From Bitcoin to Big Coin:

The Impact of Social Media on Bitcoin Performance

Feng Mai*
School of Business
Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030, USA
Email: feng.mai@stevens.edu
Phone: +1 (201)-216-3815

Qing Bai
Department of International Business & Management
Dickinson College
28 N College St, Carlisle, PA 17013, USA
Email: baiq@dickinson.edu
Phone: +1 (717)-254-8714

Zhe Shan
Department of Operations, Business Analytics, and Information Systems
Carl H. Lindner College of Business
University of Cincinnati
Cincinnati, Ohio 45221-0130, USA
Email: zhe.shan@uc.edu
Phone: +1 (513)-556-7006

Xin (Shane) Wang
Marketing Discipline
Ivey Business School, Western University
London, Ontario, Canada
Email: xwang@ivey.uwo.ca
Phone: +1 (519)-661-4272

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
Email: roger.chiang@uc.edu
Phone: +1 (513)-556-7086

* **Corresponding author.**

From Bitcoin to Big Coin:
The Impact of Social Media on Bitcoin Performance

Abstract

Bitcoin's emergence might pave the way for a technological revolution in financial markets, especially as it spreads widely through social media. This empirical study therefore examines the dynamic relationship between social media, which consist of mixed signals from various users and platforms, and the value of Bitcoin. A combination of textual analysis with econometric models demonstrates that more bullish posts are associated with higher future Bitcoin returns. The social media effects on Bitcoin are mostly driven by the silent majority, namely, the 95% of users who are less active and whose contribution amounts for less than 40% of total messages. Messages on the Internet forum, relative to tweets, have stronger impacts on future Bitcoin returns. Several mechanisms produce these outcomes, as revealed through text analysis of the pertinent messages. The results reveal that not all social media messages are created equal in terms of explaining the adoption of new financial technologies.

Keywords: Bitcoin, cryptocurrency, online communities, social media, user-generated content

1. INTRODUCTION

Since the 2008 invention of Bitcoin, this virtual currency has functioned as an emerging digital phenomenon in the financial technology realm. In June 2016, the value of all Bitcoins in the world surpassed US$8 billion; the New York Stock Exchange has created a Bitcoin index; well-known retailors 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 recent estimate [16], 12 million trading accounts and over 100,000 retailers worked with Bitcoin in the fourth quarter of 2015. Thus, in just a few years, Bitcoin has emerged from the fringes to promise a transformative financial platform with implications for consumers' daily lives.

Considering the impossibility of ignoring Bitcoin and its explosive growth, investors, policy makers, and institutional adopters all need further information about the forces that influence virtual currencies, especially noting the concerns about Bitcoin's considerable price volatility and risk. According to one survey by PricewaterhouseCoopers LLP [70], 82% of consumers express concerns about fluctuations in the Bitcoin market. Its monetary value offers an objective proxy that describes the popularity and adoption of the virtual currency and its underlying blockchain technology. The current study thus seeks to investigate the mechanisms of Bitcoin price formation, and thereby gauge its long-term success prospects, by providing insights into whether and to what extent social media can fuel its diffusion.

Specifically, we investigate the dynamic relationship between social media and Bitcoin. Social media capture the "wisdom of the crowd" and provide convenient, low-cost platforms that enable early adopters to interact and provide feedback about the ecosystem. Even if Bitcoin is a cryptocurrency, or virtual/digital currency, its status as "currency" is somewhat disputable, especially because its daily exchange rates exhibit greater volatility than and virtually no correlation with conventional currencies [89]. Some researchers suggest that Bitcoin actually resembles a financial investment instrument, like stock, rather than a currency [30]. Thus, traditional currency valuation models, based on monetary theory, are ineffective for explaining the value of Bitcoin. To develop a deeper understanding of its valuation and adoption, we need data sources beyond the Bitcoin exchange market. Furthermore, if users can earn monetary gains by investing in Bitcoin—the most popular reason users hold Bitcoin according to one recent survey [88]—social media might provide useful perspectives for investors who seek advice in financial markets [11, 80]. Because of Bitcoin's decentralized structure and exclusively online presence, value derives not from gold or government fiat but from what people assign to it. Therefore, pertinent discussions and opinions expressed in social media represent the sentiment of the crowd, and can impact or predict the future dynamics of Bitcoin prices.

On the flipside, Bitcoin provides a unique opportunity to understand the economic value of social media and its role in catalyzing the emergence of this innovation. Media spotlights and a generation of investors who are vocal on social media make investor emotions increasingly visible. Online user-generated content (UGC) about Bitcoin is abundant in quantity and diverse in form. Yet previous literature usually considers social media as a whole, disregarding the mixed signals from various users and channels. Therefore, to go beyond quantifying the impacts of social media on Bitcoin value, we explicitly analyze the heterogeneous effects of active users who contribute most contents (vocal minority) and the relatively inactive users who contribute less often (silent majority). We also reveal how messages on two major platforms (Internet forum and Twitter) affect Bitcoin prices differently. Integrating these new aspects into financial models can enrich understanding of how online UGC interacts with the markets. Two main research questions guide our investigation:

- Can user-generated content, available through social media, affect Bitcoin prices?

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

To answer these questions, we assemble 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. In the empirical analysis of the Bitcoin market with social media variables, we use vector error correction models (VECM), which extend the popular vector autoregression (VAR) models used to study the dynamic relationships between social media and market performance [57, 58]. In addition to leveraging the benefits of VAR models, such as accounting for endogeneity, autocorrelation, and reverse causality, VECM controls for cointegration and long-run dependencies, as we confirm in our analyses. Our findings in turn affirm that social media help shape the future value of Bitcoin, in that more bullish (bearish) forum posts are associated with higher (lower) next-day Bitcoin returns. Yet not all social media are created equal. Content contributed by relatively inactive users (silent majority) has a surprisingly stronger impact on Bitcoin returns than that from active users (vocal minority). Finally, at a daily frequency, forum sentiment offers a better indicator of future returns than Twitter sentiment. A series of post-hoc analyses, in which we study message contents, linguistic styles, and posting behaviors using computational methods, also pinpoints the underlying mechanisms that lead to the observed differences. In particular, we leverage machine learning techniques to demonstrate that the silent majority and vocal minority are concerned with different aspects of Bitcoin. Furthermore, in the forum and on Twitter, the way that interactions take place and information gets presented differ fundamentally.

This research therefore makes two main contributions to extant literature. First, we offer an initial examination of social media's role in Bitcoin's ecosystem. The empirical evidence reveals that the value of Bitcoin is associated with social media; crowd sentiments on social media can explain the future value of Bitcoin. This insight suggests new perspectives on the emergence of Bitcoin and the diffusion of financial technological innovations in general – for example, they are subject to the same Kaynes' "animal spirits" observed in traditional markets. Adoption and investment decisions regarding new financial technology platforms thus should adopt a similar perspective. Second, we contribute to information systems theory by revealing the different influences of various social media users and platforms. We extend prior findings (Ludwig, et al. [54] and Gao, et al. [29]) to the domain of financial markets and systematically study the underlying mechanisms using mixed methods. We show that the volume of user contributions correlates with the content, and the design of the platform also shapes how messages are presented. These mechanisms in turn result in heterogeneous effects. 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.

In the next section, we review the institutional background of Bitcoin and related research on social media. Then we formally develop the hypotheses, introduce the data and measures for our empirical study, and present our empirical models. After discussing our main findings and outlining the robustness checks, we conduct post-hoc content analyses to triangulate the relevant evidence. We conclude with some implications and potential directions for future work.

