How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis

Article *in* Journal of Management Information Systems · January 2018

DOI: 10.1080/07421222.2018.1440774

How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis

FENG MAI, ZHE SHAN, QING BAI, XIN (SHANE) WANG, AND ROGER H.L. CHIANG

FENG MAI (feng.mai@stevens.edu; corresponding author) is an assistant professor of information systems in the School of Business at Stevens Institute of Technology. He received his Ph.D. from the University of Cincinnati. His research interests include social media, electronic commerce, and business analytics.

ZHE SHAN (zhe.shan@uc.edu) is an assistant professor in the Department of Operations, Business Analytics, and Information Systems in the Lindner College of Business at the University of Cincinnati. He earned his Ph.D. in business administration and operations research from Penn State University's Smeal College of Business. His research interests include fintech innovation, information security, patient-center health care, and business process analytics.

QING BAI (baiq@dickinson.edu) is an assistant professor of finance in the Department of International Business & Management, Dickinson College. She received her Ph.D. in finance from the University of Cincinnati. Her current research focuses on asset return predictability, patent-based indicators of technological innovation, and financial innovation.

XIN (SHANE) WANG (xwang@ivey.uwo.ca) is an assistant professor in marketing and statistics, and the MBA '80 Faculty Fellow at the Ivey Business School of Western University, Canada. He received his Ph.D. in marketing from the University of Cincinnati. His research focuses on machine learning with applications in marketing, social media analytics, the marketing-information systems interface, and Bayesian statistics.

ROGER H.L. CHIANG (roger.chiang@uc.edu) is a professor of information systems at Carl H. Lindner College of Business, University of Cincinnati. He received his Ph.D. in computers and information systems from the University of Rochester. His research interests focus on business intelligence and analytics, data and knowledge management, and intelligent systems. He has published over fifty refereed articles in journals and conference proceedings, including *ACM Transactions on Database Systems, ACM Transactions on Management Information Systems, Communications of the ACM,*

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/mmis.

Journal of Management Information Systems / 2018, Vol. 35, No. 1, pp. 19–52.

ISSN 0742–1222 (print) / ISSN 1557–928X (online)
DOI: https://doi.org/10.1080/07421222.2018.1440774

Journal of Management Information Systems, Marketing Science, MIS Quarterly, and others. He has also served as a senior editor or associate editor of some leading journals.

ABSTRACT: Bitcoin's emergence has the potential to pave the way for a technological revolution in financial markets. What determines its valuation is an important open question with far-reaching business and policy implications. Building on information systems and finance literature, we examine the dynamic interactions between social media and the monetary value of bitcoin using textual analysis and vector error correction models. We show that more bullish forum posts are associated with higher future bitcoin values. Interestingly, social media's effects on bitcoin are driven primarily by the silent majority, the 95 percent of users who are less active and whose contributions amount to less than 40 percent of total messages. In addition, messages on an Internet forum, relative to tweets, have a stronger impact on future bitcoin value. Overall, our findings reveal that social media sentiment is an important predictor in determining bitcoin's valuation, but not all social media messages are of equal impact. This study offers new insights into the digital currency market and the economic impact of social media.

KEY WORDS AND PHRASES: bitcoin, cryptocurrencies, digital currency, fintech, social media, text mining, vector error correction model.

Digital currency was first introduced in the 1990s in the form of stored value cards for peer-to-peer (P2P) payments that did not require bank authorization [12]. Bitcoin represents a new form of digital currency that uses cryptography and information technology (IT) to facilitate P2P transactions. Since its invention in 2008, bitcoin has captured the attention of the business world. In August 2017, the market capitalization of all bitcoins in the world surpassed US$73 billion.[1] The New York Stock Exchange has created a bitcoin index; well-known retailers such as Dell, Newegg, and Overstock accept bitcoin, as do online payment gateways such as PayPal; and hundreds of bitcoin ATMs operate on four continents. According to one estimate [13], 12 million trading accounts and over 100,000 retailers worked with bitcoin in the fourth quarter of 2015. In less than a decade, bitcoin has emerged from the fringes of the Internet to become a thriving fintech innovation, disrupting existing payment and monetary systems [5].

Accompanying the rising popularity of bitcoin is a vexing question with no clear answer: What determines its value? Finding the factors that influence bitcoin's monetary value (the market price on major bitcoin exchanges) has important practical and theoretical implications. Investors need predictors to estimate future price swings and calculate the expected return. Policymakers need to unpack the forces behind bitcoin to devise regulations and curb financial stability risks [31]. Businesses need to understand the price movement patterns before adopting bitcoin or even launching their own digital currency—what is known as an initial coin offering [48]. For information systems (IS) researchers, bitcoin's value can be viewed as a proxy for the market's confidence and perceived usefulness of the digital currency. Therefore, revealing influential factors affecting bitcoin's value can advance theory by identifying the roles of different parties in the dispersion of new financial technology.

We study whether and to what extent social media can impact bitcoin's value. Prior literature in economics provides models that can explain the worth of currency using a nation's monetary policies, macroeconomic conditions, and inflation and interest rates, among other variables [43]. However, as bitcoin is a digital currency with no government or central bank backing, traditional explanatory variables for currency valuation fall short. Recently, IS researchers suggest that bitcoin resembles a financial investment instrument like stock, rather than a currency [22]. We therefore draw from the theory on the connection between social media and equity value [37, 38] and hypothesize that social media can exert an impact on the bitcoin market. Social media can reveal information that is unobtainable from traditional media. The discussions on social media are also more timely and abundant compared with traditional media, especially in bitcoin's fledgling stages. Thus, establishing the linkage between social media and bitcoin's value could offer investors, regulators, and businesses a new indicator of digital currencies' future value.

Further, bitcoin provides a unique opportunity to understand the economic value of social media and its role in catalyzing the spread of fintech innovations. Thanks to the focus of the media and a generation of investors who are vocal on social media, emotions of bitcoin investors are increasingly visible online. Online discussions about bitcoin are also abundant in quantity and diverse in form. These characteristics make bitcoin an ideal laboratory for testing new theories. Previous literature typically considers social media as a whole, disregarding the mixed signals from various users and channels. This study explicitly analyzes the heterogeneous effects of users with different levels of activity [8]: the active users who contribute most content (the vocal minority), and the relatively inactive users who contribute less often (the silent majority). We also reveal how messages on two major platforms (an Internet forum and Twitter) affect the bitcoin market differently. Integrating these new aspects into economic models can improve our understanding of how social media interacts with the markets. We investigate two research questions:

Is there a predictive relationship between social media and bitcoin value?

Does social media information created by different user cohorts and published on different platforms exhibit the same effect?

To answer these questions, we assembled diverse data from bitcoin trading markets, traditional Internet measures, and social media. We conduct sentiment analyses of messages on an Internet forum (Bitcointalk.org) and Twitter. We use vector error correction models (VECMs) to empirically test the relationship between bitcoin value and social media variables. VECM extends the traditional vector autoregression (VAR) models that are used to study a system of interdependent variables [37]. VECM shares many of the benefits of VAR models. Specifically, VECM accounts for endogeneity, autocorrelation, and reverse causality. It allows us to model the bidirectional causality between pairs of variables. Also, VECM controls for cointegration—a form of long-run dependencies between variables.

Overall, our findings show that social media is an important predictor of future values of bitcoin. More bullish (or bearish) forum posts are significantly associated with higher (or lower) next-day bitcoin market price. Yet not all social media are created equal. Content contributed by relatively inactive users has a larger effect than that from active users. Furthermore, at a daily frequency, forum sentiment offers a better indicator of future values than Twitter sentiment. Variance decomposition analyses suggest that social media metrics explain a significant amount of future variations of bitcoin value. Finally, our social media metrics can improve out-of-sample forecasts of bitcoin values in a three-month test period. Our findings are robust to alternative sentiment metrics, different sampling periods, and fluctuations caused by local government policies.

This research makes two main contributions. First, we develop a more comprehensive understanding of the various factors behind the monetary value of bitcoin. We show that social media sentiment is a meaningful source of variation that can explain and predict bitcoin value. These findings offer a new perspective on the emergence of bitcoin and the diffusion of fintech innovations—for example, prices of digital currencies are subject to the same Keynesian "animal spirits" observed in traditional markets. Theories and empirical models on fintech adoptions should take this perspective into consideration.

Second, we contribute to IS theory by highlighting the different influences of various social media users and platforms. We extend prior findings in Gao et al. [20] to the domain of financial markets and quantify the dynamic effects of different user cohorts. We show that the volume of user contributions and platform differences correlate with the impact of the messages. Therefore, in addition to asking generic questions such as "Does social media affect *X*?" researchers should pay closer attention to the complex and subtle forces that lead to the creation of various social media messages.

Research Background and Hypothesis Development

This research draws primarily on two streams of research in IS and finance: (1) market characteristics of bitcoin, and (2) the impact of social media on the financial markets. We begin by reviewing studies on bitcoin's exchange market and lay out the reasons for incorporating social media metrics in predicting Bitcoin's value (H1). We then highlight the gaps in social media research that motivate our investigation of user cohorts (H2) and platform (H3) differences.

Predictive Relationship Between Social Media and Bitcoin Value

Although the literature on bitcoin has underscored the need to model bitcoin as a financial asset [5], there is no consensus on how bitcoin's monetary value should be determined. One main reason is whether bitcoin—or digital currencies in general—qualifies as currency is in dispute. Government agencies have not provided clear guidelines on how to treat virtual currencies.[2] Yermack [62] tests bitcoin in terms of the three functions of money—as a measure of exchange, store of value, and unit of account—and concludes that it faces challenges in meeting all three criteria. Böhme et al. [5] compare the coefficient of variation for the daily USD–BTC (bitcoin) exchange rate with other currency exchanges; they find that bitcoin is 41 times more volatile than the USD–EUR exchange rate. This extreme price volatility makes it even more important to find meaningful predictors because such predictors could protect individual and business adopters against future price swings.

Bitcoin's market characteristics hint at the possibility that we can study bitcoin using models for stocks. Glaser et al. [22] examine users' motivations for holding bitcoin and conclude that most users treat their bitcoin investment as an asset rather than as a means of payment. In addition, Kristoufek [33, 34] has shown that bitcoin's price correlates with conventional online behavioral metrics such as Google search. The popularity of the search term "Bitcoin" among U.S. Google users correlates highly with both Bitcoin exchange rates (80.6 percent) and weekly total transaction volume at the four largest exchanges (89.1 percent). The strong contemporaneous relationship between attention from Internet users and Bitcoin valuation is similar to the relationship between Web visits and firm equity value [15].

