

Amidst this hype, growing concern has focused on bitcoins' considerable price volatility and

associated risks. Since 2011, several significant prices adjustments have affected bitcoins,

including a price drop of 23% when Mt. Gox, the oldest and then-largest exchange, went offline

and "lost" 650,000 members' bitcoins. According to PricewaterhouseCoopers (2014), 82% of

consumers express concern about fluctuations in the bitcoin market, and some analysts suggest

the presence of a bitcoin bubble waiting to burst (Salmon 2013). Bitcoin also has come under

scrutiny for its potential uses in illegal activities, such as money laundering, leaving users

anxious about its legal status and the possibility of a government crackdown (Grinberg 2011). In

this sense, it remains unclear whether Bitcoin is sufficiently stable to last.

Despite widespread descriptions of bitcoin as a cryptocurrency, virtual currency, or digital

currency, its status as a "currency" is disputable. Yermack (2013) notes that bitcoin's daily

exchange rates exhibit greater volatility and virtually no correlation with conventional

---

[1] We use "Bitcoin" to refer to the online payment system; "bitcoin" is the unit of currency.

currencies. In this sense, bitcoin resembles a financial investment, such as an Internet stock, rather than a currency. Moreover, the factors that determine the bitcoin's growth and influence its value are unclear; Ciaian et al. (2014) argue that Wikipedia views have a statistically significant impact on bitcoin prices, but the effect is complex.

Because of Bitcoin's decentralized structure and exclusively online presence, the value of bitcoins derives not from gold or government fiat but from the value that people assign. The dynamics of the bitcoin price thus should relate to pertinent discussions and opinions on online social media, where investors and business adopters interact and provide feedback about the market. Tirunillai and Tellis (2012) have posited that social media and user-generated content (UGC) constitute important determinants of investments. Social media capture the "wisdom of the crowd" and provide low-cost platforms for connecting with target markets. Bitcoin provides a unique opportunity to observe and understand the interplay of social media with the value of a financial instrument. Accordingly, we examine the influence of social media metrics on bitcoin returns and strive to answer the following research questions:

- What factors characterize trading behavior surrounding bitcoins? Does it behave like a currency or resemble a speculative investment, similar to Internet stocks?

- Can user-generated content, available through social media, predict value in the bitcoin market? Do distinctive social media characteristics (e.g., user and platform differences) affect this predictive relationship?

We conduct an empirical analysis of the bitcoin market, using vector autoregressive (VAR) and vector error correction (VECM) models with exogenous control variables. In cases when some variables share common trends, the VECM provides a more appropriate framework than standard VAR models, because it can explore dynamic movements among cointegrated focal

4

variables. We assemble diverse data sources, from bitcoin and stock markets, traditional Internet measures, and social media; we also construct social media metrics using data from an Internet forum (bitcointalk.org) and a microblogging service site (Twitter) to assess the volume of posts, user sentiments, and measures of bullishness and disagreement across forum contributors. To incorporate the influence analysis in our predictive models, we stratify the sample by the characteristics of social media users, dividing users of the Internet forum and Twitter into the "silent majority" and "vocal minority," according to their contribution levels.

A summary of our major findings is as follows. We determine that correlations among bitcoin market variables (return, volatility, trading volume) are consistent with the well-documented relationships that appear in the stock market. To the extent that social media can predict future bitcoin prices and trading volume, we find that the number of bullish (bearish) forum posts has a positive (negative), statistically significant relationship with future bitcoin returns when we consider daily frequencies. At an hourly frequency, the number of bullish tweets reveals a positive, statistically significant relationship with future bitcoin returns. Disagreement among forum contributors also predicts future bitcoin trading volume at a daily frequency. The UGC obtained from different users and platforms exerts different impacts on bitcoin returns, such that UGC contributed by the vocal minority and silent majority display distinct relationships with the future bitcoin market. Finally, regarding the relative predictive power of both types of social media, the daily data indicate that the Internet forum variables influence future bitcoin prices significantly, but Twitter variables do not. The hourly data instead show that Twitter variables offer more predictive value.

With these findings, our study makes several unique contributions. First, to the best of our knowledge, this study is the first investigation of predictors of bitcoin returns. Prior studies

mainly discuss bitcoin in general (Barber et al. 2012; Grinberg 2011; Kroll et al. 2013; Moore and Christin 2013) or explore its price formation in particular (Kristoufek 2013; 2014). Although research has identified several factors that affect bitcoin prices, such as supply and demand or attractiveness for investors, none of these factors have been explained using standard economic theory (Kristoufek 2014). Noting Yermack (2013) claim that bitcoin resembles a speculative investment, we offer empirical evidence that bitcoin trading exhibits characteristics similar to stock trading, which provides insights for regulators and potential investors.

Second, there exists the significant impact of social media on future bitcoin returns. Social media offer substantial information about bitcoin's acceptance among the general public, as well as daily fluctuations in its market sentiments. Investors thus can gain insights into bitcoin's value from this rich information environment. We also reveal how different users and platforms affect the influence of UGC on bitcoin performance, leading to our recommendation that investors choose an appropriate information channel and source to gather insights for their specific investment decisions.

Third, this study contributes to research into the impact of social media on firm or market performance more generally. The UGC available through social media provide valuable and timely information for investors about market acceptance and performance prediction. Prior studies note the relationship between digital user metrics and product sales (Chevalier and Mayzlin 2006; Dellarocas et al. 2007; Dhar and Chang 2009; Ghose and Yang 2009; Moe and Fader 2004), link social media to firm equity (Luo et al. 2013), or stock market performance (Tirunillai and Tellis 2012). We extend this stream by examining the relationship between UGC and bitcoin returns. Bitcoin's exclusively online presence makes it an ideal setting to investigate the impacts of social media on firm performance and market acceptance.

Fourth, we employ multiple sources (Internet forum and microblogging site), various data formats (numeric and textual), and different facets of social media metrics (sentiment and influence) to address their relative effects. Many studies rely on a single source; we explicitly consider how different types of social media create distinct relational ties among users and thus varied networking behaviors. Moreover, content sentiment and social influence may interact, generating disparate effects on information cascading and reception across different social media platforms, such as the two we study. These differences shape the formation and characteristics of social media networks, which suggest new measures for capturing user behaviors in online communities.

Fifth, spanning the disciplines of information systems (IS), marketing, and finance, the current research can help individuals, firms, and government agencies develop strategies for using big data and their analytics. As a disruptive digital technology, Bitcoin is fundamentally changing the financial ecosystem, which affects every area of society in the modern, networked business environment. The distributed, self-regulated paradigm of Bitcoin already has invoked the emergence of a participatory market economy. In this socio-economic transformation, online social media integrate people into networks, thereby enabling collective intelligence (Helbing 2014). The user bases of Bitcoin and social media are well aligned, suggesting the great potential for using them in conjunction (McNulty 2014). Our study, which combines the value streams of Bitcoin and big data, provides a new potential methodological paradigm for predictive analytics.

Sixth, we apply data mining, sentiment analysis, and econometric techniques to explore the societal and managerial impacts of big data. The findings can help financial investors and government legislators mitigate the risk and harness the potential of virtual currencies. For big data analytics in the IS field (Shmueli and Koppius 2011), we develop new social media

7

measurements that incorporate sentiment, influence, and dynamics signals; we also consider

cointegration among time-series variables by applying VECM models. By assessing the

predictability of empirical phenomena in bitcoin returns, our work contributes to the

development of a new virtual currency theory.

In the next section, we develop our theoretical background and hypotheses. We then

introduce the measures and data sample for our empirical study, after which we develop our

empirical models (VAR and VECM). Subsequent to the discussion of our findings and

robustness checks, we conclude this article with some implications and insights.

THEORETICAL BACKGROUND AND HYPOTHESES

Bitcoin: Currency or Stock?

Because bitcoin is the most popular virtual currency, a reasonable expectation is that it

behaves similarly to traditional currencies, such that its price would be driven by its use in

transactions, its supply, and the price level (of tradable goods and services) (Kristoufek 2014). In

a narrow sense, a digital currency is just another medium of exchange that happens to be

electronically created and stored. However, virtual currency is a fundamentally different

financial phenomenon, defined as "a type of unregulated, digital money, which is issued and

usually controlled by its developers, and used and accepted among the members of a specific

virtual community" (European Central Bank 2012). That is, a digital currency can be the

electronic form of a traditional currency, while a virtual currency could be a totally distinct

economic system. Unfortunately, government agencies have not reached consensus on how to

treat virtual currencies,[2] and no systematic study of virtual currencies has appeared in finance literature.

Although some technology communities assert that bitcoin is digital, not virtual, financial authorities generally consider it virtual money (European Banking Authority 2014; Network Financial Crimes Enforcement 2013). Transactions on the Bitcoin network are not denominated in dollars or other currency, so bitcoins are a virtual currency, supported by a decentralized payment network. Because, as we stated previously, the value of bitcoin derives not from gold or government fiat but from the value that people assign to it, its dollar value gets determined on an open market, similar to the exchange rate among world currencies (Brito and Castillo 2013). Yet the bitcoin's daily exchange rates do not correlate with traditional currencies (Kristoufek 2014), and its exchange rate volatility is orders of magnitude greater than the volatilities of those more widely used currencies (Yermack 2013). In these traits, bitcoin seemingly mimics an Internet stock, rather than a currency.