2. INSTITUTIONAL BACKGROUND AND RELATED WORK

2.1 Characteristics of the Bitcoin Market

Many studies discuss Bitcoin in general [4, 35, 49, 64], and a few explore price formation [47, 48], identifying several factors that affect Bitcoin prices, such as supply and demand or attractiveness for investors. Yet none of these factors has been explained using standard financial theory [48]. Bitcoin's design and regulatory environment distinguish it in four main ways from traditional consumer products and financial instruments. First, there is a strong link between Bitcoin value and public attention. As Böhme, et al. [8] show, the popularity of the search term "Bitcoin" among U.S. Google users correlated highly with both Bitcoin exchange rates (ρ = 0.806) and weekly total transaction volume at the four largest exchanges (ρ = 0.891). The strong contemporaneous relationship between attention from Internet users and Bitcoin valuation implies a potentially predictive relationship between social media content and Bitcoin returns that is stronger than that for stocks or consumer products.

Second, no consensus exists about the fair value of Bitcoin. Unlike stocks, for which the fair market value can be estimated by discounting future cash flows (dividends), an equilibrium Bitcoin value is hard to calculate. Without a universally accepted pricing model, online messages can disclose new or private information that fundamentally alters Bitcoin evaluations, such as when new stores accept Bitcoins or forthcoming regulations appear likely to limit its use. The interpretation of such news can vary across users. Interactive social media then can serve as a platform for contesting these different interpretations, which may induce trading in the market Harris and Raviv [37] and provide another influence channel on future Bitcoin movements.

Third, as an investment, Bitcoin's return is extremely volatile. Böhme, et al. [8] compare the coefficient of variation for the daily USD–Bitcoin exchange rate with other currency exchanges and stock market indices; they find that Bitcoin is 41 times as volatile as the USD–EUR exchange rate and 10 times as volatile as S&P 500 returns. This extreme return volatility makes it even more important to find predictors that suggest daily frequency, because such predictors could significantly improve the investment performance of Bitcoin investors and protect business adopters against future price swings.

Fourth, the Bitcoin market has limited depth and a lack of short-selling or derivative instruments. Bitcoin thus can be hard to trade; there are not enough buyers and sellers in an exchange. A buyer seeking to purchase a large amount of Bitcoin at the prevailing exchange rate often cannot do so quickly, without increasing the price. In addition, the largely undeveloped swap, forward, and other derivative markets make it difficult to hedge Bitcoin against the value of other currencies. With limited market depth, the price is more likely to reflect large buy and sell orders by specific investors who are speculating or liquidating. The lack of short selling also may create more arbitrage opportunities [28].

These unconventional characteristics make it difficult to characterize Bitcoin and digital currencies in general. According to the U.S. IRS, Bitcoin should be treated as property for federal tax purposes.[1] Testing Bitcoin in terms of the three functions of money—namely, a measure of exchange, store of value, and unit of account—Yermack [89] concludes that it faces challenges in meeting all three criteria. Glaser, et al. [30] examine users' motivations for holding Bitcoin and conclude that most users treat their Bitcoin investments as assets rather means of payment. Furthermore, the correlations between Bitcoin and other asset classes suggest that Bitcoin investments can offer diversification benefits [9]. Rather than attempt to reconcile the contradictory views held by regulatory agencies and researchers, we note that this ambiguity is not uncommon for a new technology [62]. A lack of understanding of the new technology can lead stakeholders into traps, such as deferring their participation due to uncertainty, committing too quickly, or failing to persist after an initial foray [18]. Social cues can have strong influences on the adoption and dissemination of such new technologies though [1], leading us to predict that the transformative power of digital currency, as a financial technology innovation, could be

[1] http://www.irs.gov/uac/Newsroom/IRS-Virtual-Currency-Guidance

harnessed more readily if we understood the social mechanisms underlying its adoption. Further, it is unlikely that all users share the same perspectives and motivation, so we analyze online digital traces to quantify the role of social media and dissect the distinctive impacts of different user cohorts and platforms. In so doing, we aim to enhance theoretical understanding of the emergence and adoption of new technologies, as well as help businesses prepare for the opportunities and threats brought about by this innovation.

2.2 Social Media's Impacts on Product Adoption and the Financial Market

In modern society, new information often becomes available through social media and Web 2.0 applications, which fundamentally have changed the interactions of consumers and firms [27]. For example, online word of mouth, consumer reviews, and blogs are prominent sources of information for consumer purchase decisions [12, 36]. Prior studies thus validate the relationship between digital user metrics and sales of products such as movies (Liu [50], Dellarocas, et al. [19], Duan, et al. [23], Chintagunta, et al. [14]), books (Forman, et al. [25], Chevalier and Mayzlin [13]), music (Dhar and Chang [22], Dewan and Ramaprasad [21]), television shows (Godes and Mayzlin [31]), video games (Zhu and Zhang [91]), and airline services (Luo [55], Luo [56]). However, the question of whether social media influence the underlying behavior of financial markets remains subject to ongoing debate. A sample of the 50 firms with the greatest Yahoo message board posting volume reveals that changes in daily posting volume are associated with both earnings announcement events and shifts in stock trading volume and returns [87]. In their study of Internet stocks, Tumarkin and Whitelaw [82] find that message board activity cannot predict stock returns, but instead, the causality appears to run from the market to the forums. In contrast, Antweiler and Frank [2] indicate that a positive shock to message board posting predicts negative stock returns on the next day. An analysis of articles published on a social media platform indicates that the views expressed, in both articles and comments, predict future stock returns and earnings surprises [11]. Although stock messages may reflect information quickly, they seemingly have no ability to forecast stock returns [17]. In another examination of the dynamic relationship between social media (consumer ratings and Web blogs) and firm equity value, social media metrics have significant predictive power for firm equity value [58]. These authors also examine the relative effects of social media metrics compared with conventional online behavior metrics (e.g. Google searches, Web traffic) and find that predictions based on social media are faster [8]. Overall then, the relationship between social media and financial market appears inconsistent across prior studies.

The inconsistent findings might result from the tendency in extant literature to treat social media as a single source, even though content gets generated on multiple platforms by users with varying intents. A few studies address contextual factors related to UGC. For example, Ludwig, et al. [54] show that a user's linguistic style correlates with posting quantity and quality. In a healthcare domain, physicians with poorer quality, as perceived by patients, are less likely to be reviewed online, and both high and low quality physicians are less likely to receive informative ratings online [29]. Yet critical gaps remain, especially with regard to the economic impacts of contextual differences in UGC in financial markets. Nor has any study, to the best of our knowledge, taken a systematic view of how social media content correlates with behavioral and platform factors.

To address these gaps, we consider Bitcoin, a financial technological innovation, through the lens of social media. The unique traits of Bitcoin enable us to develop and test novel hypotheses, unconstrained by the limitations of traditional financial market data. We combine econometric models and text mining methods to demonstrate *how* user contribution quantity and platform differences lead to distinct economic impacts, as well as identify *why* these differences occur.

3. HYPOTHESES DEVELOPMENT

3.1 Can Social Media Influence the Bitcoin Market?

Online messages often disclose new or private information that might fundamentally alter Bitcoin evaluations, such as when new stores accept Bitcoins or forthcoming regulations limit its use. Online discussions thus offer good indications of general market sentiment toward Bitcoin. The design of Bitcoin's algorithm also 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. In addition, speculative investors tend to follow trends, which may exaggerate the effects of any such information. Hence, we anticipate that the UGC on online platforms should affect the investment returns and trading volume of Bitcoins. In particular, the decentralized nature of Bitcoin meant that most early users were individuals, rather than large institutional investors. These early adopters arguably contribute to social media more frequently and are more likely to be influenced by social media. Further, an important motivation for early institutional Bitcoin adopters then 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" [70]. When the social gaming company Zynga added Bitcoin to its most popular games in 2014, it garnered thousands of media mentions for example. Finally, according to a recent survey [24], Bitcoin users largely exhibit the demographic characteristics of being heavy social media users. For these reasons, we postulate:

H1. Social media metrics have significant effects on future Bitcoin returns, such that increased positive (negative) sentiments indicate higher (lower) future Bitcoin prices.