If bitcoin's price formation process indeed resembles that of stock, can we use social media to predict its value? Financial theory asserts that new information will change expectations of investors and thereby affect the stock price [18]. In other words, theory would predict that Bitcoin's price movements follow new information. In modern society, social media has fundamentally changed how information disseminates and has become a valuable source of novel information. Internet forums can disclose new or private information that fundamentally alters bitcoin evaluations, such as when new stores accept bitcoin or forthcoming regulations limit its use. Thus, it is plausible that social media serves as a channel through which information and expectations become reflected in the bitcoin price.

A growing body of empirical studies has examined the interactions between social media and asset value. An analysis of articles published on a social media platform indicates that the opinions expressed in both articles and comments predict future stock prices and earnings surprises [9]. In another examination of the dynamic relationship between social media (online consumer ratings and Web blogs) and firm equity value, social media metrics are found to have significant predictive power for firm equity value [38]. In contrast, in their study of Internet stocks, Tumarkin and Whitelaw [57] find that message board activity cannot predict stock prices, but instead, the causality appears to run from the market to the forums. Antweiler and Frank [1] further indicate that a positive shock to message board posting predicts negative stock returns on the next day. Overall, the relationship between social media and financial market is inconsistent across prior studies.

What is more, there are salient differences between bitcoin and stock. For example, bitcoin has no discounted future cash flows (e.g., dividends) and hence no intrinsic value to speak of. The bitcoin market also has limited depth and a lack of short-selling or derivative instruments, meaning that it is costly to trade. Therefore, the connection between social media information and bitcoin's value cannot be automatically assumed from previous research.

On the other hand, several unique features of bitcoin lead us to hypothesize that social media metrics will have a significant predictive relationship to bitcoin's value. First, in bitcoin's earlier stages, social media has been the most prominent channel through which new information is shared and discussed. If social media can predict the stock price of firms [37] for which many other information disclosure channels (annual reports, financial analysts, etc.) are available, we may expect the same to apply for bitcoin. Second, the design of bitcoin's algorithm ensures that the supply of new coins gets created at a known, geometrically decaying rate, so demand from businesses and individuals represents the main driver of its value. As a new fintech product with strong network effects [47], the attention bitcoin garners on social media can translate to new adopters and positive externalities, consequently increasing its value. The third reason concerns the demographics of users. A survey shows that bitcoin users largely exhibit the demographic characteristics of heavy social media users [17]. Social media messages may naturally have an impact on the bitcoin users' behavior due to more exposure [61]. In addition, peer influence plays a crucial role in how social media impacts asset prices [9]. Such a peer influence effect should be stronger among investors with more shared characteristics because of the homophily in the network [21]. For these reasons, we postulate:

> *Hypothesis 1: (The Social Media Metrics Effects Hypothesis). Social media metrics have significant effects on future bitcoin prices, such that increased positive (negative) sentiments indicate higher (lower) future bitcoin prices.*

Distinctive Impacts of User Cohorts and Platforms on Social Media

When considering the influence of social media, the existing literature tends to use social media as an all-inclusive term even though content is generated on multiple platforms by users with varying behavior. In the previous section, we mentioned that studies examining the relationship between social media and financial markets report inconsistent findings. The mixed results may be an artifact of treating content generated by all users, and from different platforms, as a single source. A few recent studies that dissect social media show that user behaviors correlate with the content they generated. For example, Ludwig et al. [36] show that a user's linguistic style correlates with posting quantity and quality. In the health-care domain, patients who have lower-quality physicians are also less likely to post online reviews [20]. Yet critical gaps remain, especially with regard to whether differences within the social media realm have any bearing on their predictive value in financial markets. With a growing interest in developing online media strategies and integrating social media metrics in business decision making, the distinctive impacts of different user cohorts and platforms are worth investigating. The vast digital footprints created by bitcoin users allow us to test these differences.

We first look at the differences associated with user activities. The power law nature of social media suggests that most users contribute little content as the silent majority, and a small proportion of highly active users contribute the most as the vocal minority. This phenomenon has been empirically verified for online social media such as Twitter and online reviews [41, 44]. Yet the evidence about which cohort is more valuable in terms of reflecting market sentiments and affecting future prices remains inconclusive.

On the one hand, critical mass theory [41] predicts that “the group of active contributors is a minority of the population, but this minority makes the most useful contributions,” thus indicating the vocal minority’s contribution should be of higher quantity and higher quality. Quality aside, the sheer quantity of content produced by the vocal minority should amplify its messages, resulting in disproportional influence [10]. This is because for the online community, more posts are associated with a higher probability of becoming a leader [30]. Early bitcoin adopters who also elect to post large amounts naturally should emerge as community leaders. Research based on social network theory and word-of-mouth theory highlight the importance of these influential users through social media. As Trusov et al. [56] show, community members differ in the frequency, volume, type, and quality of digital content they generate and consume. Leaders have a disproportionate influence on others [23], partly because they have greater exposure to mass media than their followers [49].

Further, from a financial market point of view, the vocal minority also has a crucial role for information cascades, which can lead to herding behavior. That is, opinions and decisions by community leaders are widely observed and assumed to be conveying localized or private information by followers [14]. For instance, groups of mutual funds tend to adopt the investment choices of their successful counterparts [19]. Jiao and Ye [26] show strong evidence that mutual funds collectively enter or exit stocks, following the herd of hedge funds. Thus, the vocal minority may be more influential as an information source.

On the other hand, the opinions of the silent majority may be just as important, if not more so, than those of the vocal minority. First, by definition, the silent majority users contribute to conversations sporadically, usually after highly salient events, and they are not particularly interested in generating buzz [44]. The sentiments of the silent majority, as market measures, thus tend to be more concise and relevant.

Second, the decentralized nature of bitcoin has meant that most grassroots users can be categorized as the silent majority. If its market price reflects the valuation of crowds, then the diversity prediction theorem [45]—collective error diminishes as the diversity of the crowd increases—may apply to the bitcoin market. When it comes to predicting the future movement of asset prices, the silent majority that consists of many independent individuals may outperform the collective of like-minded experts and fanatics.

Third, the silent majority users are less likely to engage in groupthink [25], defined as self-deception, wishful thinking, and conformity to group values that lead to willful blindness and collective denial [3]. Bitcoin has been subjected to criticism that its value may depend on its most zealous users [42]. It is plausible that the vocal minority users engage in numerous discussions of bitcoin, get caught up in glorified ideas, and are more prone to groupthink. If so, they may hold biased views of the future return of the investment and deny any downside risk [54]. In sum, any or a combination of these mechanisms may lead to the result that messages from the silent majority is a more compelling metric for actual investors. Recognizing both sides of the argument, we propose two competing hypotheses:

Hypothesis 2a: (The Vocal Minority Hypothesis). The vocal minority has a stronger impact than the silent majority on bitcoin value.

Hypothesis 2b: (The Silent Majority Hypothesis). The silent majority has a stronger impact than the vocal minority on bitcoin value.

In addition to the differences brought by user activity levels, we propose that various social media platforms affect financial markets differently. The mechanisms of information diffusion, visibility, and representation differ by platform. As examples, we use an Internet forum and Twitter, which differ in three main ways. First, Internet forums generally seek diverse opinions, and reaching consensus is not a primary objective. In contrast, on Twitter, most communications propagate from the sender to followers, who spread the information further by retweeting. Limited by length restrictions, these followers may add brief, general sentiments, but they cannot engage in thorough discussions of the original content. Any dissent can be expressed only via a reply, which is unlikely to receive the same publicity as the original tweet. On forums though, the act of reading a message brings up all replies to that message. According to the theory of social exchange motivations [51], the lack of latent benefit of publicity should suppress critical, in-depth discussions on Twitter. Thus, forum discussions are likely to reflect the complete picture.

Second, a forum is designed to be an archive of all messages; by design, Twitter focuses more on timeliness. It is not uncommon for forum users to engage in a discussion that was started days or months ago, whereas the average life cycle of a tweet is much shorter, and it is difficult for users to trace earlier tweets from an active account. The Twitter search function, for example, does not return messages that are more than a few weeks old. In turn, the information search cost for a nonrecent tweet is much higher, which should reduce the efficiency of the market for information at the intraday level. In addition, behavioral finance scholars note that investors have limited attention capacities, so they respond asymmetrically to more visible information [2]. Since aggregate daily information is more visible and accessible on forums in the form of discussion threads, investors are more likely to respond to it.

Third, a tweet is limited to 140 characters, so information generally must be condensed. A forum does not have this strict limitation. This condensing process creates two limitations in terms of analyzing the impact of these social media. For one, adding external URLs to tweets is a common practice [11], and essential information then gets encapsulated at an external site; it cannot be decoded solely by analyzing (or reading) the tweets themselves. Apart from the URLs, because of the length limitation, contributors on Twitter also are more likely to use numerical expressions to present information in an exact form. Yet numbers lack inherent meaning; they are clear only when used relative to other numerical information [58]. To determine the full implications of a current trading price on Twitter, users would need to know the linguistic context (e.g., increased/decreased) and/or temporal context (e.g., last available price, momentum). If numerical information is indeed more salient on Twitter, whereas verbal information is more salient on forums, we expect the aggregated sentiment measure on Twitter to have less impact. Formally,

> *Hypothesis 3: (The Internet Forum-Content Bitcoin Value Impact Hypothesis). User-generated content from Internet forums, rather than Twitter, has a stronger impact on bitcoin value at a daily level.*

Data

Measures for Monetary Value of Bitcoin

The focal point of our empirical analysis is the monetary value (market price) of bitcoin. We study the dynamic relationship between the natural logarithm of price and other variables. A nice property of *ln* (price) is that the continuously compounded return in bitcoin is the first difference of *ln* (price). If P_t is the bitcoin market price at the end of day t, then the daily continuously compounded return is:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}). \tag{1}$$

This specification means that our model is constant returns to scale; in other words, the model coefficients can be interpreted as the effects of one-unit changes in explanatory variables on investment outcomes, measured by the continuously compounded percentage rate of return. Changes in log price have been widely used in asset pricing studies [7].