To test empirically whether bitcoin shares the characteristics of stocks, we examine the dynamic relationship between its prices and volume, and compare it against the price–volume relationship that Antweiler and Frank (2004) establish in stock trading: (1) positive correlation between conditional volatility and volume, (2) large price movements followed by high volume, (3) conditioning on lagged volume that substantially attenuates the leverage effect, and (4) conditioning on lagged volume that produces a positive risk–return relation. Although we cannot test the latter two features in the bitcoin context, we expect to observe similar volume-related behaviors and therefore hypothesize:

---

[2] For example, the U.S. Internal Revenue Service treats bitcoin and other virtual currencies like property, similar to stocks, whereas the Australian Taxation Office regards bitcoin transactions as akin to barter arrangements.

H1: Bitcoin trading exhibits features similar to those of stock trading, such that (a) its

conditional volatility and trading volume correlate positively, and (b) large bitcoin price

movements are followed by high trading volume.

Can Social Media Predict Bitcoin Returns?

The efficient market hypothesis asserts that new information may change market

expectations about a firm and thereby affect its stock price (Fama 1970). If bitcoin resembles

stocks, bitcoin price movements should follow new information. In modern society, such new

information often becomes available through social media and Web 2.0 applications, which

fundamentally have changed the interactions between consumers and firms (Gallaugher and

Ransbotham 2010). For example, online word of mouth, in the form of consumer reviews or

ratings and blogs, constitute prominent sources of consumer and investor information about a

firm's future performance prospects (Chen and Xie 2008; Gu et al. 2012). Prior studies examine

the relationship between digital user metrics and product sales: Liu (2006), Dellarocas et al.

(2007), Duan et al. (2008a; 2008b), and Chintagunta et al. (2010) consider movies; Forman et al.

(2008) and Chevalier and Mayzlin (2006) investigate books; Dhar and Chang (2009) and Dewan

and Ramaprasad (2012) examine music; Godes and Mayzlin (2004) assess television shows; Zhu

and Zhang (2010) address video games; and Luo (2007; 2009) study airline services. Yet all

these studies focus on offline services; bitcoin represents a unique study context, due to its

inherently online character.

Researchers have tested whether the Internet in general, and UGC specifically, influences the

underlying behavior of stock markets. Based on a sample of the 50 firms with the greatest Yahoo

message board posting volume, Wysocki (1999) finds that changes in daily posting volume are

associated with both earnings announcement events and changes in stock trading volume and

returns. Tumarkin and Whitelaw (2001) examine Internet stocks and find that message board activity cannot predict stock returns; rather, the causality appears to run from the market to the forums. In contrast, Antweiler and Frank (2004) indicate that a positive shock to message board posting predicts negative stock returns on the next day, though the effect is economically small. Both the level of message posting and disagreement among messages seemingly predict subsequent trading volume. In their analysis of articles published on a social media platform, Chen et al. (2014) reveal that the views expressed in both articles and comments predict future stock returns and earnings surprises, with an effect that is both statistically and economically significant. Although Das and Chen (2007) find, among a sample of nine firms, that stock messages reflect information quickly, they uncover no ability to forecast stock returns. Luo et al. (2013) examination of the dynamic relationship between social media (consumer ratings and Web blogs) and firm equity value suggests that social media metrics have significant predictive power for firm equity value. They also examine the relative effects of social media metrics compared with conventional online behavior metrics (e.g. Google searches, Web traffic, etc.) and find that predictions based on social media are faster than those from conventional online media.

We anticipate that the association between social media and Bitcoin is similar to, if not stronger than, that of Internet stocks and firm equity for the following reasons. The decentralized nature of Bitcoin meant that most early users were individuals, rather than large institutional investors, who arguably contribute to social media more frequently and are more likely to be influenced by social media. An important motivation for early institutional Bitcoin adopters was to capture positive public relations through social media, in that "being noted as a Bitcoin innovator can potentially generate favorable press and social media mentions" (PricewaterhouseCoopers 2014). For instance, when the social gaming company Zynga added

Bitcoin to its most popular games in 2014, it garnered thousands of media mentions. Furthermore, the design of Bitcoin's algorithm ensures that the supply of new coins gets created at a known, geometrically decaying rate, so demand from both businesses and individuals represents the main driver of the bitcoin's value. Finally, according to a recent survey (Duggan and Brenner 2013), Bitcoin users largely share the demographic characteristic of being heavy social media users.

The UGC on online platforms thus should predict the investment returns and volume of bitcoins. First, online messages can disclose new or private information that fundamentally alters bitcoin evaluations, such as when new stores accept bitcoins or forthcoming regulations limit its use. Second, online discussions offer good indications of the general market sentiment toward bitcoins. Third, speculative investors tend to follow the trends, which may exaggerate the effects of such information. In turn, we postulate:

H2. Social media metrics have significant effects on future bitcoin returns, such that (a) increased positive (negative) sentiments indicate higher (lower) future bitcoin prices and (b) disagreements on social media indicate greater future trading volume.

The Significant Role of Social Influence in Predicting Bitcoin Prices

The power of social influence in financial markets is well documented, manifested as herding behavior. Friend et al. (1970) note the significant tendency of groups of mutual funds to adopt the investment choices of their more successful counterparts, which they call follow-the-leader behavior. Jiao and Ye (2013) find strong evidence that mutual funds collectively enter or exit stocks, following the herd of hedge funds, whereas hedge funds do not follow mutual funds. According to Brown et al. (2013), mutual fund managers follow analyst recommendation revisions when they trade stocks, and these analyst-motivated trades move stock prices. Mutual

funds herd into stocks following consensus analyst upgrades and even more evidently herd out of stocks with consensus downgrades. We expect to observe some degree of similar follow-the-influencer behavior in the bitcoin market.

We also rely on marketing and IS literature, which highlights the importance of identifying influential users on social media. As Trusov et al. (2010) observe, community members differ in the frequency, volume, type, and quality of digital content they generate and consume. Influential people, such as opinion leaders, have disproportionate influence on others (Godes and Mayzlin 2009; Goldenberg et al. 2009), largely because they have greater exposure to mass media than their followers, They are more cosmopolitan, engage in more social participation, have a higher socioeconomic status, and are more innovative (Rogers 2010). The effectiveness and functioning of an online community strongly depends on the presence and activities of a vocal minority of opinion leaders, who can induce effects in various ways (Mehra et al. 2006). To leverage the effect of social influence on product adoption, many companies seek to initiate and control the diffusion process by targeting the most influential people in a social network (Bonchi et al. 2011; Hinz et al. 2011; Libai et al. 2010). However, the power-law nature of social media implies that most social media users contribute little content; this "silent majority" contributes to conversations sporadically, mostly after important events, and are not particularly interested in generating buzz (Metaxas and Mustafaraj 2012; Mustafaraj et al. 2011). In this sense, the UGC from the silent majority may be a more compelling metric for actual investors. Therefore, we hypothesize:

H3. The silent majority and vocal minority have distinct impacts on the bitcoin market.

The Distinct Predictive Value of UGC from Internet Forum versus Microblogging

In addition to user-level influence differences, we predict that different social media platforms affect financial markets differently. First, Internet forums generally seek to achieve diverse opinions, and consensus is not a primary objective. These forums are geared toward creating a collaborative environment that answers different users' similar queries about a topic. In contrast, on a microblogging site such as Twitter, communications move from the sender to his or her followers, who can spread the information further by retweeting. Limited by length restrictions, these followers might add brief, general sentiments, but they cannot engage in a thorough discussion of the original content. Second, a discussion forum is not a typical social network application, because it enables users to engage in discrete, transitory exchanges, without respect to proximity, social relations, or flows. But as a classical social networking service, Twitter supports a richer range of possible relational ties, including social relations in which users establish persistent connections with (i.e., follow) other users. Third, most users access an Internet forum through Web browsers. Twitter offers both web and mobile access, but users mainly engage through mobile devices. These distinctive features in turn may have significant impacts on the predictive power of UGC on these social media for bitcoin returns.

Accordingly, the link between the nature of a social media platform and the adoption of UGC has received attention. A one-sided message presents positively or negatively valenced information; a two-sided message can include both (Cheung and Thadani 2012). Kamins and Assael (1987) observe that two-sided information enhances information completeness, invoking greater credibility perceptions. Finance scholars also note that because investors have limited attention capacities, they respond asymmetrically to more visible information (Barber and Odean 2008; Hirshleifer and Teoh 2009): when information is more visible and accessible, investors are

more likely to respond to it. The relationship between discussion patterns and bitcoin prices thus should appear at a daily level, whereas the responses of the bitcoin market to the spread of news on Twitter may occur at an intra-day level, due to Twitter's mobile nature (Tafti et al. 2013). We hypothesize:

H4. User-generated content from Internet forum and microblogging site have different predictive values for bitcoin returns.