3.2 Vocal Minority and Silent Majority

The power law nature of social media implies that most users contribute little content (silent majority), and a small proportion of highly active users contribute the most (vocal minority). This phenomenon has been empirically verified for Twitter [65], online reviews [68], Wikipedia [40], and websites in general [83]. Yet evidence about which cohort is more valuable in terms of reflecting market sentiments and affecting future returns remains ambivalent.

On the one hand, critical mass theory [59] 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. This is because for the online community, more posts are associated with a higher probability of becoming a leader [45]. Early Bitcoin adopters who also elect to post large amounts naturally should emerge as community leaders. Social network and word-of-mouth research highlight the importance of these influential users through social media. As Trusov, et al. [81] 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 [32, 33], partly because they have greater exposure to mass media than their followers [72].

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 [20]. For instance, groups of mutual funds tend to adopt the investment choices of their successful counterparts [26]. Jiao and Ye [41] 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 might be just as important, if not more so, than those of the vocal minority. First, by definition, silent majority users contribute to conversations sporadically, usually after highly significant events, and they are not particularly interested in generating buzz [61, 65]. The sentiments of the silent majority, as market measures, thus might tend to be more concise, relevant, and less noisy.

Second, the decentralized nature of Bitcoin has meant that most grassroots users can be categorized as the silent majority. If its value reflects the valuation of crowds, then the diversity prediction theorem [67, 76]—namely, that collective error diminishes as the diversity of 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 could outperform the collective of like-minded experts and fanatics.

Third, silent majority users are less likely to engage in groupthink [39], defined as self-deception, wishful thinking, and conformity to group values that leads to willful blindness and collective denial [5]. Bitcoin has been subjected to criticism that its value might depend on its most zealous users [60]. It is plausible that the vocal minority users engages in vast discussions of Bitcoin, gets caught up in the glorified ideas, and is more prone to groupthink. If so, they may hold biased views of the future value of the investment and denies any downside risk [78]. In sum, any or a combination of these mechanisms could lead to the result that UGC from the silent majority is a more compelling metric for actual investors. Taking both sides of the argument, we propose two competing hypotheses:

H2a. The vocal minority has a stronger impact than the silent majority on the Bitcoin market.

H2b. The silent majority has a stronger impact than the vocal minority on the Bitcoin market.

3.3 Platform Differences

In addition to user-level influences, we propose that various social media platforms affect financial markets differently, because the mechanisms of information diffusion, visibility, and representation differ. As examples, we use an Internet forum and Twitter, which differ in three main ways. First, Internet forums generally seek to achieve diverse opinions, and 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 might 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 at the same time. According to the theory of social exchange motivations [74], the lack of the latent benefit of publicity should suppress critical, in-depth discussions on Twitter. Therefore, forum discussions are more 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 lifecycle 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 non-recent tweet is much higher, which should reduce the efficiency of the market for information at the intraday level. In addition, finance scholars note that investors have limited attention capacities, so they respond asymmetrically to more visible information [3, 38].

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 needs to be condensed. A forum does not have this strict limitation. This condensing process creates two limitations in terms of analyzing the impacts of these social media. For one, adding external URLs to tweets is a common practice [15], and essential information then gets encapsulated in 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; it is clear only relative to other numerical information [73, 85]. To determine the full implication 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,

H3. User-generated content from Internet forums, rather than Twitter, has a stronger impact on the value of Bitcoin at a daily level.

4. DATA

4.1 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,[2] from Bitcoincharts.com. The exchange trading volume measure refers to the amount of Bitcoin traded for USD; transaction volume indicates the amount transferred within the Bitcoin economy. 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 *t*. The continuously compounded return in Bitcoin is the first difference of the log price:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln(S_t) - \ln(S_{t-1}) \,. \tag{1}$$

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 choose λ = .94, the same value used by RiskMetrics (owned by J.P. Morgan), 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.

4.2 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 [75], 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 343,769 posts and 15,420 topics to retain for further analysis. Each post contained textual

[2] 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.

content, an author, and a timestamp. Among the 17,215 unique users who posted, the most active 5% of users generated 63.11% 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.

[Insert Figure 1 about here]

For the sentiment analysis, we applied a finance sentiment dictionary [52], which includes 2,329 negative and 297 positive sentiment words. We used Natural Language Toolkit 3.0 [6] for the language processing tasks, such as sentence segmentation, word tokenization, and lemmatization, and 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 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 wider ranges 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–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 [52] 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.

4.3 Other Variables

We included a set of traditional Internet activity measures and exogenous 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 gathered 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.

[Insert Table 1 about here]

5. EMPIRICAL METHODOLOGY

To study the dynamic relationship between Bitcoin and social media, we use vector error correction models (VECM) to capture the inter-dependencies across time series. These models extend the vector autoregression (VAR) system when cointegration is present, meaning that there are long-term common stochastic trends among the non-stationary time series [43]. We choose VECM rather than a more traditional multiple regression (cf. Antweiler and Frank [2], Wysocki [87]) for several reasons. First, we can treat all of the key variables as jointly endogenous, without creating ad hoc model restrictions. Nor do we need 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 the variables. Third, we can interpret the estimated VECM model using Granger causality, and test whether the past values of social media variables are useful for predicting the Bitcoin market variables.

In our empirical study, we examine models in which the variables include daily observations of Bitcoin market activities, namely, returns (r_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). Lastly, we include the relevant control variables defined in Table 1.

To determine the appropriate models, we first test the stationarity of the variables. Conventional regression estimators, including VAR, encounter problems when applied to nonstationary processes, so the regression of two independent random walk processes would yield a spurious significant coefficient, even if they were not related [34]. We used an augmented Dickey-Fuller unit root test of each variable. Among the time series in the model, the number of positive/negative posts, VIX, and investor sentiments are stationary; the others have one order of integration. Next, we determined the appropriate lag length *p* using Akaike's information criterion, as is standard in econometrics literature [53].

In a VAR system, we can model the interrelationship of the variables by taking first differences of each non-stationary series and including the differences, yet this approach can suffer misspecification biases if cointegration is present. Instead, VECM yields more efficient estimators of cointegrating vectors. We performed a Johansen test [42] and confirmed the presence of cointegration in our daily frequency data. Therefore, VECM is the appropriate model for our data. It 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 [44]. We estimated the order of cointegration *r* of the model using Johansen's multiple trace test procedure.

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

$$\Delta \boldsymbol{Y}_t = \sum_{j=1}^{k-1} \Gamma_j \Delta \boldsymbol{Y}_{t-k} + \alpha\beta' \boldsymbol{Y}_{t-1} + \boldsymbol{\mu} + \epsilon_t, \tag{2}$$

where Δ is the first difference operator, $\boldsymbol{Y}_t$ is a $p\times1$ vector with order of integration 1, $\boldsymbol{\mu}$ is a $p\times1$ constant vector representing the linear trend, k is the a lag structure, 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 model 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 and can be 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.

6. RESULTS

6.1 VECM Analyses

To test H1, we examine the effects of the bullishness of forum messages using a VECM. The model includes daily measures of the Bitcoin market variables r, 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 information criteria. The trace test reveals five cointegrating relations in the system. Table 2 presents the estimated coefficients in the VECM, highlighting the relationship between social media metrics and Bitcoin market variables. We thus can observe several characteristics of the Bitcoin market. First, the relationship among Bitcoin market variables is consistent with findings from stock market studies. Trading and transaction volume exhibit a strong autoregressive relationship, such that high trading (transaction) volume days tend to precede days of high trading (transaction) volume. Trading volume also helps predict volatility. Second, the two forum metrics work as we predicted in H1. 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. To confirm this result, we also performed a Granger causality test between Bitcoin returns and lagged social media metrics. The social media metrics are individually ($\chi^2 = 6.37$, $p = 0.012$ for POS^F; $\chi^2 = 5.48$, $p = 0.019$ for NEG^F) and jointly ($\chi^2 = 7.00$, $p = 0.030$) significant, meaning the past values of forum sentiments are useful for predicting r. Finally, the Google Trend measure is the only control variable that affects future Bitcoin returns. Therefore, forum posts contain new information about the 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.