Our data set comprises daily market prices (BTC–USD exchange rates) from BitStamp Ltd., the top bitcoin exchange by volume. To control for other observable variations in the bitcoin trading market, we also include transaction volume, trading volume, and volatility in our model. We collected bitcoin-to-bitcoin transaction volume, defined as the total value of all transaction outputs per day,[3] from Bitcoincharts.com. Transaction volume indicates the amount transferred within the bitcoin economy, while the trading volume measure refers to the amount of bitcoin traded for U.S. dollars. We denote the trading volume and transaction volume of day t as V_t and V_t^{TX}.

To capture the effects on bitcoin price brought about by uncertainty, we include a risk measure of bitcoin value using the volatility of bitcoin returns. To measure the volatility of the return, we apply the exponentially weighted moving average model, which tracks changes in volatility with the formula $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2$. The estimate of volatility on day t, σ_t^2, is calculated from σ_{t-1}^2 and the most recent daily percentage change in price. The value of λ governs the responsiveness of the estimate to the most recent daily percentage change. We chose λ = .94, the same value used by RiskMetrics (previously a JPMorgan subsidiary, and now owned by MCSI Inc., which changed its name from Morgan Stanley Capital International and MSCI Barra), which has 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.

Social Media Metrics

We implemented a Python-based Web crawler to collect discussion content from Bitcointalk.org between January 1, 2012, and December 31, 2014. We chose this forum for two reasons: it was rated the most popular bitcoin community in a recent survey [52], and it appears first in the community section of the official Bitcoin website. We limited our data collection to the Bitcoin discussion board, to which users post general news, community developments, innovations, and so forth. After filtering out content beyond our study period, we gathered 343,769 posts and 15,420 topics for further analysis. Each post contained textual content, an author, and a time stamp. Among the 17,215 unique users who posted, the most active 5 percent of users generated 63.11 percent of the content. The average number of posts generated by a single user in the sample period was 19.97; the median was 3. As Figure 1 reveals, the distribution of the number of messages by users follows a typical power law distribution. Most users belong to the silent majority, and a small proportion of the vocal minority generated most of the content.

For the sentiment analysis, we applied a finance sentiment dictionary [35], which includes 2,329 negative and 297 positive sentiment words. We used Natural Language Toolkit 3.0 [4] for the language-processing tasks, such as sentence

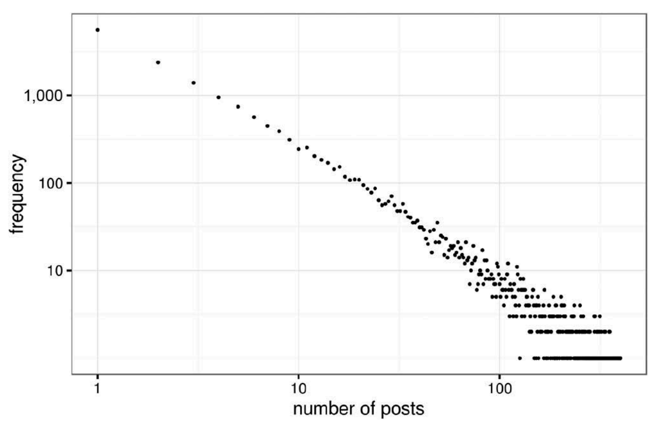

Figure 1. Distribution of Posting Activities among Forum Users (Log-Log Scale)

segmentation, word tokenization, and lemmatization. We counted the number of positive and negative words for each message. If a message contains more positive than negative words, it constitutes a positive post, and vice versa.

To compare the impacts of Twitter and the Internet forum, we also collected tweets that contained the hashtag (#Bitcoin) from the public application program interfaces (API) of Twitter. Twitter's search APIs allow queries within the indices of recent or popular tweets, and also 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 September 16 and December 16, 2014. We thus gathered 3,348,965 unique tweets from 339,295 unique users. On average, 21,910 users tweeted 27,227 messages per day. With these data, we again applied the sentiment dictionary [35] to count the number of positive and negative words in each tweet. If the number of positive words is greater than the negative words, the tweet is classified as positive, and vice versa.

Other Variables

We included a set of traditional Internet activity measures and control variables from the financial market. To measure search interest related to bitcoin, we collected data from Google 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. We also obtained the Web traffic measure *website rank* (traffic rankings of the website) related to Bitcoin.org from the Alexa Web Information Service. External instruments from the financial market include the S&P 500 index, stock market volatility (VIX index from Chicago Board Options Exchange), COMEX gold price, and AAII investor sentiment survey. Because Google Trends and AAII Investor Sentiment provide only weekly data, we used the previous week's measure applied to each day in the subsequent week. Finally, we searched the Thomson Reuters News Analytics (TRNA) database for news articles that contained the word "bitcoin" in the title or full text. We included daily TRNA news sentiment scores in our analyses; these scores are calculated using a proprietary system to give financial professionals an idea of how average sentiment is shifting in the news. Table 1 summarizes the key measures.

Empirical Methodology

To study the dynamic relationship between bitcoin and social media, we use VECMs to capture the interdependencies across time-series. These models extend the VAR system when cointegration is present, meaning that there are long-term common trends among the nonstationary time series [28]. We chose VECM rather than a more traditional multiple regression (cf. [1, 60]), for several reasons. First, as an extension of the VAR model, VECM also allows us to model the recursive relationship

Table 1. Key Measures and Summary Statistics

Variable	Definition	Mean	SD	Median	Min	Max
Bitcoin Market Variables						
ln(P)	Bitcoin price (log)	4.32	1.84	4.17	1.47	7.05
σ	Volatility of bitcoin returns	0.05	0.03	0.04	0.01	0.21
V	Log daily trading volume	11.96	0.48	11.97	10.58	13.74
V^{TX}	Log daily transaction volume	14.40	1.51	14.51	11.09	18.09
Social Media Activities						
POS^F	Number of positive posts	55.58	32.38	49	3	225
NEG^F	Number of negative posts	88.30	58.19	75	3	509
POS^T	Number of positive tweets	3,669	761.9	3,604	955	5,780
NEG^T	Number of negative tweets	3,050	956.9	2,862	1,009	6,716
Control Variables						
rank	Bitcoin.org web traffic rank (log)	9.66	0.94	9.49	7.14	11.64
googletrend	Google Trend for bitcoin	16.33	18.53	12	2	100
sp500	Log S&P 500 closing price	7.40	0.15	7.41	7.15	7.65
vix	COBE Volatility Index	15.33	2.88	14.68	10.32	26.66
gold	Log COMEX gold price	6.73	0.13	6.69	6.50	6.95
investor_sentiment	AAII investor sentiment	13.70	8.94	0	0	38.60
news_sentiment	TRNA Bitcoin news sentiment	0.02	0.14	0	–0.76	0.81

between interdependent variables. We can treat the variables as jointly endogenous, without creating ad hoc model restrictions by separating them as endogenous and exogenous variables. Nor do we need a priori knowledge about the mechanisms 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 the variables. Third, as a time-series model, we can interpret an estimated VECM model using Granger causality. This allows us to test whether the past values of social media variables are useful for predicting the bitcoin market variables and establish the causality between variables.

In our empirical study, we examine models in which the variables include daily observations of bitcoin market activities, namely, price ($ln\ P_t$), volatility (σ_t^2), transaction volume (V_t^{TX}), and trading volume (V_t). The models also include measures of relevant social media activities: number of forum posts or tweets expressing both positive/bullish opinions (POS^F, POS^T) and negative/bearish opinions (NEG^F, NEG^T). Last, we include the relevant control variables defined in Table 1.

We now outline how we determine the appropriate model. Appendix A provides more details on the model specification tests. We first test the stationarity of the variables. Conventional regression estimators, including VAR, encounter problems when applied to nonstationary processes. The regression of two independent random-walk processes would yield a spurious significant coefficient, even if they were not related [24]. We used an augmented Dickey–Fuller unit root test on each variable. Among the time series in the model, news sentiment, VIX, and investor sentiments are stationary; the others are nonstationary with one order of integration. Next, we determined the appropriate lag length p using the Akaike information criterion, which is standard in econometrics literature [40].

Given nonstationary variables, we can model their relationship using VAR by taking the first differences of each time series. Yet this approach can suffer misspecification biases if cointegration is present [40]. Instead, VECM yields more efficient estimators of cointegrating time series using a vector of error correction terms that is equal in length to the number of cointegrating relationships added to the relationship [29]. We performed a Johansen test [27] and confirmed the presence of cointegration in our daily frequency data and concluded that VECM is the appropriate model. We estimated the order of cointegration rank = 5 using Johansen's multiple trace test procedure.

Formally, a VECM with p variables, k lags, and cointegration order r has the following form:

$$\Delta \boldsymbol{Y}_t = \sum_{j=1}^{k-1} \Gamma_j \Delta \boldsymbol{Y}_{tt-k} + \alpha\beta' \boldsymbol{Y}_{tt-1} + \boldsymbol{\mu} + \epsilon_t, \tag{2}$$

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 lag length, and ϵ is the residual vector. Furthermore, 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 with which the variables adjust to the long-term equilibria. The difference between the VECM and the VAR model with first-differenced variables is the additional $\beta' Y_{t-1}$, known as the error correction term. Thus, the VECM model is a special case of the general VAR system expressed as an equivalent VAR:

$$\boldsymbol{Y}_t = (I_k + \alpha\beta' + \Gamma_1)\boldsymbol{Y}_{t-1} + \sum_{j=2}^{k-1}(\Gamma_j - \Gamma_{j-1})\boldsymbol{Y}_{t-j} + \Gamma_{k-1}\boldsymbol{Y}_{t-k} + \boldsymbol{\mu} + \epsilon_t, \tag{3}$$

where I_k is a $k \times k$ identity matrix.

Analyses and Results

VECM Analyses

To test the Social Media Metrics Effects Hypothesis (H1), we examine the effects of the bullishness of forum messages using a VECM. The model includes daily measures of the bitcoin market variables $\ln(P)$, σ, V, and V^{TX} and the social media variables POS^F and NEG^F, as well as all the controls in Table 1. We selected the model with lag length $k = 3$, according to the Akaike information criteria. Table 2 presents the estimated coefficients in the VECM, highlighting the relationship between social media metrics and bitcoin market variables.

We can observe several characteristics of the bitcoin market in Table 2.