DATA AND VARIABLES

Bitcoin Market Variables

The data set comprises daily market prices (i.e., USD exchange rate) and trading volume series (in USD) from BitStamp Ltd., the top bitcoin exchange by volume. We also collected bitcoin-to-bitcoin transaction volume, defined as the total value of all transaction outputs per day,[3] from bitcoincharts.com. The exchange trading volume measure refers to the amount of bitcoin traded for other currencies; transaction volume indicates the amount spent in the bitcoin economy. Because the total output volume includes coins returned to the sender as change, we use the adjusted transaction volume, net of change, which should offer a more accurate reflection of the true transaction volume. We denote the trading volume and transaction volume of day t as V_t and V_t^{TX}, respectively.

In addition, we define S_t as the market price of bitcoin at the end of day (hour) t. The change in bitcoin price/value is the first difference of the price:

---

[3] A transaction is a signed section of data, broadcast to the network and collected in blocks. It typically references previous transaction(s) and dedicates a certain number of bitcoins to one or more new public key(s) (i.e., Bitcoin address). It is not encrypted; nothing in Bitcoin is encrypted.

$$\Delta P_t = S_t - S_{t-1}. \tag{1}$$

To measure the volatility of the daily/hour price change, we use multiple measures, including volatility at a daily frequency. To define the continuously compounded return per day on day t, r_t, we use $\ln \frac{S_t}{S_{t-1}}$; applying the exponentially weighted moving average (EWMA) model, we can estimate the volatility of daily return. The EWMA model tracks changes in the volatility, with the formula:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda)u_{t-1}^2. \tag{2}$$

The estimate of volatility on day t, σ_n^2 (obtained at the end of day $t - 1$), is calculated from σ_{t-1}^2 (i.e., the estimate at the end of day $t - 2$ of volatility for day $t - 1$) and u_{t-1}^2 (most recent daily percentage change). The value of λ governs the responsiveness of the estimate to the most recent daily percentage change. We choose $\lambda = .94$, the value used by RiskMetrics.[4] In addition, at an hourly frequency, we measure the absolute value of the hourly return $|r_t|$.

Social Media Metrics

We implemented a Python-based Web crawler to collect discussion content from bitcointalk.org between November 22, 2009, and August 18, 2014. We chose this forum for two reasons: It was rated the most popular Bitcoin community in a recent survey (Smyth 2013), and it appears first in the community section of the official Bitcoin website. We limit our data collection to the Bitcoin discussion board, to which users contribute general news, community developments, innovations, and so forth. After filtering out content beyond our study period, we gathered 119,847 posts and 51,269 topics to retain for further analysis. Each post contained

---

[4] The RiskMetrics database, originally created by JPMorgan and made publicly available in 1994, uses a EWMA model and $\lambda = .94$ to update daily volatility in its database. The company demonstrated that, across a range of market variables, this value of λ results in variance rate forecasts that come closest to the realized variance rate.

textual content, an author, and a timestamp. Among the 69,671 unique users who posted, the most active 5% of users generated 62.6% of the content. The average number of posts generated by a single user in the sample period was 12.75; the median was 3. As Figure 1 reveals, the distribution of the number of messages by users has a very long tail, such that most users are in the silent majority, and a small proportion of the vocal minority generated the most contents.
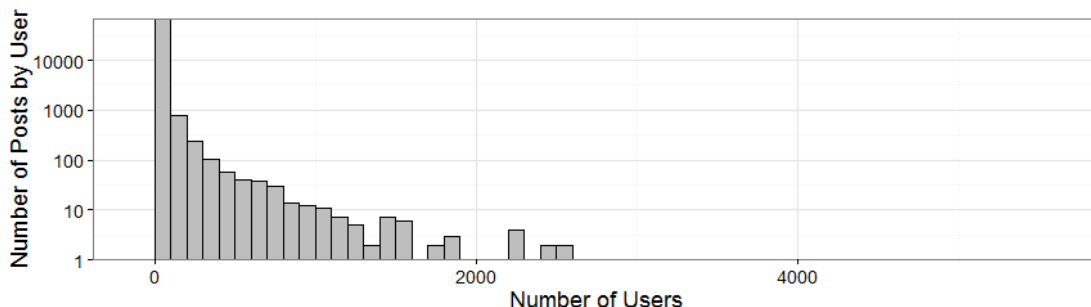

Figure 1: Distribution of posts by users

For the sentiment analysis, we applied a finance sentiment dictionary (Loughran and McDonald 2014), which includes 2,329 negative and 297 positive sentiment words. For this implementation, we used Natural Language Toolkit 3.0 (Bird 2006) for the language processing tasks, such as sentence segmentation, word tokenization, and lemmatization. For each post, we counted the number of positive and negative words. If a post contains more positive than negative words, it constitutes a positive post, and vice versa.

Because, unlike common stocks, bitcoin is traded constantly, such that it has no daily closing price per se, we chose 18:15:05 universal coordinated time (UTC) as a cut-off point, so that we could match UGC with the bitcoin daily return data. That is, if a post was submitted after 18:15:05 UTC, it belongs to the content of the next day. This particular cut-off time reflects the moment of the very first bitcoin transaction log, at 18:15:05 UTC on January 3, 2009.

Next, we collected tweets (i.e., microblogging text messages of no more than 140 characters) that contained the hashtag "#Bitcoin" from the public application program interfaces (API) of

17

Twitter. Twitter's search APIs allow queries against the indices of recent or popular tweets and can collect a wider range of data, such as latest favored or retweeted counts. Using a Python-based Web crawler, we collected data from the search API at its highest frequency (limited to 180 queries per 15-minute window) between April 18th and August 18th, 2014, during which we gathered 3,348,965 unique tweets from 339,295 unique users. In average 21,910 users tweeted 27,227 messages per day. Each tweet contains textual content, author information, a timestamp, and propagation-related flag and counting data. With these retrieved data, we again applied the sentiment dictionary (Loughran and McDonald 2014) and the NLTK toolkit to count the number of positive and negative words in each tweet. If the number of positive words in a post is greater than the number of negative words, the tweet is classified as positive, and vice versa. Also, we applied the same cut-off point (18:15:05 UTC) to summarize the daily information.

Finally, we considered the distinct influence of each tweet, according to its characteristics. In our data set, for each tweet, we collected *favorite_count*, to assess how many times a tweet had been listed as a "favorite" by Twitter users; *retweet_count*, or the number of times a tweet had been retweeted; and *followers_count* for the tweet's author, equivalent to the number of followers the account had at that moment. However, *favorite_count* and *retweet_count* are continuously updated along the tweet lifecycle, and are difficult to track their updates at the hourly and even daily frequencies. Therefore, we identified *follower-count* as the measure that offered insightful influence information. For each day, we ranked all tweets according to this measure, chose the top 20 tweets in the ranked list (following the practice in (Shi et al. 2014)), and calculated their sentiment scores.

Other Variables

We included a set of traditional Internet activity measures and stock market returns (S&P500 and NASDAQ composite) as exogenous control variables. To measure search interest related to Bitcoin, we collected data from Google Trends (www.google.com/trends/). The measure of *interest over time* indicated the popularity of a given keyword (in our case, *bitcoin*) in Google's search engine, using a 0–100 scale and normalized values. Because Google Trends provides only weekly data, we used the previous week's search interest measure to apply to each day in the subsequent week. We also gathered Web traffic data from the Alexa Web Information Service (aws.amazon.com/awis/). A Python program fetched traffic data related to bitcoin.org, including *reach* (number of unique visitors), *page views per user* (average number of webpages a user visits), and *traffic rank* (estimated daily ranking of the website). Table 1 summarizes our key measures.

Table 1: Key Measures

Variable	Meaning
P_t	Daily bitcoin price in USD
r_t	Bitcoin returns, continuously compounded
σ_t^2	Volatility of bitcoin returns, updated daily
V_t	Daily trading volume (logged, liner trend removed) from top exchanges
V_t^{TX}	Estimated daily transaction volume (logged, liner trend removed)
POS^F	Number of positive forum posts in a day
NEG^F	Number of negative forum posts in a day
POS^T	Number of positive tweets in a day
NEG^T	Number of negative tweets in a day

EMPIRICAL METHODOLOGY

We are interested in both the contemporaneous and the dynamic relationships between social media and the bitcoin market. With our contemporaneous analysis, we seek to determine whether

variation in social media activities is just noise or is associated with underlying market activities. Therefore, we examined four fundamental measures of bitcoin market activities: bitcoin returns, volatility of bitcoin returns, trading volume on major bitcoin exchanges, and bitcoin transaction volume. Changes in trading volume and volatility can serve as proxies for actual news and reactions to this news by market participants (Andersen 1996), so we expect these measures to relate positively to changes in message posting volume if information in the bitcoin market gets reflected quickly in social media activities (e.g., message posts, tweets). Hirshleifer (1977), Diamond and Verrecchia (1981), and Harris and Raviv (1993) propose models to describe how the differential interpretations of information among market participants induces trading. If disagreements on a message board reflect differences in investor opinion, we expect a measure of disagreement to be positively associated with trading volume. With a quick inspection of message content on the Bitcoin forum, we identified points when high proportions of messages contained explicit assertions that the bitcoin's price was likely to rise or fall. We aggregated these opinions into a single measure of bullishness. If, as we have predicted, social media contain new information about the fundamental value of bitcoins, the bullishness measure should relate positively to (future) bitcoin returns. Even if the messages contain no new information, they may capture a general market sentiment and thus be positively correlated with bitcoin returns. Similar to Antweiler and Frank (2004), we consider the simple pairwise correlations of bitcoin market variables with social media variables.