[Insert Table 2 about here]

To examine H2, 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% by posting volume); the other model uses the vocal minority (top 5% by posting volume). The silent minority generated a mere 36.89% of the messages, whereas the vocal majority generated 63.11%. Table 3 presents the split sample results. For the posts by the silent majority, the estimates for their impacts on returns are much greater than those in the full sample model (Table 2). The impact of their posts on Bitcoin returns grows stronger, even though posts from these users account for a smaller proportion of the total posting volume. Posts by the vocal minority instead provide indicators of future transaction volumes only, not returns. The coefficient estimates for POS^F and NEG^F are lower in value and not statistically significant. Overall, these results lend support to H2b: The return predictability available from social media mostly depends on content created by the silent majority.

[Insert Table 3 about here]

Having established the overall impact of social media and the stronger effects of the silent majority users' sentiment on Bitcoin returns, we can study platform differences and test H3. To determine if 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 returns than are Twitter messages. The Granger causality tests confirm this finding, such that the forum sentiment of the previous day Granger-causes changes in future Bitcoin prices ($\chi^2 = 18.58$, $p < 0.001$), whereas there is no Granger causality from Twitter sentiment to daily Bitcoin returns ($\chi^2 = 2.60$, $p = 0.27$).

[Insert Table 4 about here]

6.2 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 bullishness, in line with an existing method for aggregating message clarifications [2]. We define the bullishness measure as *(POS – NEG)/(POS +NEG)* and re-estimate our models using this single measure. Table 5 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.

[Insert Table 5 about here]

To ensure that our results are not driven by certain events in a specific timeframe, we divided our sample periods into two equivalent sub-periods and estimated our model again. The coefficient estimates in Table 6 largely are consistent with our main findings, 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% and 2.5% user activity cutoff levels, in addition to the 5% level in our main study. The results in Table 7 show that the impacts of the vocal minority and silent majority exhibit similar disparities with the new definitions: Posts from less active users carry more weight for indicating future price changes.

[Insert Table 6 and 7 about here]

7. MECHANISMS BEHIND THE DIFFERENT EFFECTS

Having established the heterogeneous effects of social media measures due to differences in user posting quantity and platform, we now turn to understanding the mechanisms that prompt these differences, through an analysis of the content of the discussions and users' behavior. Our goal is to present as much evidence as possible from our data, rather than select specific data, to understand the observed phenomena.

7.1 Explaining Vocal Minority vs. Silent Majority

Why does the opinion of the silent majority exert a stronger influence on future Bitcoin returns? We previously laid out three possible explanations in Section 3.2. The first explanation is the "relevance hypothesis": It is possible that silent majority group is concerned with aspects that are more relevant to our dependent variables – the monetary value of Bitcoin. The second explanation is the "diversity perspective hypothesis": the silent majority group consists of more users and therefore represents a more diverse view. A final explanation is the "groupthink hypothesis". The psychological drive for consensus is less prominent for less active users and therefore, their opinion is more unbiased. Here, we detail the evidence in support of and contrary to each of these three explanations, and thereby determine that the data are most consistent with the relevance and diversity perspective hypotheses.

The relevance explanation states that the silent majority group discusses more relevant content related to the monetary value of Bitcoin. We verified this explanation with an analysis of the textual content of posts using structural topic model (STM) [71]. As a recent development in machine learning with textual data, STM incorporates metadata about the document (e.g. attributes of the authors) into content analysis. STM makes two modifications to the notable latent Dirichlet allocation (LDA) model [7]: it allows topic proportions to correlate and permits the prevalence of the topics to depend on a set of covariates. Thus we can determine the major semantic themes (topics) in a collection of text and whether the topic proportions vary across groups. (For technical details, see Roberts, et al. [71].) We use STM to study whether and how the topical prevalence of posts differs when users are in the silent majority or vocal minority.

Specifically, with STM, we discovered seven topics in the forum discussions, which we list in Table 8. We chose the number of topics using the standard of semantic coherence [63], and we used high probability and high lift words [77] to label the topics. In Figure 2, to compare the prevalence of topics across groups, we depict the point estimates and 95% confidence intervals for the intergroup differences for each topic. Silent majority users are more likely to discuss exchange, payment, content sharing, and social acceptance topics, and the discrepancies are statistically significant. Because the supply rate for Bitcoin is controlled by design [66], most shifts in its value can be attributed to demand side effects. Therefore, the demand-side related topics favored by the silent majority such as adoption as a payment system, appeal as an investment vehicle on exchanges, and acceptance in the society are more likely to correlate with higher future returns.

The theoretical roots of the topics favored by the silent majority also reflect perceived usefulness (PU) and subjective norm (SN) constructs—two key predictors in the extended technology acceptance model (TAM2) [84]. For example, exchange and payment reflect the perceived usefulness of the primary functions of Bitcoin. Social influences are closely related to subjective norms, in the sense that users tend to accept a technology if others have done so. According to the TAM2, positive sentiments along these lines should drive the acceptance of Bitcoin and pertain to its higher valuation. Overall, the results lend support to the relevance explanation.

[Insert Table 8 about here]

[Insert Figure 2 about here]

The diversity perspective also emerges as pertinent. In this case, since discussions in internet forums are separated into threads under which a single subject is discussed, we constructed a content diversity measure for each user, using the average number of threads each user participated in on the Internet forum divided by the number of posts by this user. This measure takes a value between 0 and 1. The greater the diversity measure is, the more likely a different post by the same user addresses a different subject. The average diversity measure is 0.549 (SD = 0.188) for the vocal minority group but 0.813 (SD = 0.251) for the silent majority group. A t-test rejects the null hypothesis that the diversity measure is the same across two groups ($p < 0.001$). This finding also is congruent with previous research suggesting that diverse ideas result in better predictions [51, 67].

In contrast with our prediction that the vocal minority group is more likely to engage in groupthink, we found evidence that is either equivocal or even contests the groupthink explanation. We follow [54], who indicate that linguistics symbolically reflect the feelings and subsequent behaviors of participants in online communities. Specifically, we used Linguistic Inquiry and Word Count (LIWC) to measure psychological processes, based on word use [69], by counting the percentages of words that belong to certain linguistic and psychological categories. The LIWC's word categories have been psychometrically validated and widely applied to measure social behavior [79]. Yilmaz and Peña [90] argue that lack of negation in language can reveal groupthink, in that it indicates avoidance of divergent thinking processes. In other words, members simply exchange agreement, without thoroughly discussing other evidence or possibilities. In addition, Taffler and Tuckett [78] point out that uncertainty and anxiety are closely related to groupthink in financial markets. Therefore, we ran LIWC on all forum posts to calculate the average *negation*, *tentative*, and *anxiety* scores by user, then compared the scores for users in the silent majority versus the vocal minority group. In contrast with our groupthink explanation, the silent majority group exhibited significantly lower ($p < 0.001$) *negation* scores (M = 2.071, SD = 4.071) than the vocal minority (M = 2.215, SD = 4.154). The two groups did not differ statistically significantly in *tentative* ($p = 0.433$) or *anxiety* ($p = 0.609$) scores. We did not find evidence that supports the groupthink hypothesis.

7.2 Explaining Forum vs. Twitter

From our data, we found empirical support for all three main reasons we previously offered to explain the stronger impact of the forum, relative to Twitter, on returns. In particular, the limited interaction on Twitter is exemplified by the finding that only 2.3% of tweets are replies, whereas in the forum data, 95.6% of posts are replies. Thus, any topic is more likely to be repeatedly discussed, reinforced, or rebutted on the forum, making it a more appropriate platform from which to glean crowd sentiments after an interaction has taken place.