First, price and volatility exhibit a strong autoregressive relationship: days with higher prices and volatility tend to precede days of higher prices and volatility. Trading and transaction volume exhibit a strong negative autoregressive relationship, such that higher trading (transaction) volume days tend to precede days of lower trading (transaction) volume. Second, the two social media metrics work as we predicted in H1. Days with unexpected increases in the number of positive (bullish) posts tend to precede days with higher prices and high transaction volume. One more positive forum post is associated with an increase in bitcoin price by 3.53 basis points (1 basis point = one-hundredth of a percentage) next day. Days with unexpected increases in the number of negative (bearish) posts tend to be followed by days with lower bitcoin prices (1.63 basis points). All these relationships are statistically significant. To confirm this result, we also performed a Granger causality test between bitcoin price changes and lagged social media metrics. The social media metrics are individually (χ^2 = 6.37, p = 0.012 for POS^F; χ^2 = 5.48, p = 0.019 for NEG^F) and jointly (χ^2 = 7.00, p = 0.030) significant, meaning the past values of forum sentiments cause the changes in bitcoin value. Finally, the Google Trend measure is the only control variable that affects future bitcoin value. Therefore, forum posts contain new information about the monetary value of bitcoin and provide a better indication of general market sentiment than what is already contained in the trading record, in support of H1.

Table 2. VECM Estimates for Forum Sentiments and Bitcoin

	Dependent Variables (Bitcoin Market)			
Indep Vars	$\ln(P)$	σ	V	V^{TX}
$\ln(P)$ $(t-1)$	0.138***	−0.007***	−0.017	0.152
	(0.030)	(0.002)	(0.346)	(0.190)
$\sigma(t-1)$	0.380	0.140***	4.464	−4.203
	(0.544)	(0.030)	(6.165)	(3.382)
$V(t-1)$	−0.009**	5.76E-4***	−0.209***	0.128***
	(0.004)	(2.10E-4)	(0.043)	(0.024)
$V^{TX}(t-1)$	−2.84E-4	2.48E-4	0.304***	−0.207***
	(0.006)	(3.15E-4)	(0.065)	(0.036)
POS^F $(t-1)$	3.53E-4**	−8.15E-6	6.02E-5	0.004***
	(1.40E-4)	(7.68E-6)	(0.002)	(8.70E-4)
NEG^F $(t-1)$	−1.63E-4**	−1.61E-6	−4.66E-4	−3.94E-4
	(6.98E-5)	(3.83E-6)	(7.91E-4)	(4.34E-4)
$rank(t-1)$	0.012	−5.90E-4	−0.029	0.086
	(0.009)	(4.88E-4)	(0.101)	(0.055)
$googletrend(t-1)$	0.002***	5.67E-6	0.012*	−0.002
	(5.36E-4)	(2.94E-5)	(0.006)	(0.003)
$sp500(t-1)$	−0.557	0.021	9.895*	5.340*
	(0.448)	(0.025)	(5.086)	(2.790)
$vix(t-1)$	−0.002	1.92E-4	0.055	0.050***
	(0.003)	(1.64E-4)	(0.034)	(0.019)
$gold(t-1)$	0.066	0.015	0.770	−1.125
	(0.170)	(0.009)	(1.928)	(1.058)
$investor\_sent(t-1)$	6.42E-4	4.76E-6	−0.004	9.41E-4
	(4.61E-4)	(2.53E-5)	(0.005)	(0.003)
$news\_sent(t-1)$	−0.003	2.32E-4	0.212	−0.022
	(0.016)	(8.75E-4)	(0.181)	(0.099)

Notes: $T = 1{,}901$. Lag length $k = 3$. The first lag estimates are displayed. The controls are not displayed among the dependent variables. Standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To test the Vocal Minority and Silent Majority Hypotheses (H2a, H2b), we estimate two separate VECM models by splitting the forum messages according to user posting activities. One model uses sentiment measures generated from messages posted by the silent majority of users (bottom 95 percent by posting volume); the other model uses the vocal minority (top 5 percent by posting volume). The silent minority generated a mere 36.89 percent of the messages, whereas the vocal minority generated 63.11 percent. Table 3 presents the split sample results.

For the posts by the silent majority, the estimates for their impacts on bitcoin prices are much greater than those in the full sample model (Table 2). An unexpected increase in positive forum posts will predict a surge in bitcoin price by 8.74 basis points ($p < 0.01$). The effect of their posts grows stronger, even though posts from

Table 3. VECM Estimates for Comparing the Silent Majority and Vocal Minority

	Dependent Variables (Bitcoin Market)							
	ln(*P*)		σ		*V*		V^{TX}	
Indep Vars	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
ln(*P*) (*t* – 1)	0.136***	0.143***	–0.007***	–0.007***	0.084	–0.052	0.175	0.186
	(0.030)	(0.030)	(0.002)	(0.002)	(0.347)	(0.345)	(0.191)	(0.189)
σ(*t* – 1)	0.417	0.333	0.139***	0.142***	3.523	4.869	–4.314	–4.491
	(0.543)	(0.544)	(0.030)	(0.030)	(6.181)	(6.159)	(3.399)	(3.379)
V(*t* – 1)	–0.009**	–0.008**	5.51E-4***	5.94E-4***	–0.218***	–0.209***	0.138***	0.125***
	(0.004)	(0.004)	(2.09E-4)	(2.10E-4)	(0.043)	(0.043)	(0.024)	(0.024)
V^{TX}(*t* – 1)	–1.59E-4	–0.002	1.90E-4	2.89E-4	0.313***	0.311***	–0.217***	–0.197***
	(0.006)	(0.006)	(3.13E-4)	(3.13E-4)	(0.065)	(0.065)	(0.036)	(0.035)
POS^F (*t* – 1)	8.74E-4***	2.50E-4	–1.79E-5	–8.89E-6	–0.003	0.001	0.005***	0.005***
	(2.61E-4)	(2.09E-4)	(1.44E-5)	(1.14E-5)	(0.003)	(0.002)	(0.002)	(0.001)
NEG^F (*t* – 1)	–4.27E-4***	–1.31E-4	3.65E-6	–5.93E-6	–5.95E-4	–9.93E-4	9.02E-5	–5.58E-4
	(1.52E-4)	(1.03E-4)	(8.34E-6)	(5.62E-6)	(0.002)	(0.001)	(9.50E-4)	(6.38E-4)
rank(*t* – 1)	0.011	0.014	–5.70E-4	–6.47E-4	–0.049	–0.025	0.096*	0.081
	(0.009)	(0.009)	(4.86E-4)	(4.89E-4)	(0.101)	(0.101)	(0.055)	(0.056)
googletrend(*t* – 1)	0.002***	0.002***	9.67E-6	3.11E-6	0.012*	0.012*	–0.002	–0.002
	(5.35E-4)	(5.37E-4)	(2.94E-5)	(2.94E-5)	(0.006)	(0.006)	(0.003)	(0.003)
sp500(*t* – 1)	–0.571	–0.476	0.022	0.019	9.404*	9.897*	5.578**	5.345*
	(0.447)	(0.449)	(0.025)	(0.025)	(5.082)	(5.084)	(2.794)	(2.789)
vix(*t* – 1)	–0.002	–0.002	2.10E-4	1.81E-4	0.057*	0.054	0.051***	0.052***
	(0.003)	(0.003)	(1.64E-4)	(1.64E-4)	(0.034)	(0.034)	(0.019)	(0.019)
gold(*t* – 1)	0.083	0.056	0.014	0.015	1.023	0.651	–1.144	–1.068

	(0.170)	(0.170)	(0.009)	(0.009)	(1.931)	(1.925)	(1.062)	(1.056)
investor_sent(*t* – 1)	6.86E-4	6.08E-4	3.44E-6	4.27E-6	–0.004	–0.004	8.27E-4	7.18E-4
	(4.61E-4)	(4.61E-4)	(2.53E-5)	(2.52E-5)	(0.005)	(0.005)	(0.003)	(0.003)
news_sent(*t* – 1)	–0.001	–9.35E-4	2.04E-4	2.01E-4	0.225	0.193	0.014	–0.037
	(0.016)	(0.016)	(8.75E-4)	(8.73E-4)	(0.181)	(0.180)	(0.100)	(0.099)

Notes: T = 1,901. Lag length k = 3. The first lag estimates are displayed. Standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

these users account for a smaller proportion of the total posting volume. In contrast, posts by the vocal minority instead provide indicators of future transaction volumes only, not of prices.

The estimates for POS^F and NEG^F are lower in value (2.5 and 1.31 basis points, respectively) and are not statistically significant on prices. The effects of bullishness of messages on future transaction volume are similar between the two groups: an increase in the number of bullish posts predicts higher transaction volume in the following day. Overall, these results support the Silent Majority Hypothesis (H2b): the predictability available from social media depends mostly on content created by the silent majority.

Having established the overall impact of social media and the stronger effects of the silent majority users' sentiment on bitcoin prices, we can study platform differences and test the Internet Forum-Content Bitcoin Value Impact Hypothesis (H3). To determine whether forum messages and tweets have the same impacts, we look at observational days when we collected both forum and Twitter data. By modifying the VECM model, we can include the normalized number of bullish and bearish messages on both the forum and Twitter. The relevant estimates in Table 4 reveal that, when aggregated at the interday level, the sentiments on forum messages are more telling indicators of future bitcoin prices than are tweets. The forum variables (POS^F and NEG^F) predict the prices one day in the future. A 1-standard deviation increase in bullish forum posts is associated with 2.2 percent higher price, and a 1-standard deviation increase in bearish forum posts is associated with a 3.6 percent decrease in price. Both coefficient estimates are statistically significant. In contrast, the Twitter variables (POS^T and NEG^T) have no significant predictive power for bitcoin prices. The Granger causality tests confirm this finding. The forum sentiment of the previous day Granger causes changes in future bitcoin prices ($\chi^2 = 18.58$, $p < 0.01$), whereas there is no Granger causality from Twitter sentiment to daily bitcoin prices ($\chi^2 = 2.60$, $p = 0.27$). In addition, no social media variables exhibit significant predictive power for trading volume and transaction volume during the sample period, though days with more bearish tweets precede days with high volatility. Overall, these results lend support to Internet Forum-Content Bitcoin Value Impact Hypothesis (H3): user-generated content from Internet forums, rather than from Twitter, has a stronger impact on bitcoin value at the daily level.