For the dynamic analysis, we adopt a vector autoregression (VAR) system to capture linear interdependencies across time series. We choose the VAR approach rather than a more traditional multiple regression (cf. Antweiler and Frank (2004); Wysocki (1999)) for several reasons. First, with a VAR model, we can treat all of the key variables as jointly endogenous,

20

without creating ad hoc model restrictions. Nor do we need the extensive knowledge about the forces influencing a variable, as required by structural models with simultaneous equations. Second, the model allows for both autocorrelation and cross-correlation, so we can better understand the dynamic relationships among all key variables. For example, in their joint analysis of intraday prices and volume, Gourieroux and Jasiak (2001) explain that an examination of the univariate series of S&P500 returns suggests that past market history is not relevant, whereas a joint analysis of trading volume and returns shows that lagged volume can improve linear predications of future returns. Third, we can interpret the estimated model using Granger causality, impulse response functions, and forecast error variance decomposition. In prior finance literature, VAR models often support portfolio analyses of various risky assets or assessments of domestic and foreign interest rates jointly with foreign exchange rates. In Tumarkin and Whitelaw (2001) application, they use a linear VAR model with one lag to examine the dynamic relationship of daily stock returns, trading volume, Internet message posting volume, and changes in opinions expressed in the messages.

Thus, in our empirical study, we first examine a model in which all four endogenous variables measure bitcoin market activities, namely, returns (r_t), volatility (σ_t^2), transaction volume (V_t^{TX}) and trading volume (V_t). Then we integrate measures of Bitcoin forum activities. The message posting volume and opinions expressed are both potentially relevant for determining bitcoin market activities, so we consider two combinations of variables. The first combines the number of messages expressing positive opinions (POS^F) and the number of messages expressing negative opinions (NEG^F), which captures the level and bullishness of Internet forum activities (see (Luo et al. 2013)). The second combination links the number of messages (M), the bullishness measure, and the agreement index (AI), similar to Antweiler and

Frank (2004) proposal for aggregating message classifications to obtain a single bullishness measure. Let M=POS^F+NEG^F be the total number of value-relevant messages, and R=POS^F/NEG^F be the ratio of bullish to bearish messages. Then our first bullishness measure (bullishness I) is:

$$Bullishness\ I = \frac{POS^F - NEG^F}{POS^F + NEG^F} = \frac{R-1}{R+1}. \tag{3}$$

This measure is independent of the total number of messages (or posting volume) and bound between −1 and 1. The second bullishness measure (bullishness II) is:

$$Bullishness\ II = Ln\left[\frac{1+POS^F}{1+NEG^F}\right] \approx Bullishness\ I * \ln(1 + M). \tag{4}$$

Our third measure (bullishness III) is:

$$Bullishness\ III = POS^F - NEG^F = M * \frac{R-1}{R+1} = M * Bullishness\ I. \tag{5}$$

The latter two measures both increase with the number of messages and with the ratio of bullish to bearish messages. However, only bullishness II discounts excessively large message numbers.

To measure disagreement among forum contributors, we constructed an agreement index (Antweiler and Frank (2004):

$$AI = 1 - \sqrt{1 - bullishness\ I^2}. \tag{6}$$

This index is bound between 0 and 1 and decreases with greater disagreement levels. Suppose all messages are bullish, such that bullishness I equals 1; then it is easy to verify that AI = 1. If all messages are bearish, bullishness I equals −1, and once again AI = 1. If instead two-thirds of the messages are bullish and one-third are bearish, bullishness I is equal to 1/3, AI = .057, and agreement is low, or disagreement is high. If half of the messages are bullish and half are bearish, bullishness I equals 0, AI = 0, and agreement is at its lowest value, or disagreement is highest.

A fourth model examines bitcoin market activities and Twitter activities over short time intervals. The three bitcoin market measures are returns, absolute returns, and trading volume. The two Twitter measures are the number of tweets expressing positive opinion (POS^T) and the number of tweets expressing negative opinion (NEG^T). This last model includes the bitcoin market measures and Twitter activities, as well as Internet forum activities (POS^F, NEG^F, POS^T, NEG^T). We summarize all the models in Table 2.

Table 2: Model Summary

Model	Endogenous Variable	Exogenous Variables	Sample Period	Sampling Frequency
1a	$P_t, \sigma_t^2, V_t, V_t^{TX}$	Yes	1/1/2011-8/18/2014	Daily
1b	$r_t, \sigma_t^2, V_t, V_t^{TX}$	No	4/18/2014-8/18/2014	Hourly
2	$P_t, \sigma_t^2, V_t, V_t^{TX}, POS^F, NEG^F$	Yes	1/1/2011-8/18/2014	Daily
3	$P_t, \sigma_t^2, V_t, V_t^{TX}, M, bullishness, AI$	Yes	1/1/2011-8/18/2014	Daily
4	$r_t, |r_t|, V_t, POS^T, NEG^T$	No	4/18/2014-8/18/2014	Hourly
5a	$P_t, \sigma_t^2, V_t, V_t^{TX}, POS^F, NEG^F, POS^T, NEG^T$	Yes	4/18/2014-8/18/2014	Daily
5b	$r_t, |r_t|, V_t, POS^F, NEG^F, POS^T, NEG^T,$	No	4/18/2014-8/18/2014	Hourly

To determine an appropriate VAR system, we first test the stationarity of the variables. Conventional regression estimators, including VAR, encounter problems when applied to nonstationary processes, such that the regression of two independent random walk processes would yield a spurious significant coefficient, even if they were not related (Granger and Newbold 1974). We used an augmented Dickey-Fuller unit root test of each variables, with lag numbers chosen according to the Schwert (1989) rule. Among the time series in the model, the number of positive posts, S&P500 returns, and NASDAQ returns are stationary; the others have one order of integration.

Next, we determined the appropriate lag length *p* using Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), as is standard in the VAR literature (Love and Zicchino (2006)). For each model, we calculated the AIC and BIC values for the sample period

and chose the lag length that minimized both criteria. If they indicated conflicting optimal lag

lengths, we chose the length that minimized the BIC.

Although in a VAR system, we could model the interrelationship of the variables by taking

first differences of each non-stationary series and including the differences in a VAR, this

approach can suffer misspecification biases if cointegration is present. In that case, VAR

expresses only the short-run responses between variables, without providing information about

the long-run equilibrium in the case of cointegration between two or more series. We performed

a Johansen test (Johansen and Juselius 1990) and confirmed the presence of cointegration in our

daily frequency data. Therefore, in Models 1a, 2, 3, and 5a, we extended the VAR model to an

vector error correction model (VECM), which can fit the first differences of the non-stationary

variables, using a vector of error correction terms that is equal in length to the number of

cointegrating relationships added to the relationship (see Johansen (1995)). By taking potential

long-term relationships into account, the VECM model with p variables, k lags, and

cointegration rank r has the following form:

$$\Delta \boldsymbol{Y}_t = \sum_{j=1}^{k-1} \Gamma_j \Delta \boldsymbol{Y}_{t-k} + \alpha \beta' \boldsymbol{Y}_{t-1} + \boldsymbol{\mu} + \epsilon_t \tag{7}$$

where Δ is the first difference operator, $\boldsymbol{Y}_t$ is a $p \times 1$ vector with order of integration 1 , $\boldsymbol{\mu}$ is

a $p \times 1$ constant vector representing the linear trend, k is the a lag structure, and ϵ is the residual

vector. Γ_j is a $p \times p$ matrix that indicates short-term relationships among variables, β is a $p \times r$

matrix that represents the long-term relationships between the cointegrating vectors, and α is a

$p \times r$ matrix denoting the speed of variables adjusting to the long term equilibriums. The

difference between the VECM model and the VAR model with first differenced variables is the

additional $\beta' Y_{t-1}$, known as the error correction term. In addition, in the VECM, the non-

differenced variables (i.e., price (P_t) instead of return (r_t), number of posts instead of increase of posts) are used in the estimation because the model itself has first differences built in. Note that the VECM model is a special case of the general VAR system as it can be expressed as an equivalent VAR:

$$Y_t = (I_k + \alpha\beta' + \Gamma_1)Y_{t-1} + \sum_{j=2}^{k-1}(\Gamma_j - \Gamma_{j-1})Y_{t-j} + \Gamma_{k-1}Y_{t-k} + \mu + \epsilon_t \qquad (8)$$

where I_k is a $k \times k$ identity matrix. In addition, when the variables are not cointegrated, Eq. (7) reduces to a stationary VAR($k-1$) model in first differences:

$$\Delta Y_t = \sum_{j=1}^{k-1}\Gamma_j\Delta Y_{t-1} + \mu + \epsilon_t \qquad (9)$$

RESULTS

Contemporaneous Relation

Table 3 contains the simple pairwise correlations that differ significantly from zero. The correlations at the top of the table (rows 1-3) are among the bitcoin market variables and appear consistent with well-known, stylized facts pertaining to the stock market. For example, the correlation between returns and trading volume is positive but small, but the correlation between volatility and trading volume is large (.369 for daily frequency). As a comparison, the correlation between stock trading volume and volatility is only .063 according to (Antweiler and Frank 2004). Thus, bitcoin trading volume and transaction volume appear highly correlated, at .732. Seemingly, people who actively participate in bitcoin exchange trading are more likely to use bitcoin as currency and carry out bitcoin transactions simultaneously.