Relatedly, the limited interactions on Twitter usually happen shortly after an original tweet gets posted; the median time between a reply and the original tweet is only 16.6 minutes. On the forum, this measure is 889.0 minutes. The stark contrast is a reflection of, for better or worse, the timeliness of Twitter. The design of this microblog hinders user interactions in relation to earlier posted messages, though it also encourages viral transmission of intraday breaking news. To test the timeliness explanation explicitly, we estimated an intraday VECM that includes Bitcoin market variables, forum sentiments, and Twitter sentiments measured hourly. The Granger causality tests indicate that Twitter sentiment measures provide significant information about the next hour's return ($\chi^2 = 6.22$, $p = 0.045$), whereas there is no causality from forum sentiment measures ($\chi^2 = 0.07$, $p = 0.97$). Therefore, Twitter only reflects information rapidly and at shorter time intervals. Due to data availability though, we could not control for whether Twitter sentiments offer information in addition to other possible channels, such as news or traditional Internet activities. Our hourly result also serves as a falsification test, ruling out the possibility that our sentiment measures only work for forum data but not for Twitter.

Finally, the information content is condensed on Twitter, with considerable information embedded in the external links within tweets. Using a regular expression matcher, we found that 80.4% of the tweets in our sample contained external URLs, whereas only 13.1% of the forum messages do. These embedded URLs have two key implications. First, to capture the information transmitted in tweets, it is necessary to backtrack the links to external websites. Second, posting tweets with URLs is a defining characteristic of bots on Twitter [67]; our observation of such as high proportion of tweets with URLs implies that Twitter sentiment is a better indicator of up-to-date information but not of market participants' sentiment. Similarly, numerical expressions are more pervasive in tweets. When we compare the percentage of characters that are digits (0-9) in tweets and forum posts, we find 3.7% of the total tweet volume is numerical, whereas only 0.79% is for the forum. The higher proportion of numerical information, as we argued previously, may undermine readers' ability to understand the literal interpretation of a message. This limitation applies to human investors and sentiment classification algorithms alike.

8. DISCUSSION AND CONCLUSIONS

8.1 Findings and Implications

Even amid concerns about its volatile value, Bitcoin can provide unique benefits, including lower transaction costs, the potential to combat poverty and oppression, and stimuli for financial innovation [10]. By breaking down existing payment barriers and liberating global trades, Bitcoin could generate enormous wealth and social welfare for the economy. Today, the diffusion of a financial technology innovation of this scale—or of any significant technology innovation for that matter—cannot be completed without social media. We accordingly seek to assess and quantify the dynamic relationship of social media with the monetary value of Bitcoin. To the best of our knowledge, this study is the first effort to connect these two IT artifacts together. The results suggest that social media sentiment is an important indicator of future Bitcoin returns. Yet the relationship also is complex, because the silent minority exerts a more significant effect, and forum sentiment appears to be a better indicator at the inter-day level than tweets. We thus lend credence to the TAM in a new setting and show that the silent minority, engaged in diverse discussions, is more relevant to the acceptance of Bitcoin. The discussions on Twitter are more timely but also less interactive and more condensed. Thus, both the content and how this content is presented correlate with observed user behavior. By offering insights into the drivers of new financial technologies, this study illuminates the fine-grained relationships among social media content characteristics and their economic impacts.

The findings also have implications for virtual currency adopters, investors, and policy makers. First, social media offer substantial information about Bitcoin's acceptance among the general public, as well as daily fluctuations in its market sentiments. These signals factor in to the price forming process and influence future returns. Investors thus can discern Bitcoin's value from this rich information source, and this greater predictability relative to virtual currency returns 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 and carefully evaluate their decision to adopt Bitcoin payments. It must involve more than the marketing consideration of the potential for generating positive buzz. The dynamic relationship between social media content and Bitcoin value means the future value of accounts receivable can 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 correctly, social media also can drive profitable innovations in the future.

Although we conduct our study in the context of a financial market (i.e., Bitcoin economy), the broader implications also can influence business practices in online social media. Our results suggest ways to help consumers address innovations; in particular, companies should analyze user behaviors and textual contributions in social media communication more closely. We show that social media messages are not created equal and should not be treated equally. The practice of exploiting emotions and influences for marketing purposes is not novel; businesses have long recognized the value of lead users [86] and opinion leaders [46] for example. But our empirical findings highlight the value of the silent, yet salient majority of inactive users. Compared with the vocal minority that dominates UGC, silent majority users are more diverse and appear to care about more pragmatic aspects. Their opinion cannot be overlooked. More marketing and analytic efforts should seek to identify this "heavy tail" of the online world. Moreover, companies are advised to monitor discussions on various social platforms and devise unique strategies for them. For example, instantaneous buzz on mobile-oriented media (e.g., Twitter) may prompt quick actions, but in-depth discussions on Internet forums can paint a more comprehensive picture for participants and thus are more likely to trigger final adoption or purchase decisions.

8.2 Limitations and Further Work

Our research has several limitations in its data sources and analysis methods, which suggest extensions to this study. We used secondary data to identify the association between social media sentiments and future bitcoin returns. Well-designed, randomized experiments could enhance our understanding of the specific findings. We also report on the dynamics between social media and Bitcoin, a single cryptocurrency. Additional research could investigate whether and to what extent our findings generalize to other financial technology products. Finally, 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 and contrasting messages written in other languages may lead to insights about the potential effects of cultural differences.

In our data analysis, we used financial sentiments as a sole indicator of information in social media. Further studies might identify more subtle human emotions (e.g., fear, surprise) in the textual data and investigate their role. The mechanisms we proposed to explain the prominence of the silent majority and the stronger impact of the forum messages also are not exhaustive. Subsequent analyses of user social networks and information diffusion may create new perspectives. Finally, our research highlights the unexpected status of the silent majority, so researchers should work to devise data-driven algorithms that can dynamically select the most impactful social media users—not necessarily the most active ones. Given the volume and velocity of modern social media data, such a sampling process might reduce noise in the model significantly.

REFERENCES

1. Abrahamson, E., and Rosenkopf, L. Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science*, 8, 3 (1997), 289-309.
2. Antweiler, W., and Frank, M.Z. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59, 3 (2004), 1259-1294.
3. 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.
4. Barber, S., Boyen, X., Shi, E., and Uzun, E. Bitter to better—how to make bitcoin a better currency. *Financial cryptography and data security*: Springer, 2012, pp. 399-414.
5. Bénabou, R. Groupthink: Collective delusions in organizations and markets. *The Review of Economic Studies* (2012).
6. Bird, S. Nltk: The natural language toolkit. *The COLING/ACL on Interactive Presentation Sessions*, Sydney, Australia, July, 2006.
7. Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3 (2003), 993-1022.
8. Böhme, R., Christin, N., Edelman, B., and Moore, T. Bitcoin: Economics, technology, and governance. *The Journal of Economic Perspectives*, 29, 2 (2015), 213-238.
9. Brière, M., Oosterlinck, K., and Szafarz, A. Virtual currency, tangible return: Portfolio diversification with bitcoin. *Journal of Asset Management*, 16, 6 (2015), 365-373.
10. Brito, J., and Castillo, A. *Bitcoin: A primer for policymakers*. Mercatus Center at George Mason University, 2013.
11. 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.
12. Chen, Y., and Xie, J. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54, 3 (2008), 477-491.
13. Chevalier, J.A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43, 3 (2006), 345-354.
14. Chintagunta, P.K., Gopinath, S., and Venkataraman, S. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29, 5 (2010), 944-957.
15. 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.
16. CoinDesk. *State of bitcoin and blockchain 2016*. Jan 28, 2016. Avavailable at http://www.coindesk.com/state-of-bitcoin-blockchain-2016/.