Forecast Error Variance Decomposition and Forecast Accuracy

Given the estimated effects of forum social media on bitcoin value, we now examine two more practical questions: To what extent does forum sentiment explain the future variance of bitcoin values? More important, do social media help forecast future bitcoin value?

To answer the first question, we derive the forecast-error variance decomposition (FEVD) measures [39]. FEVD can measure the percentage contribution of each type of shock to the forecast error of bitcoin value. Therefore, it is comparable to R^2 in

Table 4. VECM Estimates for Comparing Forum and Twitter

	Dependent Variables			
Independent Variables	ln(P)	σ	V	V^{TX}
ln(P) (t – 1)	–0.077	0.013	–1.736	–0.434
	(0.109)	(0.008)	(1.752)	(1.031)
σ(t – 1)	0.128	0.135	21.750	13.320
	(1.847)	(0.136)	(29.750)	(17.510)
V(t – 1)	–0.022	0.002*	0.349	0.415***
	(0.016)	(0.001)	(0.256)	(0.151)
V^{TX}(t – 1)	0.014	–0.002	0.108	–0.466**
	(0.020)	(0.001)	(0.319)	(0.188)
POS^T (t – 1)	0.012	–6.54E-4	0.007	0.041
	(0.008)	(5.64E-4)	(0.123)	(0.073)
NEG^T (t – 1)	–0.004	7.21E-4*	–0.030	–0.050
	(0.005)	(3.68E-4)	(0.081)	(0.047)
POS^F (t – 1)	0.022***	7.86E-4	0.119	0.027
	(0.008)	(6.02E-4)	(0.132)	(0.077)
NEG^F (t – 1)	–0.036***	–2.31E-4	–0.159	0.033
	(0.009)	(6.29E-4)	(0.137)	(0.081)
rank(t – 1)	–0.009	0.002	0.628	0.500**
	(0.027)	(0.002)	(0.431)	(0.254)
googletrend(t – 1)	0.001	–5.56E-4	–0.115	0.008
	(0.007)	(4.88E-4)	(0.107)	(0.063)
sp500(t – 1)	–1.156	–0.067	–3.668	–5.462
	(1.357)	(0.100)	(21.850)	(12.860)
vix(t – 1)	0.002	–1.57E-4	0.107	–0.021
	(0.007)	(5.05E-4)	(0.110)	(0.065)
gold(t – 1)	0.315	–0.009	–12.250	0.244
	(0.541)	(0.040)	(8.712)	(5.127)
investor_sent(t – 1)	0.003**	5.12E-5	0.008	–0.006
	(0.001)	(9.44E-5)	(0.021)	(0.012)
news_sent(t – 1)	–0.140***	–0.003	–0.620	–0.085
	(0.035)	(0.003)	(0.559)	(0.329)

Notes: $T = 89$. Lag length $k = 3$. The first lag estimates are displayed. Estimates for controls are not displayed. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

regression models and provides insights into the relative importance of the variables. FEVD is defined as:

$$FEVD_{jk}, s = \sum_{i=0}^{s-1} \frac{p_{jk,i}^2}{MSE_k(s)}. \tag{4}$$

The $MSE_k(s)$ is the mean squared error of s-step forecast of variable k, and $p_{jk,i}$ is the effect of a one-unit shock to variable j on k given by the impulse responses function.

FEVD has recently been used in a number of VAR/VECM applications in the IS literature [37, 55]. We follow Luo and Zhang [37] and evaluate the FEVD values at 20 days. We calculate FEVD for three models: a model that includes metrics from all forum messages, a model that includes forum metrics from the silent majority users, and a model that includes forum metrics from the vocal minority.

Table 5 provides a breakdown of forecast error variance of Bitcoin value that can be attributed to shocks to itself or other variables in our system. As would be expected, the Bitcoin price variable accounts for the largest fraction of its own forecast error variance (84.25 percent to 86.66 percent). Consistent with prior research, shocks to search and Internet traffic together can explain between 5.64 percent to 6.27 percent of the variation. When all forum messages are used, the shock in forum sentiment explains 3.60 percent of the variance, which is the third strongest source among all variables. The explanatory power for the social media variables increases to 4.54 percent when we select only the silent majority group. Given that only about 12 percent of the variance can be explained using variables outside of the bitcoin market, these findings point to an economically significant effect. In terms of explaining the variation in future price swings, sentiment on a single forum is comparable in scale to the aggregate behavior of all Google users. On the contrary, the sentiment of vocal minority users only accounts for 0.45 percent of the total variation of bitcoin value—about one-tenth that of the silent majority. Overall, the FEVD analysis further emphasizes that social media sentiments add meaningful explanatory power for bitcoin value, after controlling for bitcoin market variables, Internet and search traffic, and other control variables.

Table 5. Variance of Bitcoin Value Explained by Different Variables

	Model		
	All Forum Messages (percent)	Silent Majority	Vocal Minority
Bitcoin Market			
Price (log)	85.49	84.25	86.66
Volatility	1.86	2.18	1.79
Trading Vol	0.29	0.23	0.38
Transaction Vol	1.00	0.97	1.11
Total	88.64	87.62	89.94
Social Media			
Positive Posts	2.29	2.64	0.02
Negative Posts	1.31	1.90	0.43
Total	3.60	4.54	0.45
Search and Internet Traffic			
Google Trend	5.19	5.07	5.83
Website Rank	0.52	0.57	0.44
Total	5.71	5.64	6.27
Other Controls	2.06	2.20	3.33

To answer the second question, we test the predictive power of social media variables by conducting out-of-sample forecasting. Out-of-sample forecasting is regarded as the ultimate test of a model [53, p. 571]. In this test, we reserve the last quarter of our observations period, from October 1 to December 31, 2014, as the test period. First the model is estimated with the observations prior to the test period. The model is then reestimated period by period through to the last day of the entire sample as the updated parameters are used to generate new one-day ahead forecasts. Such recursive rolling forecasts mimic the actual behavior of a practitioner in real time and are routinely used in economics [43]. We measure the forecasting accuracy using root mean square error (RMSE) and the mean absolute error (MAE). The RMSE is defined as $\sqrt{\frac{1}{n}\sum(\text{actual} - \text{predicted})^2}$, and the MAE is defined as $\frac{1}{\text{n}}\sum|\text{actual} - \text{predicted}|$, where n is the number of forecasting periods (92 days). Smaller RMSE and MAE indicate better model performance.

A three-day moving average model is used as a benchmark for judging the accuracy of VECM forecasts. We estimate two VECMs for each forecast, one uses all the variables but excludes the forum sentiments, and another is the full model that includes the number of positive and negative forum posts generated by the silent majority users. Table 6 presents the results.

In our test period, VECM with social media metrics has the lowest RMSE and MAE. When compared with the three-day moving average model, the RMSE and MAE are reduced by approximately 16 percent (from 16.60 to 13.92) and 14 percent (from 10.89 to 9.35). When compared with the VECM model with no social media metrics, the RMSE and MAE are reduced by 10 percent (from 15.47 to 13.92) and 6 percent (from 9.96 to 9.35). Again, the results provide compelling evidence that social media sentiment has an important bearing on the determination of future bitcoin values.

Robustness Checks

We conducted a series of robustness checks for our results. To remove bias from the specific sentiment measures we used, we considered a combined measure of

Table 6. Comparison of Forecasting Accuracy

	Model		
	3-Day Moving Average	VECM (No Social Media)	VECM (With Social Media)
RMSE	16.60	15.47	13.92
MAE	10.89	9.96	9.35

Notes: The forecasting accuracy measures are calculated using a 92-day period from October 1, 2014 to December 31, 2014. For each day, a model is estimated using all the data prior to that day. The model's parameters are used to forecast next day's bitcoin monetary value.

Table 7. Robustness Checks Using the Alternative Sentiment Measure

	(1)	(2)	(3)	(4)
	All Users	Silent Majority	Vocal Minority	All Users
Forum Bullishness	1.10E-4**	2.80E-4***	8.35E-5	0.012***
	(4.45E-5)	(8.17E-5)	(6.96E-5)	(0.004)
Twitter Bullishness				0.003
				(0.005)

Notes: This table shows the VECM estimates of previous day's social media bullishness on bitcoin prices. The first lag estimates are displayed. T = 1,901 for Models 1–3; T = 89 for Model 4. Lag length $k = 3$. Estimates for controls are not displayed. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

bullishness [1]. We define the bullishness measure as (*POS* – *NEG*) / (*POS* + *NEG*) and reestimate our models using this single measure. Table 7 shows that all the coefficients are in line with our findings using both POS^F and NEG^F: social media bullishness on the forum is associated with future bitcoin returns, and the result is mainly driven by the silent majority users. When we combine forum and Twitter bullishness measures, the forum measure is the more important predictor.

To ensure that our results are not driven by certain events in a specific time frame, we interacted a time dummy with social media metrics and estimated our model again. The time dummy takes a value of one if it is after July 2013. The estimates in Table 8 are largely consistent with our main findings.

Also, the interaction effects are not significant, thereby ruling out the possibility that our results are time specific. As a check of the robustness of our results with respect to the definition of the vocal minority and the silent majority, we adopted 10 percent and 2.5 percent user activity cutoff levels, in addition to the 5 percent level in our main study. The results in Table 9 show that the impact of the vocal minority and that of the silent majority exhibit similar disparities with the new definitions: posts from less active users carry more weight for indicating future price changes.

Finally, recent evidence suggests that the monetary value of bitcoin may be impacted by government regulations and laws. Although we control for news sentiment in our model using the TRNA database, it is possible that actions of foreign governments—especially the Chinese government—are not promptly included in English news. We include bitcoin-related Baidu news trend (data provided by Baidu, the largest search engine in China) in our model as a robustness check. We find that our results still stand with this alternative control. (See Appendix B for details.)