The correlations between the bitcoin market measures and the Internet forum measures are large, with magnitudes comparable to the strong correlations of different market measures. For example, the number of relevant messages correlates at .312 with volatility, at .556 with transaction volume, and at .445 with trading volume. Because the number of posted messages is highly autocorrelated, its one-day lag value also correlates strongly with bitcoin volatility, transaction volume, and trading volume. This result confirms Wysocki (1999) finding regarding the relationship between message board activities and stock market activities: Daily variation in real market activity and information events can explain much of the variation in daily change in message postings. Because the magnitude of the relationships between bitcoin market measures and message board posting volume are not trivial, we conclude that social media reflect market information rapidly.

The midsection of Table 3 (rows 6-11) contains the correlations of various bullishness measures with the bitcoin market measures. Bullishness III correlates positively with contemporaneous returns and negatively with volatility and volume. According to the agreement index, disagreement among Internet forum contributors is associated with more trading. Finally, the bottom section of Table 3 reveals the strong correlations between traditional Internet activity measures and bitcoin market measures; they also are highly autocorrelated over time.

Stock returns are famously difficult to predict; with regard to bitcoin returns. Table 3 indicates a significant but negative contemporaneous correlation between social media posting volume and bitcoin returns. The economic significance is small, similar to the findings of (Antweiler and Frank 2004). Although we find a significant, positive, contemporaneous correlation between the third bullishness measure and returns, this magnitude also is small.

Table 3: Pairwise Correlations

	Return	Volatility	Transaction Volume	Trading Volume
Volatility				
Transaction volume		0.420		
Trading volume	0.084	0.369	0.732	
# Post	-0.066	0.312	0.556	0.445
# Post, 1d lag		0.311	0.531	0.425
Bullishness I			-0.113	-0.096
Bullishness I, 1d lag			-0.110	-0.076
Bullishness II			-0.112	-0.093
Bullishness II, 1d lag			-0.108	-0.072
Bullishness III	0.078	-0.252	-0.424	-0.312
Bullishness III, 1d lag		-0.248	-0.402	-0.280
Agreement				-0.056
Agreement, 1d lag				-0.072
Rank		-0.251	-0.399	-0.268
Rank, 1d lag			-0.245	-0.399
Reach		0.329	0.543	0.408
Reach, 1d lag		0.334	0.540	0.401
Page view	0.068	0.477	0.106	-0.076
Page view, 1 d lag		0.472	0.119	-0.066
Google trend		0.165	0.318	0.276
Google trend, 1d lag		0.163	0.323	0.272

Notes: This table contains only correlations that are significantly different from 0 at a 95% confidence level. The time period is one day. The four market variables are the log difference in bitcoin exchange rate (in USD) from the end of the previous day to the current day (return); daily volatility estimated with the EWMA model (volatility); the natural log of bitcoin daily transaction volume in USD, with the linear time trend removed (transaction volume); and the natural of log bitcoin daily trading volume in USD, with the linear time trend removed (trading volume).

Dynamic Relationships among Bitcoin Returns, Volatility, Trading Volume, and Transaction Volume

We provide the *t*-statistics for each coefficient of the first lagged variables (Φ_1) in Model 1 (*a* and *b*) in Table 4. We report *t*-statistics instead of coefficients to avoid interpretation issues associated with scale effects. The autoregressive coefficients are on the diagonal; all four coefficients are positive and significant at the 1% level. At a daily frequency (Panel a), trading volume and transaction volume exhibit a strong autoregressive relationship. That is, high trading (transaction) volume days tend to precede days of high trading (transaction) volume. Other significant *t*-statistics are in bold, though the magnitudes for the majority of coefficients are small. Other than past returns, no market variables indicate predictive power for bitcoin returns

(negative direction). Trading volume helps predict volatility, consistent with findings from stock

market studies. We also note some interesting results for the bitcoin transaction volume variable:

The higher the returns and trading volume today, the higher the transaction volume tomorrow;

but the higher the volatility today, the lower the transaction volume tomorrow. The performance

of bitcoin, as an investment asset, thus appears to affect its popularity as a currency.

Panel b reports the t-statistics for the coefficients at an hourly frequency. We lack hourly

transaction volume data, so the vector of variables of interest is reduced to $(r_t, \sigma_t^2, V_t)'$. Both

returns and volatility show positive autoregressive relations, whereas trading volume indicates a

strong negative autoregressive relationship. The signs of t-statistics are generally similar to those

in Panel a.

Table 4: t-Statistics from Model 1 *a* and *b*

Panel a: Daily Frequency (January 1, 2011–August 23, 2014, VECM)

Dependent Variable	Independent Variable			
	ΔP_{t-1}	$\Delta\sigma_{t-1}^2$	ΔV_{t-1}	ΔV_{t-1}^{TX}
$\Delta P_t,$	7.10***	-0.23	-1.06	0.05
$\Delta\sigma_t^2$	0.87	0.36	4.78***	-1.50
ΔV_t	-0.27	-1.60	-13.10***	-0.96
ΔV_t^{TX}	2.11**	-1.12	1.31	-14.67***

Panel b: Hourly Frequency (April 18, 2014–August 18, 2014, VAR)

Dependent Variable	Independent Variable		
	r_{t-1}	$|r_t|$	ΔV_{t-1}
r_t	8.57***	-0.35	2.56***
$|r_t|$	-0.68	11.73***	4.38***
ΔV_t	1.94**	-3.54***	-22.88***

Notes: Variables are as defined in Table 1.

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

Dynamic Relationships among Bitcoin Activity and Forum Activity

We examine the predictive power of the number and bullishness of forum messages posted during the whole sample period using Models 2 and 3 (Table 5). The *t*-statistics for the market variables $r_t, \sigma_t^2, V_t, V_t^{TX}$ are similar to those in Table 4, suggesting little multicollinearity between market variables and forum variables. The two forum metrics generally work as we anticipated. Days with unexpected increases in the number of positive (bullish) posts tend to precede days with high bitcoin returns and high transaction volume. Days with unexpected increases in the number of negative (bearish) posts tend to precede days with lower bitcoin returns and lower transaction volume. All these relationships are statistically significant. Panel b shows the results from the same model, using only messages posted by the silent majority of users (excluding top 5%). In this case, the predictive power of posting volume for bitcoin returns becomes stronger, but its power in terms of predicting transaction volume diminishes. Panel c contains the subsample results, which reveal that the signs of the t-statistics of POS^F and NEG^F are consistent across all three years, but the relative predictive power varies from year to year. Therefore, it is possible that forum posts contain new information about the value of bitcoin. It is also possible that the posts contain no new information but still provide a better indication of general market sentiment than is already contained in the trading record.

With Panel d, we examine the predictive power of the posts from the vocal minority (top 5%). In contrast with the results for the silent majority, these posts are strong indicators of future transaction activities. Specifically, an increase in the number of bearish posts by the vocal minority predicts lower future transaction volume, and an increase in the number of bullish posts predicts higher future transaction volume. These users likely are also active traders on the bitcoin market. Moreover, the number of messages posted by the vocal minority is more likely to depend

on market activity: Higher volatility predicts fewer posts on the next day, whereas higher

transaction volume predicts more posts on the next day.

Table 5: t-Statistics from VECM Analysis (Model 2)

Panel a: Full Sample Period, All Users

Dependent Variable	Independent Variable					
	ΔP_{t-1}	$\Delta \sigma^2_{t-1}$	ΔV_{t-1}	ΔV^{TX}_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
ΔP_t	6.66***	-0.09	-0.94	-0.12	1.94*	-1.81*
$\Delta \sigma^2_t$	0.77	0.94	1.40	1.86**	1.03	-0.85
ΔV_t	-0.05	-1.11	-6.94***	-2.54***	-0.01	0.30
ΔV^{TX}_t	1.99	-0.82	0.42	-10.95***	2.41***	-1.27
ΔPOS^F_t	0.05	-1.88*	-2.12**	2.80***	-3.23***	-1.14
ΔNEG^F_t	-0.68	-2.08**	-2.26**	2.71***	0.86	-3.67***

Panel b: Full Sample Period, Excluding Top 5% of Users

Dependent Variable	Independent Variable					
	ΔP_{t-1}	$\Delta \sigma^2_{t-1}$	ΔV_{t-1}	ΔV^{TX}_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
ΔP_t	6.44***	0.04	-0.78	-0.13	2.31**	-2.21**
$\Delta \sigma^2_t$	1.16	0.89	1.46	1.91*	-0.54	0.33
ΔV_t	0.00	-1.11	-7.04***	-2.40**	-0.53	0.54
ΔV^{TX}_t	2.26**	-0.87	0.57	-10.93***	0.32	0.71
ΔPOS^F_t	-0.54	0.60	-1.53	1.07	-2.94***	-0.74
ΔNEG^F_t	-0.71	-0.75	-1.48	0.74	0.10	-3.78***

Panel c: Subsample Periods, Excluding Top 5% of Users

Dependent Variable	Independent Variable					
	2011		2012		2013	
	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
ΔP_t	2.57***	-2.01**	0.26	-0.99	0.66	-1.77*
$\Delta \sigma^2_t$	-1.24	0.97	-1.24	-0.12	-0.58	-0.09

Panel d: Full Sample Period, Only Top 5% of Users

Dependent Variable	Independent Variable					
	ΔP_{t-1}	$\Delta \sigma^2_{t-1}$	ΔV_{t-1}	ΔV^{TX}_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
ΔP_t	6.79***	-0.12	-0.91	-0.15	1.54	-1.38
$\Delta \sigma^2_t$	0.72	0.96	1.35	1.86*	1.32	-1.09
ΔV_t	-0.08	-1.04	-6.90***	-2.55***	-0.03	0.28
ΔV^{TX}_t	2.01	-0.75	0.38	-10.92***	2.64***	-1.58
ΔPOS^F_t	0.14	-2.30**	-1.99*	3.04***	-4.06***	-1.54
ΔNEG^F_t	-0.69	-2.24**	-2.31**	3.11***	0.34	-3.95***

Notes: The full sample period is 1/1/2011–8/18/2014. The variables are as defined in Table 1.