17. Das, S.R., and Chen, M.Y. Yahoo! For amazon: Sentiment extraction from small talk on the web. *Management Science*, 53, 9 (2007), 1375-1388.
18. Day, G.S. Organizing for interactivity. *Journal of Interactive Marketing*, 12, 1 (1998), 47-53.
19. Dellarocas, C., Zhang, X.M., and Awad, N.F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21, 4 (2007), 23-45.
20. Devenow, A., and Welch, I. Rational herding in financial economics. *European Economic Review*, 40, 3 (1996), 603-615.
21. Dewan, S., and Ramaprasad, J. Music blogging, online sampling, and the long tail. *Information Systems Research*, 23, 3-part-2 (2012), 1056-1067.
22. Dhar, V., and Chang, E.A. Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23, 4 (2009), 300-307.
23. Duan, W., Gu, B., and Whinston, A.B. Do online reviews matter?—an empirical investigation of panel data. *Decision Support Systems*, 45, 4 (2008), 1007-1016.
24. Duggan, M., and Brenner, J. *The demographics of social media users, 2012*. Washington, DC: Pew Research Center's Internet & American Life Project, Feb 14, 2013, 2013. Avavailable at http://www.lateledipenelope.it/public/513cbff2daf54.pdf.
25. Forman, C., Ghose, A., and Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19, 3 (2008), 291-313.
26. Friend, I., Blume, M., Crockett, J., and Fund, T.C. *Mutual funds and other institutional investors: A new perspective*. McGraw-Hill, 1970.
27. Gallaugher, J., and Ransbotham, S. Social media and customer dialog management at starbucks. *MIS Quarterly Executive*, 9, 4 (2010), 197-212.
28. Gandal, N., and Halaburda, H. Competition in the crypto-currency market. *Workshop on the Economics of Information Security*, State College, PA, June, 2014.
29. 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.
30. Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C., and Siering, M. Bitcoin - asset or currency? Revealing users' hidden intentions. *Proceedings of the 22nd European Conference on Information Systems*, Tel Aviv, Israel, 2014.
31. Godes, D., and Mayzlin, D. Using online conversations to study word-of-mouth communication. *Marketing Science*, 23, 4 (2004), 545-560.
32. Godes, D., and Mayzlin, D. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28, 4 (2009), 721-739.

33. 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.
34. Granger, C.W., and Newbold, P. Spurious regressions in econometrics. *Journal of econometrics*, 2, 2 (1974), 111-120.
35. Grinberg, R. Bitcoin: An innovative alternative digital currency. *Hastings Science & Technology Law Journal*, 4 (2011), 160-208.
36. Gu, B., Park, J., and Konana, P. Research note-the impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research*, 23, 1 (2012), 182-196.
37. Harris, M., and Raviv, A. Differences of opinion make a horse race. *Review of Financial studies*, 6, 3 (1993), 473-506.
38. Hirshleifer, D., and Teoh, S.H. Thought and behavior contagion in capital markets. *Handbook of financial markets: Dynamics and evolution*, 2009, pp. 1-46.
39. Janis, I. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. *Boston: Houghton Mifflin*, 3 (1972), 4.
40. Javanmardi, S., Ganjisaffar, Y., Lopes, C., and Baldi, P. User contribution and trust in wikipedia. *Proceedings of 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2009, pp. 1-6.
41. Jiao, Y., and Ye, P. Mutual fund herding in response to hedge fund herding and the impacts on stock prices. *Working Paper*, 2013. Avavailable at http://faculty.ucr.edu/~yawenj/mfhf.pdf.
42. 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.
43. Johansen, S. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society* (1991), 1551-1580.
44. Johansen, S. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, 1995.
45. Johnson, S.L., Safadi, H., and Faraj, S. The emergence of online community leadership. *Information Systems Research*, 26, 1 (2015), 165-187.
46. King, C.W., and Summers, J.O. Overlap of opinion leadership across consumer product categories. *Journal of Marketing Research* (1970), 43-50.
47. Kristoufek, L. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, 3 (2013).
48. Kristoufek, L. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PloS ONE*, 10, 4 (2015), e0123923.

49. Kroll, J.A., Davey, I.C., and Felten, E.W. The economics of bitcoin mining, or bitcoin in the presence of adversaries. *Proceedings of the Twelfth Workshop on the Economics of Information Security*, Washington, DC, 2013.
50. Liu, Y. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70, 3 (2006), 74-89.
51. Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108, 22 (2011), 9020-9025.
52. Loughran, T., and McDonald, B. Measuring readability in financial disclosures. *The Journal of Finance*, 69, 4 (2014), 1643-1671.
53. Love, I., and Zicchino, L. Financial development and dynamic investment behavior: Evidence from panel var. *The Quarterly Review of Economics and Finance*, 46, 2 (2006), 190-210.
54. Ludwig, S., De Ruyter, K., Mahr, D., Wetzels, M., Bruggen, 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.
55. Luo, X. Consumer negative voice and firm-idiosyncratic stock returns. *Journal of Marketing*, 71, 3 (2007), 75-88.
56. Luo, X. Quantifying the long-term impact of negative word of mouth on cash flows and stock prices. *Marketing Science*, 28, 1 (2009), 148-165.
57. 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.
58. Luo, X., Zhang, J., and Duan, W. Social media and firm equity value. *Information Systems Research*, 24, 1 (2013), 146-163.
59. Marwell, G., and Oliver, P. *The critical mass in collective action*. Cambridge University Press, 1993.
60. McCrum, D. *Bitcoin's place in the long history of pyramid schemes*. Financial Times: Nov. 10, 2015. Retrieved from http://www.ft.com/intl/cms/s/0/1877c388-8797-11e5-90de-f44762bf9896.html#axzz45HHE2ak3, on April 8, 2016.
61. Metaxas, P.T., and Mustafaraj, E. Trails of trustworthiness in real-time streams (extended summary). *Proceedings on Design, Influence and Social Technologies Workshop*, Seattle, WA, 2012.
62. Milliken, F.J. Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. *Academy of Management review*, 12, 1 (1987), 133-143.
63. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., and McCallum, A. Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*: Association for Computational Linguistics, 2011, pp. 262-272.

64. Moore, T., and Christin, N. Beware the middleman: Empirical analysis of bitcoin-exchange risk. *Financial cryptography and data security*: Springer, 2013, pp. 25-33.
65. Mustafaraj, E., Finn, S., Whitlock, C., and Metaxas, P.T. Vocal minority versus silent majority: Discovering the opionions of the long tail. *Proceedings of 2011 IEEE International Conference on Privacy, Security, Risk and trust and 2011 IEEE International Conference on Social Computing*, Boston, MA: IEEE, 2011, pp. 103-110.
66. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. 2008.
67. Page, S.E. Making the difference: Applying a logic of diversity. *The Academy of Management Perspectives*, 21, 4 (2007), 6-20.
68. Peddibhotla, N.B., and Subramani, M.R. Contributing to public document repositories: A critical mass theory perspective. *Organization Studies*, 28, 3 (2007), 327-346.
69. Pennebaker, J.W., Boyd, R.L., Jordan, K., and Blackburn, K. The development and psychometric properties of liwc2015. *UT Faculty/Researcher Works* (2015).
70. PricewaterhouseCoopers LLP. *Digital disruptor: How bitcoin is driving digital innovation in entertainment, media and communications*. May 2014, 2014. Avavailable at http://www.pwc.com/en_SG/sg/tmt/assets/tmtnews201402/digital_disruptor.pdf.
71. Roberts, M.E., Stewart, B.M., Tingley, D., and Airoldi, E.M. The structural topic model and applied social science. *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*, 2013.
72. Rogers, E.M. *Diffusion of innovations*. Simon and Schuster, 2010.
73. Scammon, D.L. “Information load” and consumers. *Journal of Consumer Research*, 4, 3 (1977), 148-155.
74. Shi, Z., Rui, H., and Whinston, A.B. Content sharing in a social broadcasting environment: Evidence from twitter. *MIS Quarterly*, 38, 1 (2014), 123-A126.
75. Smyth, L. The demographics of bitcoin. Sumulacrum, 2013.
76. Surowiecki, J. *The wisdom of crowds*. Anchor, 2005.
77. Taddy, M. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108, 503 (2013), 755-770.
78. Taffler, R.J., and Tuckett, D.A. Emotional finance: The role of the unconscious in financial decisions. *Behavioral finance. Investors, corporations, and markets* (2010), 95-112.
79. Tausczik, Y.R., and Pennebaker, J.W. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29, 1 (2010), 24-54.
80. Tirunillai, S., and Tellis, G.J. Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, 31, 2 (2012), 198-215.