Discussion and Conclusions

Bitcoin and other digital currencies provide unique benefits, including lower transaction costs and stimulus for financial innovation [6]. By breaking down existing

Table 8. Robustness Checks: Effect of Time Periods

	Dependent Variables	
	ln(P)	
Independent Variables	Silent Majority	Vocal Minority
ln(P) (t – 1)	0.133***	0.143***
	(0.031)	(0.031)
σ(t – 1)	0.423	0.376
	(0.546)	(0.546)
V(t – 1)	–0.008**	–0.008**
	(0.004)	(0.004)
V^{TX}(t – 1)	–0.005	–0.004
	(0.005)	(0.006)
POS^F (t – 1)	0.001***	3.89E-4
	(3.05E-4)	(3.58E-4)
NEG^F (t – 1)	–4.69E-4***	–1.63E-4
	(1.67E-4)	(1.06E-4)
POS^F (t – 1) × *post*-07/2013	–4.26E-4	–1.61E-5
	(5.53E-4)	(5.50E-4)
NEG^F (t – 1) × *post*-07/2013	1.46E-5	–2.33E-4
	(4.33E-4)	(3.41E-4)

Notes: T = 1,901. Lag length k = 3. The first lag estimates are displayed. Estimates for controls are not displayed. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9. Robustness Checks: Posting Volume Thresholds

	Cutoff = top 10 percent		Cutoff = top 2.5 percent	
	ln(P)		ln(P)	
Independent Variables	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
POS^F (t – 1)	0.000995***	0.000364**	0.000805***	0.000139
	(0.000363)	(0.000181)	(0.000222)	(0.000248)
NEG^F (t – 1)	–0.000467**	–0.000145	–0.000328***	–0.000142
	(0.000209)	(8.91e-05)	(0.000124)	(0.000120)

Notes: T = 1,901. Lag length k = 3. The first lag estimates are displayed. Estimates for controls are not displayed. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

payment barriers and liberating global trades, they have the potential to generate enormous wealth and social welfare for the economy. Lack of understanding of their price fluctuations, however, could hold back bitcoin and other digital currencies from achieving their full potential. We have accordingly sought to quantify the dynamic relationship between social media and the monetary value of bitcoin. To the best of our knowledge, this study is the first research that systemically explores the economic impact of social media information on bitcoin valuation. The results suggest that social media sentiment is an important leading indicator of future bitcoin price swings. Yet the relationship is complex, because the silent majority exerts a more significant effect, and forum sentiment appears to be a better indicator at the interday level than tweets. Evidence from the Granger causality test, error variance decomposition, and out-of-sample forecasting suggests that forum sentiment has a strong predictive power for bitcoin value.

The findings also have implications for virtual currency adopters, investors, and policymakers. First, the predictive relationship suggests that social media offer substantial novel information about bitcoin's demand among the general public as well as daily fluctuations in its market sentiments. These signals are factored into the price-formation process and influence future returns. Investors thus can discern bitcoin's monetary value from this rich information source. Greater predictability of digital currency values can improve their reliability as a regular component of investment portfolios. For regulators, social media monitoring also offers timely indicators of impending movements of bitcoin prices, which can be used to address the potential systemic risks associated with this unprecedented financial innovation.

Second, companies should strategically evaluate their decision to adopt bitcoin payments. An important motivation for early institutional bitcoin adopters was to capture positive public relations through social media, because "being noted as a Bitcoin innovator can potentially generate favorable press and social media mentions" [46, p. 2]. Our results suggest that companies must think through more than just the marketing consideration of generating positive buzz. The dynamic relationship between social media content and bitcoin value means the future value of accounts receivable can also be affected. This self-fulfilling feedback loop is new for payment systems and could be a distinct feature of similar blockchain-based financial technologies such as Ripple and Ethereum. If leveraged thoughtfully, social media also can drive other fintech innovation in the future.

Although our study focuses on bitcoin, a fintech innovation, the broader implications also can influence general business practices in online social media. Companies should analyze user behaviors and activities on social media while monitoring the content. We have shown that social media messages are not created equal and therefore should not be treated in the same way. The practice of exploiting emotions and influences for marketing purposes is not novel; businesses have long recognized the value of lead users [59] and opinion leaders [32], for example. But our empirical findings highlight the value of the silent, yet influential majority of inactive users. Despite the vocal minority dominating social media, the silent majority users' opinions cannot be overlooked. More marketing and analytic efforts should seek to identify this "heavy tail" of the online community. Moreover, companies can benefit from monitoring discussions on various social media platforms and devising unique strategies for them. For example, the instantaneous buzz on mobile-oriented media (e.g., Twitter) may prompt interactions, but in-depth discussions on Internet forums can paint a more comprehensive picture of participants and thus are more likely to trigger final adoption or purchase decisions.

Our research has several limitations in its data sources and analysis methods, which suggest possible extensions to this study. We used secondary data to identify the association between social media sentiments and future bitcoin prices. Well-designed, randomized experiments could enhance our understanding of the specific findings. Second, we collected data from an English-language Internet forum and limited our Twitter data to messages in English. Bitcoin prices across the globe are highly correlated, and the market consists of investors and adopters worldwide. Comparing messages written in other languages may lead to insights about the potential effects of cultural differences. Moreover, we used financial sentiment as the sole indicator of information in social media. Further studies might identify subtle human emotions (e.g., fear, surprise) in the textual data and investigate their role. Finally, we did not explore the mechanisms that may explain the prominence of the silent majority and the stronger impact of the forum messages. Subsequent analyses of text mining, user social networks, and information diffusion may create new perspectives in understanding this unique phenomenon.

Notes

1 Calculated as the number of coins in existence available to the public multiplied by the U.S. dollar market price.

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.

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.

ORCID

Feng Mai http://orcid.org/0000-0001-6897-8935

References

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

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

3. Bénabou, R. Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies*, *80*, 2 (2012), 429–462. doi:10.1093/restud/rds030.

4. Bird, S. NLTK: the Natural Language Toolkit. *In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*. Sydney, Australia: Association for Computational Linguistics, July 2006, pp. 69–72.

5. Böhme, R.; Christin, N.; Edelman, B.; and Moore, T. Bitcoin: economics, technology, and governance. *Journal of Economic Perspectives*, *29*, 2 (2015), 213–238. doi:10.1257/jep.29.2.213.

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

7. Campbell, J.Y. Understanding risk and return. *Journal of Political Economy*, *104*, 2 (1996), 298–345. doi:10.1086/262026.

8. Chen, A.; Lu, Y.; Chau, P.Y.; and Gupta, S. Classifying, measuring, and predicting users' overall active behavior on social networking sites. *Journal of Management Information Systems*, *31*, 3 (2014), 213–253. doi:10.1080/07421222.2014.995557.

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

10. Chen, J.; Xu, H.; and Whinston, A.B. Moderated online communities and quality of user-generated content. *Journal of Management Information Systems*, *28*, 2 (2011), 237–268. doi:10.2753/MIS0742-1222280209.

11. Chu, Z.; Gianvecchio, S.; Wang, H.; and Jajodia, S. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, *9*, 6 (2012), 811–824. doi:10.1109/TDSC.2012.75.

12. Clemons, E.K.; Croson, D.C.; and Weber, B.W. Reengineering money: The Mondex stored value card and beyond. *International Journal of Electronic Commerce*, *1*, 2 (1996), 5–31. doi:10.1080/10864415.1996.11518281.

13. Hileman, G. *State of Bitcoin and Blockchain 2016*. New York: CoinDesk, January 28, 2016. https://www.coindesk.com/state-of-bitcoinblockchain-2016/

14. Devenow, A., and Welch, I. Rational herding in financial economics. *European Economic Review*, *40*, 3–5 (1996), 603–615. doi:10.1016/0014-2921(95)00073-9.

15. Dewan, R.M.; Freimer, M.L.; and Zhang, J. Management and valuation of advertisement-supported web sites. *Journal of Management Information Systems*, *19*, 3 (2002), 87–98. doi:10.1080/07421222.2002.11045737.

16. Dickey, D.A., and Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 366a (1979), 427–431.

17. Duggan, M.; and Brenner, J. *The Demographics of Social Media Users. Pew Research Center's Internet & American Life Project, 2012*. Washington, DC: Pew Research Center, February 14, 2013.

18. Fama, E.F. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*, 2 (1970), 383–417. doi:10.2307/2325486.

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

20. Gao, G.; Greenwood, B.; McCullough, J.; and Agarwal, R. Vocal minority and silent majority: How do online ratings reflect population perceptions of quality? *MIS Quarterly*, *39*, 3 (2015), 565–589. doi:10.25300/MISQ/2015/39.3.03.

21. Garg, R.; Smith, M.D.; and Telang, R. Measuring information diffusion in an online community. *Journal of Management Information Systems*, *28*, 2 (2011), 11–38. doi:10.2753/MIS0742-1222280202.

22. Glaser, F.; Zimmermann, K.; Haferkorn, M.; and Weber, M.C. Bitcoin: asset or currency? Revealing users' hidden intentions. In M. Avital, J. M. Leimeister and U. Schultze. *Proceedings of the 22nd European Conference on Information Systems*. Tel Aviv: Association for Information Systems, 2014.

23. Goldenberg, J.; Han, S.; Lehmann, D.R.; and Hong, J.W. The role of hubs in the adoption process. *Journal of Marketing*, *73*, 2 (2009), 1–13. doi:10.1509/jmkg.73.2.1.

24. Granger, C.W., and Newbold, P. Spurious regressions in econometrics. *Journal of Econometrics*, *2*, 2 (1974), 111–120. doi:10.1016/0304-4076(74)90034-7.

25. Janis, I. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Boston: Houghton Mifflin, 1972.

26. Jiao, Y., and Ye, P. Mutual fund herding in response to hedge fund herding and the impacts on stock prices. *Journal of Banking and Finance*, *49* (2014), 131–148. doi:10.1016/j.jbankfin.2014.09.001.

27. Johansen, S., and Juselius, K. Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, *52*, 2 (1990), 169–210. doi:10.1111/j.1468-0084.1990.mp52002003.x.

28. Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*, 6 (1991), 1551–1580. doi:10.2307/2938278.

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

30. Johnson, S.L.; Safadi, H.; and Faraj, S. The emergence of online community leadership. *Information Systems Research*, *26*, 1 (2015), 165–187. doi:10.1287/isre.2014.0562.

31. Jopson, B. Regulators say bitcoin poses financial stability risks. *Financial Times*, June 21, 2016.

32. King, C.W., and Summers, J.O. Overlap of opinion leadership across consumer product categories. *Journal of Marketing Research*, *7*, 1 (1970), 43–50. doi:10.2307/3149505.

33. Kristoufek, L. Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, *3* (2013), 1–7. doi:10.1038/srep03415.

34. Kristoufek, L. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PloS ONE*, *10*, 4 (2015), e0123923. doi:10.1371/journal.pone.0123923.

35. Loughran, T., and McDonald, B. Measuring readability in financial disclosures. *Journal of Finance*, *69*, 4 (2014), 1643–1671. doi:10.1111/jofi.12162.