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

Model 3 contains an alternative set of Internet forum measures that capture posting volume and bullishness, as well as disagreement among contributors. As we show in Table 6, bitcoin returns are positively associated with the previous day's bullishness, but the relationships are not statistically significant. As we predicted, disagreement induces trading; the agreement index is negatively associated with future trading volume. The agreement index emerges as the only variable derived from the forum activities that exhibits significant predictive power for future trading volume.

Table 6: t-Statistics from VECM Analysis (Model 3)

Dependent Variable	Independent Variable						
	ΔP_{t-1}	$\Delta \sigma^2_{t-1}$	ΔV_{t-1}	ΔV^{TX}_{t-1}	M	B III	AI
ΔP_t	6.85***	0.30	-0.93	-0.06	-0.21	1.18	-0.14
ΔV_t	0.03	-1.69	-7.46***	-2.23**	-0.44	-0.69	-2.35**

Notes: This table reports on the full sample period, 1/1/2011–8/18/2014, excluding the top 5% of users. The variables are as defined in Table 1.

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

Dynamic Relationships among Bitcoin Activity and Twitter Activity

We examine the predictive power of the number and bullishness of all tweets over a four-month sample period (April 18–August 18, 2014), as summarized in Table 7 (Panel a), using

hourly sampling frequency. The magnitudes of the *t*-statistics among the market variables are similar to those reported in Table 4 (Panel b), again indicating low correlations between the market variables and social media variables. Changes in POS^T and NEG^T indicate strong negative autoregressive relationships. In terms of predicting bitcoin returns, volatility, and trading volume, both variables are mostly useless since the t-statistics in the last two columns of Table 7 are insignificant, with signs opposite our expectations. For example, more bearish tweets indicate higher future returns and trading volume.

Table 7: t-Statistics from VAR Analysis (Model 4)

Panel a: All Tweets, Hourly Frequency (April 18–August 18)

Dependent Variable	Independent Variable				
	r_{t-1}	$|r_t|$	ΔV_{t-1}	ΔPOS^T_{t-1}	ΔNEG^T_{t-1}
$r_t,$	8.57***	-0.31	2.50**	0.69	0.27
$|r_t|$	-0.72	11.76***	4.30***	0.69	0.77
ΔV_t	1.85*	-3.49***	-23.00***	1.34	1.81*
ΔPOS^T_t	-1.76*	2.37**	1.65	-30.17***	2.74***
ΔNEG^T_t	-0.25	-0.24	1.91*	0.51	-25.45***

Panel b: Tweets by Opinion Leaders, Hourly Frequency

Dependent Variable	Independent Variable				
	r_{t-1}	$|r_t|$	V_{t-1}	$\Delta top\ 20\_POS^T_{t-1}$	$\Delta top\ 20\_NEG^T_{t-1}$
$r_t,$	8.58***	-0.36	2.53**	2.05**	-0.44
$|r_t|$	-0.71	11.71***	4.47***	0.33	1.88*
ΔV_t	1.92*	-3.58***	-22.80***	0.34	0.47
$top\ 20\_POS^T_t$	-1.30	-1.15	1.16	-35.31***	-1.11
$top\ 20\_NEG^T_t$	1.03	-0.20	-0.87	-0.36	-1.48

Notes: The variables are as defined in Table 1.

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

To determine if incorporating a social influence analysis improves the predictive power of the tweet metrics, we replaced POS^T and NEG^T (all tweets) with the $top\_20\_POS^T$ and $top\_20\_NEG^T$ (tweets by people who rank in the top 20 in terms of follower counts (Shi et al.

2014)), which is similar to using weighted opinions instead of simple aggregated opinions. The results for these opinion leaders appear in Panel b and reveal that though they remain insignificant, the t-statistics move mostly in the expected directions. Hours with many bullish (bearish) tweets by opinion leaders precede hours with high (low) bitcoin returns, and the predictive relationship between bullish tweets and bitcoin returns is statistically significant at 5%. We define opinion leaders as those who rank among the top 20 in terms of follower counts, following the practice in (Shi et al. 2014). In a robustness check, we ran analyses with the 50 and 100 top ranked opinion leaders and found quantitatively similar (but weaker) results.

The Relative Predictive Power of Internet Forum versus Twitter Variables

To examine the relative predictive power of the Internet forum and Twitter variables, we used a VECM model with eight endogenous variables (Model 5a) for daily data and a VAR model with seven endogenous variables (Model 5b) for hourly data, and the results are reported in Table 8. At a daily frequency (Panel a), the forum variables predict bitcoin returns one day in the future, and the relations are statistically significant. Days with many bullish and few bearish forum posts precede days marked by higher bitcoin returns. In contrast, the Twitter variables have negligible predictive power for bitcoin returns. No social media variables exhibit significant predictive power for volatility or trading volume during the sample period, though days with more bullish tweets precede days with high bitcoin transaction volume.

Panel b of Table 8 presents the results with hourly data. For the Twitter data, we limit our sample to tweets by top 20 opinion leaders. Hours with a greater number of bullish tweets precede hours with higher bitcoin returns; this relationship is statistically significant. Hours marked by increases in the number of bearish tweets precede hours with lower bitcoin returns, but this relationship is insignificant. Forum variables instead have no predictive power for hourly

33

bitcoin returns. Changes in the hourly tweet volume and forum posting volume both predict

future bitcoin volatility. In particular, more bearish tweets and more bullish messages each

precede hours in which we find higher bitcoin absolute returns, in statistically significant

relations. These findings are consistent with our hypotheses.

Table 8: Forum versus Twitter (Model 5)

Panel a: Daily Frequency (April 18–August 14, VECM)

Dependent Variable	Independent Variable			
	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}	ΔPOS^T_{t-1}	ΔNEG^T_{t-1}
$\Delta P,$	1.72*	-2.13**	-1.47	0.62
$\Delta \sigma^2_t$	0.15	0.19	-0.01	-1.23
ΔV_t	0.23	0.25	0.84	-0.76
ΔV^{TX}_t	-0.43	0.85	2.08**	0.65

Panel b: Hourly Frequency (April 18–August 14, VAR)

Dependent Variable	Independent Variable			
	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}	ΔPOS^T_{t-1}	ΔNEG^T_{t-1}
$r_t,$	-0.16	0.55	1.77*	-0.41
$|r_t|$	2.64***	-1.55	-0.03	2.15**
ΔV_t	2.41**	0.24	0.31	0.29

Notes: The variables are as defined in Table 1.

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

Robustness Checks

Our primary motivation for choosing VECM over standard VAR models is to avoid the

potential misspecification bias that can arise due to the inability of VAR to deal with long-term

relationships. As a robustness check, we fit a VAR model for Model 2, using the same variables,

and took the first-order difference of POS^F and NEG^F to keep the variables stationary. As the

results in Table 9 indicate, we obtained similar results: An increase in the amount of positive and

negative messages posted by the bottom 95% of users had predictive power for bitcoin returns,

and an increase in posts from more active users offered more predictive power with respect to transaction volume.

Table 9: t-Statistics from VAR Analysis (Model 2)

Panel a: Bottom 95% of Users

Dependent Variable	Independent Variable					
	r_{t-1}	σ^2_{t-1}	V_{t-1}	V^{TX}_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
$r_t,$	4.38	-0.47	-3.68	0.52	1.47	-1.35
σ^2_t	17.07	34.07	2.93	2.28	-0.69	2.13**
ΔV_t	-0.07	-1.85	17.02	2.90	1.01	-1.07
ΔV^{TX}_t	3.58	-2.47	3.35	19.44	per	0.48
ΔPOS^F_t	2.14**	-0.49	0.35	1.55	-14.70	4.56
ΔNEG^F_t	0.52	-2.00**	0.57	-0.24	5.07	-9.32

Panel b: Top 5% of Users

Dependent Variable	Independent Variable					
	r_{t-1}	σ^2_{t-1}	V_{t-1}	V^{TX}_{t-1}	ΔPOS^F_{t-1}	ΔNEG^F_{t-1}
$r_t,$	4.35	-0.34	-3.63	0.38	1.17	-0.77
σ^2_t	16.85	37.06	2.95	2.37	0.89	0.07
ΔV_t	-0.20	-1.73	17.10	2.92	1.10	-1.81*
ΔV^{TX}_t	3.38	-2.55	3.52	19.53	1.88*	-1.94**
ΔPOS^F_t	0.80	-2.77**	-0.33	3.41***	-14.70	4.56
ΔNEG^F_t	0.91	-2.47**	0.01	1.93**	5.07	-9.32

Notes: Variables are as defined in Table 1. ***, **, * denote significance at 1%, 5%, and 10%, respectively.