81. 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.
82. Tumarkin, R., and Whitelaw, R.F. News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57, 3 (2001), 41-51.
83. Venkataraman, M., Subbalakshmi, K.P., and Chandramouli, R. Measuring and quantifying the silent majority on the internet. *Sarnoff Symposium (SARNOFF), 2012 35th IEEE*, 2012, pp. 1-5.
84. Venkatesh, V., and Davis, F.D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46, 2 (2000), 186-204.
85. Viswanathan, M., and Childers, T.L. Processing of numerical and verbal product information. *Journal of Consumer Psychology*, 5, 4 (1996), 359-385.
86. Von Hippel, E. Lead users: A source of novel product concepts. *Management science*, 32, 7 (1986), 791-805.
87. Wysocki, P.D. Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, 1999. Avavailable at http://ssrn.com/abstract=160170.
88. Yang, S. *Bitcoin catches on with gold bugs*. The Wall Street Journal: May 22, 2016. Retrieved from http://www.wsj.com/articles/bitcoin-catches-on-with-gold-bugs-1463959897, on June 23, 2016.
89. Yermack, D. Is bitcoin a real currency? An economic appraisal. *National Bureau of Economic Research Working Paper Series*, 2013. Avavailable at http://www.nber.org/papers/w19747.
90. Yilmaz, G., and Peña, J. How do interpersonal behaviors and social categories affect language use?: The case of virtual teams. *Communication Quarterly*, 63, 4 (2015), 427-443.
91. Zhu, F., and Zhang, X. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74, 2 (2010), 133-148.

Table 1 Key measures and summary statistics

Variables	Meaning	Mean	SD	Median	Min	Max
Bitcoin Market Variables						
r	Bitcoin returns	0.00375	0.0525	0.00114	-0.478	0.359
σ	Volatility of Bitcoin returns	0.0445	0.0293	0.0382	0.00958	0.209
V	Log daily trading volume	11.96	0.478	11.97	10.58	13.74
V^{TX}	Log daily transaction volume	14.40	1.505	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	Log Bitcoin.org web traffic rank	9.655	0.936	9.485	7.144	11.64
googletrend	Google Trend for Bitcoin	16.33	18.53	12	2	100
sp500	Log S&P 500 closing price	7.399	0.145	7.408	7.152	7.645
vix	COBE Volatility Index	15.33	2.884	14.68	10.32	26.66
gold	Log COMEX gold price	6.729	0.133	6.687	6.500	6.953
investor_sentiment	AAII investor sentiment	13.70	8.941	0	0	38.60
news_sentiment	TRNA Bitcoin news sentiment	0.0205	0.142	13.40	-0.763	0.810

Table 2 VECM estimates for forum sentiments and Bitcoin

	Dependent Variables					
Independent Variables	**r**	**Δσ**	**ΔV**	**ΔV^{TX}**	**ΔPOS^F**	**ΔNEG^F**
r(t-1)	0.138***	-0.00688***	-0.0166	0.152	-3.279	-12.94
	(0.0305)	(0.00167)	(0.346)	(0.190)	(10.48)	(20.62)
Δσ(t-1)	0.380	0.140***	4.464	-4.203	199.4	-139.6
	(0.544)	(0.0298)	(6.165)	(3.382)	(186.8)	(367.5)
ΔV(t-1)	-0.00859**	0.000576***	-0.209***	0.128***	-2.539*	-2.510
	(0.00382)	(0.000210)	(0.0433)	(0.0238)	(1.313)	(2.582)
ΔV^{TX}(t-1)	-0.000284	0.000248	0.304***	-0.207***	12.57***	14.17***
	(0.00574)	(0.000315)	(0.0651)	(0.0357)	(1.972)	(3.878)
ΔPOS^F (t-1)	0.000353**	-8.15e-06	6.02e-05	0.00364***	-0.144***	-0.0673
	(0.000140)	(7.68e-06)	(0.00159)	(0.000870)	(0.0481)	(0.0945)
ΔNEG^F (t-1)	-0.000163**	-1.61e-06	-0.000466	-0.000394	-0.0258	-0.136***
	(6.98e-05)	(3.83e-06)	(0.000791)	(0.000434)	(0.0240)	(0.0472)
Δrank(t-1)	0.0122	-0.000590	-0.0293	0.0857	-2.587	-10.45*
	(0.00890)	(0.000488)	(0.101)	(0.0553)	(3.058)	(6.014)
Δgoogletrend(t-1)	0.00181***	5.67e-06	0.0118*	-0.00212	-0.576***	-0.942***
	(0.000536)	(2.94e-05)	(0.00607)	(0.00333)	(0.184)	(0.362)
Δsp500(t-1)	-0.557	0.0212	9.895*	5.340*	345.8**	911.5***
	(0.448)	(0.0246)	(5.086)	(2.790)	(154.1)	(303.2)
Δvix(t-1)	-0.00211	0.000192	0.0554	0.0504***	2.495**	5.236***
	(0.00298)	(0.000164)	(0.0338)	(0.0186)	(1.026)	(2.017)
Δgold(t-1)	0.0660	0.0149	0.770	-1.125	30.00	180.5
	(0.170)	(0.00933)	(1.928)	(1.058)	(58.42)	(114.9)
Δinvestor_sent(t-1)	0.000642	4.76e-06	-0.00419	0.000941	0.0466	-0.134
	(0.000461)	(2.53e-05)	(0.00523)	(0.00287)	(0.158)	(0.312)
Δnews_sent(t-1)	-0.00266	0.000232	0.212	-0.0223	-2.284	-15.63
	(0.0159)	(0.000875)	(0.181)	(0.0991)	(5.477)	(10.77)

Notes: N = 1901. 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 3 VECM estimates for comparing the silent majority and vocal minority

	Dependent Variables							
	r		**Δσ**		**ΔV**		**ΔV^{TX}**	
Independent Variables	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
r(t-1)	0.136***	0.143***	-0.00679***	-0.00707***	0.0836	-0.0525	0.175	0.186
	(0.0305)	(0.0305)	(0.00168)	(0.00167)	(0.347)	(0.345)	(0.191)	(0.189)
Δσ(t-1)	0.417	0.333	0.139***	0.142***	3.523	4.869	-4.314	-4.491
	(0.543)	(0.544)	(0.0299)	(0.0298)	(6.181)	(6.159)	(3.399)	(3.379)
ΔV(t-1)	-0.00915**	-0.00809**	0.000551***	0.000594***	-0.218***	-0.209***	0.138***	0.125***
	(0.00379)	(0.00383)	(0.000209)	(0.000210)	(0.0432)	(0.0433)	(0.0237)	(0.0238)
ΔV^{TX}(t-1)	-0.000159	-0.00165	0.000190	0.000289	0.313***	0.311***	-0.217***	-0.197***
	(0.00570)	(0.00572)	(0.000313)	(0.000313)	(0.0649)	(0.0647)	(0.0357)	(0.0355)
ΔPOS^F (t-1)	0.000874***	0.000250	-1.79e-05	-8.89e-06	-0.00252	0.00115	0.00457***	0.00545***
	(0.000261)	(0.000209)	(1.44e-05)	(1.14e-05)	(0.00297)	(0.00236)	(0.00164)	(0.00130)
ΔNEG^F (t-1)	-0.000427***	-0.000131	3.65e-06	-5.93e-06	-0.000595	-0.000993	9.02e-05	-0.000558
	(0.000152)	(0.000103)	(8.34e-06)	(5.62e-06)	(0.00173)	(0.00116)	(0.000950)	(0.000638)