36. Ludwig, S.; De Ruyter, K.; Mahr, D.; Wetzels, M.; Brüggen, E.; and De Ruyck, T. Take their word for it: The symbolic role of linguistic style matches in user communities. *MIS Quarterly*, *38*, 4 (2014), 1201–1217. doi:10.25300/MISQ/2014/38.4.12.

37. Luo, X., and Zhang, J. How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, *30*, 2 (2013), 213–238. doi:10.2753/MIS0742-1222300208.

38. Luo, X.; Zhang, J.; and Duan, W. Social media and firm equity value. *Information Systems Research*, *24*, 1 (2013), 146–163. doi:10.1287/isre.1120.0462.

39. Lütkepohl, H. Testing for causation between two variables in higher dimensional VAR models. In H. Schneeweiß, and K. F. Zimmermann (eds.), *Studies in Applied Econometrics*, Heidelberg: Physica-Verlag (1993), 75–91.

40. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. Heidelberg: Springer Science & Business Media, 2005.

41. Peddibhotla, N.B., and Subramani, M.R. Contributing to public document repositories: A critical mass theory perspective. *The Organization Studies,* 28, 3 (2007), 327–346. doi:10.1177/0170840607076002.

42. McCrum, D. Bitcoin's place in the long history of pyramid schemes. *Financial Times*, November 10, 2015.

43. Meese, R.A., and Rogoff, K. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*, 1–2 (1983), 3–24. doi:10.1016/0022-1996(83)90017-X.

44. Mustafaraj, E.; Finn, S.; Whitlock, C.; and Metaxas, P.T. Vocal minority versus silent majority: Discovering the opinions of the long tail. In *Proceedings of 2011 IEEE International Conference on Privacy, Security, Risk and Trust, and the 2011 IEEE International Conference on Social Computing*. Boston, MA: IEEE, 2011, pp. 103–110.

45. Page, S.E. Making the difference: Applying a logic of diversity. *Academy of Management Perspectives*, *21*, 4 (2007), 6–20. doi:10.5465/AMP.2007.27895335.

46. PricewaterhouseCoopers. *Digital Disruptor: How Bitcoin Is Driving Digital Innovation in Entertainment, Media and Communications*. London: Digital Intelligence Series, May 2014.

47. Qiu, L.; Tang, Q.; and Whinston, A.B. Two formulas for success in social media: Learning and network effects. *Journal of Management Information Systems*, *32*, 4 (2015), 78–108. doi:10.1080/07421222.2015.1138368.

48. Ren, S., and Culpan, T. Ethereum's wild ride needs to slow. *Bloomberg Businessweek*, July 13, 2017.

49. Rogers, E.M. *Diffusion of Innovations*. New York: Simon and Schuster, 2010.

50. Schwert, G.W. Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, *7*, 2 (1989), 147–159.
51. Shi, Z.; Rui, H.; and Whinston, A.B. Content sharing in a social broadcasting environment: Evidence from Twitter. *MIS Quarterly*, *38*, 1 (2014), 123–142. doi:10.25300/MISQ.
52. Smyth, L. *The Demographics of Bitcoin*. Simulacrum, 2013.
53. Stock, J.H., and Watson, M.W. *Introduction to Econometrics*. Boston: Addison-Wesley, 2003.
54. Taffler, R.J., and Tuckett, D.A. Emotional finance: The role of the unconscious in financial decisions. In H.K. Baker and J.R. Nofsinger (eds.), *Behavioral Finance: investors, Corporations, and Markets*. New York: Wiley, 2010, pp. 95–112.
55. Thies, F.; Wessel, M.; and Benlian, A. Effects of social interaction dynamics on platforms. *Journal of Management Information Systems*, *33*, 3 (2016), 843–873. doi:10.1080/07421222.2016.1243967.
56. Trusov, M.; Bodapati, A.V.; and Bucklin, R.E. Determining influential users in Internet social networks. *Journal of Marketing Research*, *47*, 4 (2010), 643–658. doi:10.1509/jmkr.47.4.643.
57. Tumarkin, R., and Whitelaw, R.F. News or noise? Internet postings and stock prices. *Financial Analysts Journal*, *57*, 3 (2001), 41–51. doi:10.2469/faj.v57.n3.2449.
58. Viswanathan, M., and Childers, T.L. Processing of numerical and verbal product information. *Journal of Consumer Psychology*, *5*, 4 (1996), 359–385. doi:10.1207/s15327663jcp0504_03.
59. Von Hippel, E. Lead users: A source of novel product concepts. *Management Science*, *32*, 7 (1986), 791–805. doi:10.1287/mnsc.32.7.791.
60. Wysocki, P.D. Cheap talk on the web: the determinants of postings on stock message boards. Working paper no. 98025. University of Michigan Business School, Ann Arbor, 1999.
61. Xie, K., and Lee, Y.-J. Social media and brand purchase: Quantifying the effects of exposures to earned and owned social media activities in a two-stage decision making model. *Journal of Management Information Systems*, *32*, 2 (2015), 204–238. doi:10.1080/07421222.2015.1063297.
62. Yermack, D. Is bitcoin a real currency? An economic appraisal. Working paper series. National Bureau of Economic Research, Cambridge, MA, 2013.

Appendix A: VECM Model Specifications

Step 1: Stationarity of variables. We first test the variables for unit roots and determine if the variables are stationary. We perform the augmented Dickey–Fuller (ADF) test [16]. The null hypothesis is that a variable contains a unit root, which indicates that the variable follows a nonstationary process. If the series is stationary after differencing once, it is integrated of order 1 or I(1). The alternative hypothesis is that the series was generated by a stationary process; the series is integrated of order zero of I(0). When performing the ADF test, we include a lag using the rule of thumb $p = 12 \times (T/100)^{1/4}$ as recommended by Schwert [50]. As Table A1 shows, we cannot reject the null hypotheses for bitcoin market variables and social media variables. We conclude that these time-series exhibit a unit root. Among the control variables, we reject null hypothesis of a unit root for VIX,

Table A1. Results of Unit Root Tests

Variables	Meaning	Test Stats	p-value	Order of Integration
Bitcoin Market Variables				
ln(P)	Bitcoin price (log)	–1.171	0.6859	I(1)
σ	Volatility of bitcoin returns	–1.931	0.3177	I(1)
V	Log daily trading volume	–1.825	0.3679	I(1)
V^{TX}	Log daily transaction volume	–3.039	0.1215	I(1)
Social Media Activities				
POS^F	Number of positive posts	–1.664	0.4500	I(1)
NEG^F	Number of negative posts	–2.358	0.1540	I(1)
POS^T	Number of positive tweets	–2.204	0.4878	I(1)
NEG^T	Number of negative tweets	–3.112	0.1033	I(1)
Control Variables				
rank	Bitcoin.org web traffic rank (log)			
googletrend	Google Trend for Bitcoin	–1.732	0.4145	I(1)
sp500	Log S&P 500 closing price	–0.829	0.8104	I(1)
vix	COBE Volatility Index	–5.649	< 0.001	I(0)
gold	Log COMEX gold price	–0.815	0.8147	I(1)
investor_sentiment	AAII investor sentiment	–6.379	< 0.001	I(0)
news_sentiment	TRNA Bitcoin news sentiment	–5.503	< 0.001	I(0)

investor sentiment, and news sentiment, and failed to reject the null hypothesis for rank, google trend, and gold index.

Step 2: Number of lags. We use the Akaike information criterion (AIC) to choose the optimal lag length in the model. We estimate VAR models with length varying from 0 to 12 and compute the log-likelihood and the AIC. AIC for a VAR model is defined as $-2L + 2(k + 2kp)$, where L is the log-likelihood, k is the number of coefficients, and p is the lag length. A smaller AIC indicates better trade-off between model fits and complexity. Based on results in Table A2, we select the model with $p = 3$.

Step 3: Cointegration Tests. Table A3 reports the results from the Johansen trace test [27] for cointegration rank. The trace test is a sequential hypothesis testing procedure. It starts from the null hypothesis of no integration (maximum rank = 0), and compares the log-likelihood of the unconstrained model that includes one more cointegrating equation with the constrained model. The test is repeated until the first null hypothesis is not rejected. From Table A3, we reject the null hypothesis of no cointegration, which confirms that VECM is the appropriate model. The trace test stops at the null hypothesis that there are five cointegration relations in the bitcoin market. Therefore, we proceed to estimate our VECM with rank = 5.

Table A2. Selecting Optimal Lag Length

Lag	Log-Likelihood	AIC
0	−17,762.2	32.86
1	−2,226.1	4.45
2	−1,930.1	4.22
3	−1,711.1	4.12*
4	−1,555.2	4.15
5	−1,422.0	4.21
6	−1,214.5	4.14
7	−1,080.1	4.21
8	−928.6	4.24
9	−792.0	4.30
10	−675.7	4.40
11	−554.3	4.49
12	−432.8	4.57

Table A3. Trace Test for Cointegration

Rank	Log Likelihood	Eigenvalue	Trace Statistic	5 percent Critical Value
0	777.9	—	909.0	233.13
1	914.7	0.22	635.4	192.89
2	1,025.0	0.18	414.8	156.00
3	1,105.0	0.14	254.9	124.24
4	1,164.2	0.10	136.3	94.15
5	1,199.4	0.06	66.1*	68.52

* Indicates the first null hypothesis that is not rejected

Appendix B: VECM with Baidu Trend

This section shows that our results are robust when we control for the recent Chinese government meddling with bitcoin. We used the data from Baidu news monitoring (zhishu.baidu.com) and downloaded the Chinese news intensity data for bitcoin (Figure B1). Baidu is the largest search engine in China. Its news aggregation service provides broad coverage of government policy announcements through the major Chinese news outlets.