DISCUSSION AND CONCLUSION

This study investigates the predictive power of social media metrics for bitcoin returns, as well as their dynamic relationships. The results suggest that social media is an important indicator of future bitcoin returns. First, a positive shock of bullish posting predicts positive bitcoin returns on the next day, and a positive shock of bearish posting predicts negative returns on the next day (H2a). These effects are statistically significant. Second, disagreement induces trading: Greater disagreement across messages predicts higher bitcoin exchange trading volume

on the next day (H2b). Third, message posting helps predict bitcoin transaction volume. From an investment perspective, Internet forum posts thus can be viewed as bitcoin popularity indicators, offering investors a means to predict short-run price movements and thereby front-run the market.

Our analysis of the dynamic relationships between social media and bitcoin returns also indicates some interesting differences at the social influence level and specifies the unique effects of different social media platform (Internet forum vs. microblogging site). In line with H3, we note that bitcoin market predictions can gain accuracy if they integrate an influencer analysis and stratify the sample by social media user characteristics. For example, the numbers of bullish and bearish tweets by all users have negligible predictive power for bitcoin returns in the next hour. However, if we limit the sample to tweets from those users with the most followers, the predictive relationship becomes statistically significant. Posts from the most active Internet forum contributors also have stronger associations with bitcoin transaction volume on the next day than do posts from ordinary users. This finding verifies that follow-the-influencer behavior exists in the bitcoin market, extending prior finance and marketing findings to a new setting (Brown et al. 2009; Hinz et al. 2011; Jiao and Ye 2013; Libai et al. 2010). To leverage the effects of social influence on bitcoin markets, investors should actively follow the most influential people in a social network to collect market information much more efficiently. However, the power of the silent majority should not be ignored, as we show that their sentiments can be the more important metric in predicting the movement of future prices.

The relative predictive power of forum posts and tweets, using both daily and hourly data, also are distinct (H4). Forum metrics do a better job of predicting future bitcoin returns at a daily frequency; Twitter metrics outperform them at an hourly frequency. A potential explanation is

36

that the two types of social media platforms differ in their characteristics, namely, as information sharing platforms and in how users receive information. The predominantly mobile uses of Twitter make tweets more visible to investors at an intraday level, and they therefore respond in a timelier fashion. Internet forum members instead engage in more thorough discussions, whether positive or negative, which leads to consolidated market reactions at the daily level.

From a practical perspective, our results offer insights into the bitcoin economy. First, Bitcoin's algorithm will cap the number of ultimately available bitcoins at 201 million (to date, fewer than 13 million bitcoins have been mined worldwide), yet price fluctuations and the potential presence of a bubble remain key concerns of early adopters and investors. Public opinions about the real value of bitcoins also diverge, reflecting the different arguments regarding whether the use of Bitcoin as a payment network necessarily increases the value of bitcoins. We show that the price of bitcoins relates closely to communities' sentiments about their future value, similar to the situation for stocks and firm equities. This finding suggests evidence of the investment value of bitcoin, as a financial instrument.

Second, we recommend that companies strategically and carefully evaluate their decision to adopt the Bitcoin payment system. The decision must involve more than the marketing consideration of the potential for generating positive buzz, since the dynamic relationship between social media content and bitcoin value means the future value of the accounts receivable can also be affected.

Third, the predictability of bitcoin returns should strengthen their reliability as a regular medium of exchange. In addition to its unique benefits (i.e., lower transaction costs, potential to combat poverty and oppression, stimulus for financial innovation; (Brito and Castillo 2013)), "Bitcoin does not present a threat to economic activity by disrupting traditional channels of

commerce. Instead, its global transmissibility opens new markets to merchants and service providers" (Federal Advisory Council and Board of Governors (2014)). Our research presents more positive evidence that government agencies and judicial systems can use to weigh and balance their restraint or encouragement of this unprecedented financial innovation.

The implications of our results also go beyond the bitcoin market specifically. Even as research attends more to the business impacts of social media and appropriate analytical methods for dealing with big data, most studies continue to use single data sources. Yet as our study shows, not all social media content is created equally; it can have varying impacts. Social media users also exhibit distinct contribution frequency and behaviors. Different social media platforms retain information differently and result in varying interaction behaviors among users. Therefore, researchers should take particular care before aggregating or sampling large volume of social media data across users and platforms. Models that support analyses of these data through finer-grained approaches could yield interesting results.

Our research also has several limitations. We collected considerable social media data from Bitcoin, but we could quantify the impact of Twitter messages on the bitcoin market for only a four-month period. This time period does not represent a limitation for the hourly analysis, but our conclusions for the daily analysis may suffer from low statistical power. Investigations of a longer period would be beneficial. Our findings highlight the dynamics between social media and Bitcoin, a single cryptocurrency; whether they generalize to other virtual currencies remains unclear. Finally, we collected social media data from an English-language Internet forum and limited our Twitter crawl to messages in English too. Yet the bitcoin market contains adopters, investors, and speculators worldwide, so a more comprehensive study is needed to draw conclusions about UGC written in other languages.

In turn, we suggest several extensions to this study. First, our analyses show that different types of social media and different groups of users have distinct impacts on the bitcoin market, so more insights could accrue from investigations of the root causes of these differences. For example, are the different impacts due mainly to content (i.e., subjectivity of opinions) or to behavioral reasons (i.e., posting frequency, interactions with other users)? Research that addresses these questions could enhance understanding of the mechanisms of the feedback loops between social media and financial markets. Second, we used financial sentiments as the sole indicators provided in social media. To identify other features that might contribute to the dynamics, further linguistic analyses and statistical topic modeling might be conducted on the textual data. For example, many Bitcoin users worry about its security, so additional studies might investigate the implications of sudden surges of discussions about security or privacy, as well as users' opinions about these topics.

The cryptocurrency market is gaining momentum. An estimated 275 digital currencies are now in play, many of which cater to specific products or services (Mayer 2014). Whereas most prior research addresses technical issues (e.g., mining, security, privacy), the success of online payment systems relies on the credible demonstration of their business, economic, and societal value. We believe the IS community should take the lead in this emerging, interdisciplinary research area, to span information technology, marketing, finance, and social science. By revealing the association between two big data phenomena (virtual currency and social media), this study provides an initial examination of the predictive relationship between social media and bitcoin returns. It thus sheds some new light on the impact of big data and analytics on a disruptive transformation in the networked world.

REFERENCES

Andersen, T. G. 1996. "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *The Journal of Finance* (51:1), pp. 169-204.

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), pp. 1259-1294.

Barber, B. M., and Odean, T. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies* (21:2), pp. 785-818.

Barber, S., Boyen, X., Shi, E., and Uzun, E. 2012. "Bitter to Better—How to Make Bitcoin a Better Currency," in *Financial Cryptography and Data Security*. Springer, pp. 399-414.

Bird, S. 2006. "Nltk: The Natural Language Toolkit," in: *The COLING/ACL on Interactive Presentation Sessions*. Sydney, Australia: pp. 69-72.

Bonchi, F., Castillo, C., Gionis, A., and Jaimes, A. 2011. "Social Network Analysis and Mining for Business Applications," *ACM Transactions on Intelligent Systems and Technology* (2:3), p. 22.

Brito, J., and Castillo, A. 2013. *Bitcoin: A Primer for Policymakers*. Mercatus Center at George Mason University.

Brown, J. R., Fazzari, S. M., and Petersen, B. C. 2009. "Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom," *The Journal of Finance* (64:1), pp. 151-185.

Brown, N. C., Wei, K. D., and Wermers, R. 2013. "Analyst Recommendations, Mutual Fund Herding, and Overreaction in Stock Prices," *Management Science* (60:1), pp. 1-20.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media," *Review of Financial Studies* (27:5), pp. 1367-1403.

Chen, Y., and Xie, J. 2008. "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science* (54:3), pp. 477-491.

Cheung, C. M., and Thadani, D. R. 2012. "The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model," *Decision Support Systems* (54:1), pp. 461-470.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of marketing research* (43:3), pp. 345-354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. 2010. "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science* (29:5), pp. 944-957.

Ciaian, P., Rajcaniova, M., and Kancs, d. A. 2014. "The Economics of Bitcoin Price Formation," in: *ArXiv* 1405.4498. Available at http://adsabs.harvard.edu/abs/2014arXiv1405.4498C

Das, S. R., and Chen, M. Y. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* (53:9), pp. 1375-1388.

Dellarocas, C., Zhang, X. M., and Awad, N. F. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing* (21:4), pp. 23-45.

Dewan, S., and Ramaprasad, J. 2012. "Music Blogging, Online Sampling, and the Long Tail," *Information Systems Research* (23:3-part-2), pp. 1056-1067.

Dhar, V., and Chang, E. A. 2009. "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing* (23:4), pp. 300-307.