Notes: N = 1901. 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 4 VECM estimates for comparing forum and twitter

	Dependent Variables			
Independent Variables	**r**	**Δσ**	**ΔV**	**ΔV^{TX}**
r(t-1)	-0.0773	0.0130	-1.736	-0.434
	(0.109)	(0.00801)	(1.752)	(1.031)
Δσ (t−1)	0.128	0.135	21.75	13.32
	(1.847)	(0.136)	(29.75)	(17.51)
ΔV(t-1)	-0.0223	0.00196*	0.349	0.415***
	(0.0159)	(0.00117)	(0.256)	(0.151)
ΔV^{TX}(t-1)	0.0143	-0.00205	0.108	-0.466**
	(0.0198)	(0.00146)	(0.319)	(0.188)
ΔPOS^T (t-1)	0.0123	-0.000654	0.00665	0.0409
	(0.00766)	(0.000564)	(0.123)	(0.0726)
ΔNEG^T (t-1)	-0.00385	0.000721*	-0.0299	-0.0501
	(0.00500)	(0.000368)	(0.0805)	(0.0474)
ΔPOS^F (t-1)	0.0221***	0.000786	0.119	0.0272
	(0.00817)	(0.000602)	(0.132)	(0.0775)
ΔNEG^F (t-1)	-0.0363***	-0.000231	-0.159	0.0329
	(0.00854)	(0.000629)	(0.137)	(0.0809)

Notes: N = 89. 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 5 Robustness checks using the alternative sentiment measure

	(1) All Users	(2) Silent Majority	(3) Vocal Minority	(4) All Users
Forum Bullishness	0.000114**	0.000282***	8.35e-05	0.0120***
	(4.45e-05)	(8.17e-05)	(6.96e-05)	(0.00444)
Twitter Bullishness				0.00330
				(0.00467)

Notes: This table shows the VECM estimates of previous day's social media bullishness on Bitcoin returns. N = 1901 for Models 1–3; N = 89 for Model 4. Standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 6 Robustness checks: subsamples by periods

	Jan 2012–June 2013		July 2013–Dec 2014	
	r		r	
Independent Variables	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
ΔPOS^F (t-1)	0.000982***	0.000267	0.000783**	0.000516
	(0.000335)	(0.000245)	(0.000377)	(0.000400)
ΔNEG^F (t-1)	-0.000439**	-0.000112	-0.000701**	-0.000417*
	(0.000178)	(0.000116)	(0.000285)	(0.000224)

Notes: Standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 7 Robustness checks: posting volume thresholds

	Cutoff = top 10%		Cutoff = top 2.5%	
	r		r	
Independent Variables	Silent Majority	Vocal Minority	Silent Majority	Vocal Minority
ΔPOS^F (t-1)	0.000995***	0.000364**	0.000805***	0.000139
	(0.000363)	(0.000181)	(0.000222)	(0.000248)
ΔNEG^F (t-1)	-0.000467**	-0.000145	-0.000328***	-0.000142
	(0.000209)	(8.91e-05)	(0.000124)	(0.000120)

Notes: Standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 8 Topics in the forum discussion

Topic	Label	Words with High Probability	Words with High Lift
1	Exchange	account, business, mtgox, sell, exchanges, company, funds	wagner, pierce, filed, brock, paxum, lawsuit, kong
2	Payment	fiat, payment, accept, cash, banks, card, online	nail, debit, card, grocery, gyft, mortar, credit
3	Blockchain	users, blockchain, private, security, client, trust, public	hearn, client, blockchain, addresses, nodes, node, private
4	Policy and History	gold, satoshi, real, years, government, person, name	hearn, coinjoin, zerocoin, alice, trustless, colored, blacklist
5	Mining	power, miners, number, amount, fees, current, blocks	corrected, variance, hashrate, difficulty, subsidy, gpus, asic
6	Content Sharing	great, thread, read, site, help, thanks, find	clearing, audio, clicked, banner, phinnaeus, podcast, videos
7	Social Acceptance	everyone, trying, start, come, around, understand, agree	blah, yeah, shit, everyone, stupid, rich, happen

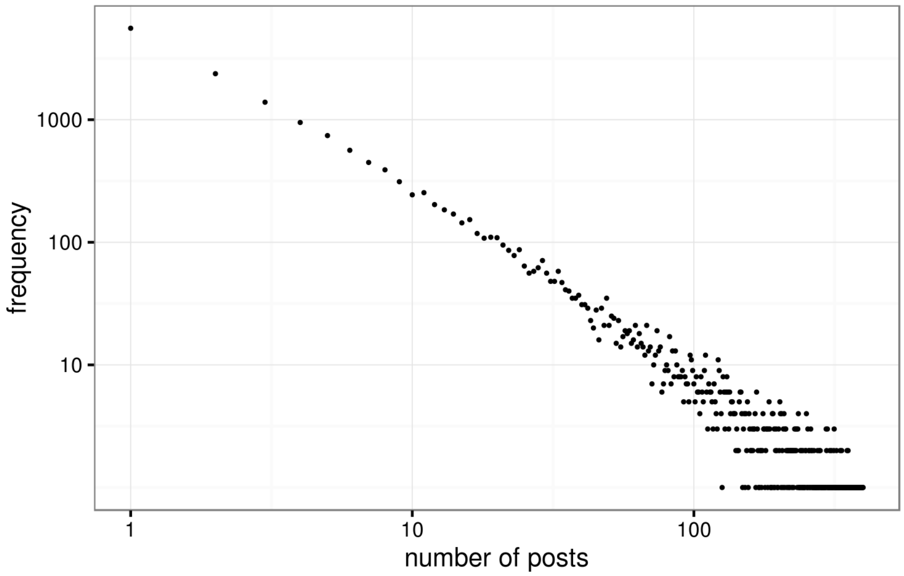

Figure 1 Distribution of posting activities among forum users (log-log scale)

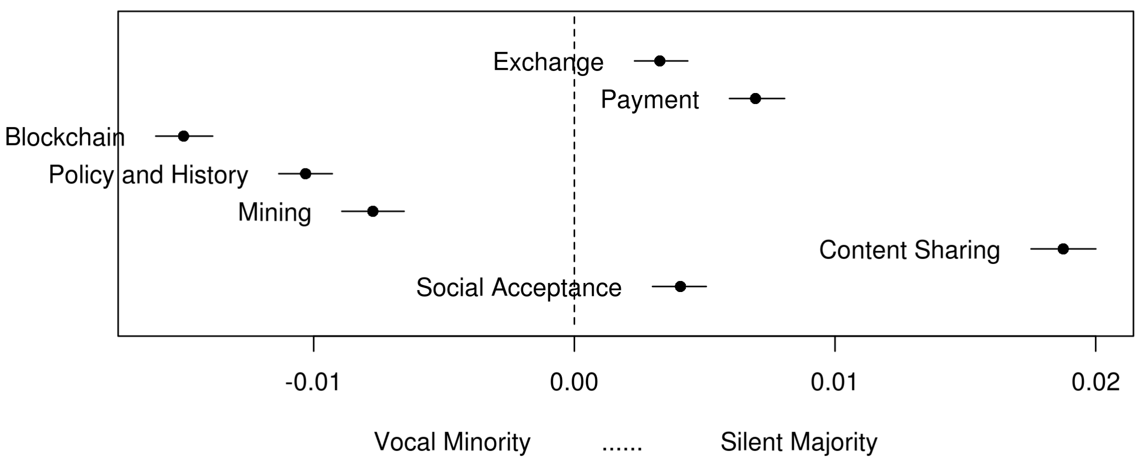

Figure 2 Differences in topics proportions (Silent Majority – Vocal Minority)