We replicated VECM analyses to examine the relationship between social media and bitcoin values. As Table B1 demonstrates, with the added control of Baidu news intensity, the social media variables remain to have a significant predictive relationship with future bitcoin prices in the Social Media Metrics Effects Hypothesis (H1). In addition, Table B2 shows the distinct effects hold in the Vocal Minority and Silent Majority Hypotheses (H2a, H2b). Finally, Table B3 suggests that when both forum

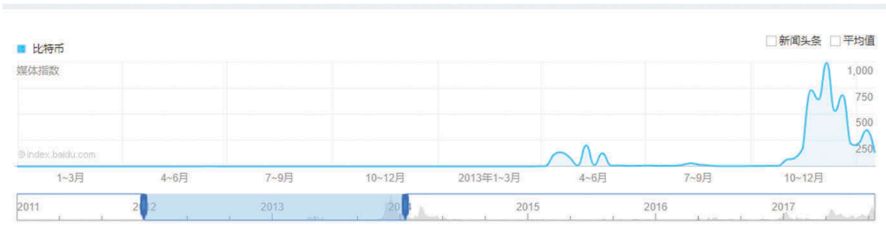

Figure B1. Baidu News Intensity

Table B1. VECM Estimates for Forum Sentiments and Bitcoin

	Dependent Variables (Bitcoin Market)			
Indep Vars	ln(P)	σ	V	V^{TX}
ln(P) (t – 1)	0.133***	–0.007***	–0.064	0.199
	(0.031)	(0.002)	(0.349)	(0.193)
σ (t – 1)	0.631	0.159***	4.158	–3.532
	(0.543)	(0.030)	(6.099)	(3.374)
V (t – 1)	–0.007*	7.07E-4***	–0.180***	0.136***
	(0.004)	(2.15E-4)	(0.044)	(0.024)
V^{TX} (t – 1)	–0.002	–1.72E-4	0.243***	–0.202***
	(0.006)	(3.28E-4)	(0.067)	(0.037)
POS^F (t – 1)	0.001***	–2.28E-5	–0.002	0.007***
	(3.66E-4)	(2.02E-5)	(0.004)	(0.002)
NEG^F (t – 1)	–4.42E-4**	1.50E-5	2.89E-4	3.29E-4
	(2.10E-4)	(1.16E-5)	(0.002)	(0.001)
Rank(t – 1)	1.84E-7	–1.64E-8	1.97E-7	1.30E-6
	(3.20E-7)	(1.77E-8)	(3.60E-6)	(1.99E-6)
Baidunews(t – 1)	1.07E-7	7.05E-9	6.68E-7	6.18E-7
	(1.28E-7)	(7.09E-9)	(1.44E-6)	(7.97E-7)
googletrend(t – 1)	0.002***	2.21E-5	0.014**	–0.002
	(5.44E-4)	(3.01E-5)	(0.006)	(0.003)
sp500(t – 1)	–0.481	0.026	10.550**	5.205*
	(0.450)	(0.025)	(5.059)	(2.798)
vix(t – 1)	–0.002	2.20E-4	0.054	0.053***
	(0.003)	(1.66E-4)	(0.034)	(0.019)
gold(t – 1)	0.069	0.014	0.555	–0.977
	(0.171)	(0.009)	(1.926)	(1.066)
investor_sent(t – 1)	5.01E-4	–6.41E-6	–0.006	3.42E-4
	(4.66E-4)	(2.58E-5)	(0.005)	(0.003)
news_sent(t – 1)	–9.74E-4	1.47E-4	0.203	0.010
	(0.016)	(8.85E-4)	(0.180)	(0.100)

Notes: T = 1,901. Lag length k = 3. The first lag estimates are displayed. The controls are not displayed among the dependent variables. Standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B2. VECM Estimates for Comparing the Silent Majority and Vocal Minority

	Dependent Variables (Bitcoin Market)							
	ln(P)		σ		V		V^{TX}	
Independent Variables	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
ln(P)(t – 1)	0.129***	0.143***	–0.007***	–0.008***	–0.072	–0.138	0.193	0.190
	(0.031)	(0.031)	(0.002)	(0.002)	(0.348)	(0.346)	(0.193)	(0.191)
σ(t – 1)	0.625	0.570	0.159***	0.159***	4.351	5.104	–3.468	–3.651
	(0.540)	(0.543)	(0.030)	(0.030)	(6.081)	(6.085)	(3.366)	(3.350)
V(t – 1)	–0.007*	–0.007*	7.24E-4***	7.33E-4**	–0.180***	–0.191***	0.135***	0.126***
	(0.004)	(0.004)	(2.14E-4)	(2.12E-4)	(0.043)	(0.043)	(0.024)	(0.024)
V^{TX}(t – 1)	–0.002	–0.002	–1.61E-4	–6.51E-5	0.241***	0.257***	–0.203***	–0.188***
	(0.006)	(0.006)	(3.29E-4)	(3.28E-4)	(0.067)	(0.067)	(0.037)	(0.037)
POS^F (t – 1)	9.61E-4***	2.47E-4	–9.54E-6	–5.21E-6	–8.67E-4	0.002	0.004**	0.005***
	(2.65E-4)	(2.14E-4)	(1.47E-5)	(1.18E-5)	(0.003)	(0.002)	(0.002)	(0.001)
NEG^F (t – 1)	–4.11E-4***	–1.17E-4	6.64E-6	–3.87E-6	9.13E-5	–6.07E-4	2.69E-4	–4.01E-4
	(1.53E-4)	(1.04E-4)	(8.46E-6)	(5.71E-6)	(0.002)	(0.001)	(9.51E-4)	(6.40E-4)
rank(t – 1)	1.77E-7	1.68E-7	–1.68E-8	–2.05E-8	2.28E-7	1.61E-7	1.27E-6	9.81E-7
	(3.19E-7)	(3.23E-7)	(1.77E-8)	(1.77E-8)	(3.59E-6)	(3.62E-6)	(1.99E-6)	(1.99E-6)
baidunews(t – 1)	8.29E-8	9.42E-8	7.16E-9	8.47E-9	6.63E-7	7.64E-7	5.35E-7	6.55E-7
	(1.28E-7)	(1.29E-7)	(7.10E-9)	(7.09E-9)	(1.44E-6)	(1.44E-6)	(7.99E-7)	(7.95E-7)
googletrend(t – 1)	0.002***	0.002***	2.25E-5	1.95E-5	0.014**	0.014**	–0.002	–0.002
	(5.43E-4)	(5.47E-4)	(3.01E-5)	(3.01E-5)	(0.006)	(0.006)	(0.003)	(0.003)
sp500(t – 1)	–0.501	–0.453	0.025	0.022	10.530**	10.600**	5.220*	5.116*
	(0.449)	(0.453)	(0.025)	(0.025)	(5.058)	(5.074)	(2.799)	(2.793)
vix(t – 1)	–0.002	–0.002	2.13E-4	1.91E-4	0.055	0.054	0.053***	0.054***

	(0.003)	(0.003)	(1.66E-4)	(1.66E-4)	(0.034)	(0.034)	(0.019)	(0.019)
gold(*t* – 1)	0.055	0.048	0.014	0.015	0.582	0.443	–1.081	–1.029
	(0.171)	(0.172)	(0.009)	(0.009)	(1.924)	(1.924)	(1.065)	(1.059)
investor_sent(*t* – 1)	5.40E-4	4.83E-4	–6.31E-6	–7.48E-6	–0.006	–0.006	5.03E-4	4.12E-4
	(4.65E-4)	(4.67E-4)	(2.58E-5)	(2.57E-5)	(0.005)	(0.005)	(0.003)	(0.003)
news_sent(*t* – 1)	–0.002	–4.03E-4	1.64E-4	1.44E-4	0.207	0.179	0.012	–0.035
	(0.016)	(0.016)	(8.87E-4)	(8.85E-4)	(0.180)	(0.180)	(0.100)	(0.099)

Notes: T = 1,901. Lag length k = 3. The first lag estimates are displayed. Standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B3. VECM Estimates for Comparing Forum and Twitter

	Dependent Variables (Bitcoin Market)			
Indep Vars	ln(*P*)	σ	*V*	V^{TX}
ln(*P*) (*t* – 1)	–0.017	0.023**	0.525	0.306
	(0.148)	(0.009)	(1.966)	(1.215)
σ (*t* – 1)	1.662	0.116	–19.910	3.294
	(2.636)	(0.160)	(35.100)	(21.690)
V (*t* – 1)	–0.020	0.001	–0.090	0.277*
	(0.018)	(0.001)	(0.240)	(0.149)
V^{TX} (*t* – 1)	0.033	–0.001	–0.196	–0.518**
	(0.026)	(0.002)	(0.341)	(0.211)
POS^T (*t* – 1)	0.007	–5.36E-4	–0.053	0.019
	(0.006)	(4.90E-4)	(0.108)	(0.067)
NEG^T (*t* – 1)	–0.009	5.36E-4	0.034	–0.046
	(0.007)	(4.03E-4)	(0.088)	(0.055)
POS^F (*t* – 1)	0.013*	–6.83E-5	0.025	0.003
	(0.008)	(4.45E-4)	(0.098)	(0.060)
NEG^F (*t* – 1)	–0.023**	1.12E-4	–0.246*	0.013
	(0.010)	(6.30E-4)	(0.138)	(0.086)
rank (*t* – 1)	1.24E-6	1.87E-7	3.59E-5	3.25E-5
	(2.46E-6)	(1.49E-7)	(3.28E-5)	(2.03E-5)
baidunews (*t* – 1)	0.005	–3.27E-4	0.015	0.035
	(0.008)	(4.95E-4)	(0.109)	(0.067)
googletrend (*t* – 1)	6.34E-6	5.82E-7*	1.44E-4*	6.61E-5
	(5.55E-6)	(3.37E-7)	(7.39E-5)	(4.57E-5)
sp500 (*t* – 1)	–0.679	–0.023	18.630	–2.615
	(1.696)	(0.103)	(22.580)	(13.960)
vix (*t* – 1)	0.007	5.12E-5	0.154	–0.014
	(0.008)	(4.74E-4)	(0.104)	(0.064)
gold (*t* – 1)	0.145	0.005	–10.270	0.258
	(0.649)	(0.039)	(8.643)	(5.341)
investor_sent (*t* – 1)	0.001	3.52E-7	–4.80E-4	–0.010
	(0.001)	(8.89E-5)	(0.019)	(0.012)
news_sent (*t* – 1)	–0.036	–5.53E-4	–0.353	–0.103
	(0.033)	(0.002)	(0.442)	(0.273)

Notes: $T = 89$. Lag length $k = 3$. The first lag estimates are displayed. Estimates for controls are not displayed. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

and Twitter sentiments are included, only forum variables have significant relationships with future Bitcoin price, as in the Internet Forum-Content Bitcoin Value Impact Hypothesis (H3). The evidence supports that our main results hold when we account for shocks in Chinese government regulations.