Diamond, D. W., and Verrecchia, R. E. 1981. "Information Aggregation in a Noisy Rational Expectations Economy," *Journal of Financial Economics* (9:3), pp. 221-235.

Duan, W., Gu, B., and Whinston, A. B. 2008a. "Do Online Reviews Matter?—an Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), pp. 1007-1016.

Duan, W., Gu, B., and Whinston, A. B. 2008b. "The Dynamics of Online Word-of-Mouth and Product Sales—an Empirical Investigation of the Movie Industry," *Journal of retailing* (84:2), pp. 233-242.

Duggan, M., and Brenner, J. 2013. "The Demographics of Social Media Users, 2012," Pew Research Center's Internet & American Life Project, Washington, DC. Available at http://www.lateledipenelope.it/public/513cbff2daf54.pdf

European Banking Authority. 2014. "Eba Opinion on Virtual Currencies." from http://www.eba.europa.eu/documents/10180/657547/EBA-Op-2014-08+Opinion+on+Virtual+Currencies.pdf

European Central Bank. 2012. "Virtual Currency Schemes." from http://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work*," *The journal of Finance* (25:2), pp. 383-417.

Federal Advisory Council and Board of Governors. 2014. "Record of Meeting on May 9th, 2014." from http://fortunedotcom.files.wordpress.com/2014/05/fac-20140513.pdf

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Friend, I., Blume, M., Crockett, J., and Fund, T. C. 1970. *Mutual Funds and Other Institutional Investors: A New Perspective*. McGraw-Hill.

Gallaugher, J., and Ransbotham, S. 2010. "Social Media and Customer Dialog Management at Starbucks," *MIS Quarterly Executive* (9:4), pp. 197-212.

Ghose, A., and Yang, S. 2009. "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science* (55:10), pp. 1605-1622.

Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-560.

Godes, D., and Mayzlin, D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science* (28:4), pp. 721-739.

Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W. 2009. "The Role of Hubs in the Adoption Process," *Journal of Marketing* (73:2), pp. 1-13.

Gourieroux, C., and Jasiak, J. 2001. *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press Princeton, NJ.

Granger, C. W., and Newbold, P. 1974. "Spurious Regressions in Econometrics," *Journal of econometrics* (2:2), pp. 111-120.

Grinberg, R. 2011. "Bitcoin: An Innovative Alternative Digital Currency," *Hastings Science & Technology Law Journal* (4), pp. 160-208.

Gu, B., Park, J., and Konana, P. 2012. "Research Note-the Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products," *Information Systems Research* (23:1), pp. 182-196.

Harris, M., and Raviv, A. 1993. "Differences of Opinion Make a Horse Race," *Review of Financial studies* (6:3), pp. 473-506.

Helbing, D. 2014. "The World after Big Data: What the Digital Revolution Means for Us," in: *SSRN* 2438957. Available at http://ssrn.com/abstract=2438957

Hinz, O., Skiera, B., Barrot, C., and Becker, J. U. 2011. "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing* (75:6), pp. 55-71.

Hirshleifer, D., and Teoh, S. H. 2009. "Thought and Behavior Contagion in Capital Markets," in *Handbook of Financial Markets: Dynamics and Evolution*. pp. 1-46.

Hirshleifer, J. 1977. "The Theory of Speculation under Alternative Regimes of Markets," *The Journal of Finance* (32:4), pp. 975-999.

Jiao, Y., and Ye, P. 2013. "Mutual Fund Herding in Response to Hedge Fund Herding and the Impacts on Stock Prices." *Working Paper*, Available at http://faculty.ucr.edu/~yawenj/mfhf.pdf

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johansen, S., and Juselius, K. 1990. "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money," *Oxford Bulletin of Economics and statistics* (52:2), pp. 169-210.

Kamins, M. A., and Assael, H. 1987. "Two-Sided Versus One-Sided Appeals: A Cognitive Perspective on Argumentation, Source Derogation, and the Effect of Disconfirming Trial on Belief Change," *Journal of Marketing Research* (24:1), pp. 29-39.

Kristoufek, L. 2013. "Bitcoin Meets Google Trends and Wikipedia: Quantifying the Relationship between Phenomena of the Internet Era," *Scientific Reports* (3).

Kristoufek, L. 2014. "What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis," in: *arXiv* 1406.0268v1. Available at http://arxiv.org/pdf/1406.0268v1.pdf

Kroll, J. A., Davey, I. C., and Felten, E. W. 2013. "The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries," in: *The Twelfth Workshop on the Economics of Information Security (WEIS 2013)*. Washington, DC.

Libai, B., Bolton, R., Bügel, M. S., De Ruyter, K., Götz, O., Risselada, H., and Stephen, A. T. 2010. "Customer-to-Customer Interactions: Broadening the Scope of Word of Mouth Research," *Journal of Service Research* (13:3), pp. 267-282.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of marketing* (70:3), pp. 74-89.

Loughran, T., and McDonald, B. 2014. "Measuring Readability in Financial Disclosures," *The Journal of Finance* (69:4), pp. 1643-1671.

Love, I., and Zicchino, L. 2006. "Financial Development and Dynamic Investment Behavior: Evidence from Panel Var," *The Quarterly Review of Economics and Finance* (46:2), pp. 190-210.

Luo, X. 2007. "Consumer Negative Voice and Firm-Idiosyncratic Stock Returns," *Journal of Marketing* (71:3), pp. 75-88.

Luo, X. 2009. "Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices," *Marketing Science* (28:1), pp. 148-165.

Luo, X., Zhang, J., and Duan, W. 2013. "Social Media and Firm Equity Value," *Information Systems Research* (24:1), pp. 146-163.

Mayer, A. 2014. "Bitcoin Has a Future, but Maybe Not as a Currency." CBCNews. Retrieved on Sep 14, 2014, from http://www.cbc.ca/news/technology/bitcoin-has-a-future-but-maybe-not-as-a-currency-1.2686045

McNulty, E. 2014. "Can We Predict the Future Value of Virtual Currency?" Retrieved on Aug 27, 2014, from http://dataconomy.com/bitcoin-big-data-can-predict-future-value-virtual-currency/

Mehra, A., Dixon, A. L., Brass, D. J., and Robertson, B. 2006. "The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation," *Organization science* (17:1), pp. 64-79.

Metaxas, P. T., and Mustafaraj, E. 2012. "Trails of Trustworthiness in Real-Time Streams (Extended Summary)," in: *Design, Influence and Social Technologies Workshop (DIST*

2012) at the 2012 ACM Conference on Computer Supported Cooperative Work. Seattle, WA.

Moe, W. W., and Fader, P. S. 2004. "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science* (50:3), pp. 326-335.

Moore, T., and Christin, N. 2013. "Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk," in *Financial Cryptography and Data Security*. Springer, pp. 25-33.

Mustafaraj, E., Finn, S., Whitlock, C., and Metaxas, P. T. 2011. "Vocal Minority Versus Silent Majority: Discovering the Opionions of the Long Tail," in: *2011 IEEE International Conference on Privacy, Security, Risk and trust and 2011 IEEE International Conference on Social Computing (PASSAT/SocialCom 2011)*. Boston, MA: IEEE, pp. 103-110.

Network Financial Crimes Enforcement. 2013. "Application of Fincen's Regulations to Persons Administering, Exchanging, or Using Virtual Currencies." from http://fincen.gov/statutes_regs/guidance/html/FIN-2013-G001.html

PricewaterhouseCoopers. 2014. "Digital Disruptor: How Bitcoin Is Driving Digital Innovation in Entertainment, Media and Communications ". Available at http://www.pwc.com/en_SG/sg/tmt/assets/tmtnews201402/digital_disruptor.pdf

Rogers, E. M. 2010. *Diffusion of Innovations*. Simon and Schuster.

Salmon, F. 2013. "The Bitcoin Bubble and the Future of Currency." Medium.com. Retrieved on Sep 27, 2014, from https://medium.com/money-banking/2b5ef79482cb

Schwert, G. W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics* (7:2), pp. 147-159.

Shi, Z., Rui, H., and Whinston, A. B. 2014. "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter," *MIS Quarterly* (38:1), pp. 123-A126.

Shmueli, G., and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *Mis Quarterly* (35:3), pp. 553-572.

Smyth, L. 2013. "The Demographics of Bitcoin." Sumulacrum. Retrieved on Sep 14, 2014, from http://simulacrum.cc/2013/03/04/the-demographics-of-bitcoin-part-1-updated/

Tafti, A., Zotti, R., and Jank, W. 2013. "Real-Time Diffusion of Information on Twitter and the Financial Markets," in: *The 2013 Conferece on Information Systems and Technology (CIST 2013)*. Minneapolis, MN.

Tirunillai, S., and Tellis, G. J. 2012. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science* (31:2), pp. 198-215.

Trusov, M., Bodapati, A. V., and Bucklin, R. E. 2010. "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research* (47:4), pp. 643-658.

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

Wysocki, P. D. 1999. "Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards." *University of Michigan Business School Working Paper*, No. 98025. Available at SSRN: http://ssrn.com/abstract=160170

Yermack, D. 2013. "Is Bitcoin a Real Currency? An Economic Appraisal." *National Bureau of Economic Research Working Paper Series*, No. 19747. Available at http://www.nber.org/papers/w19747

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.