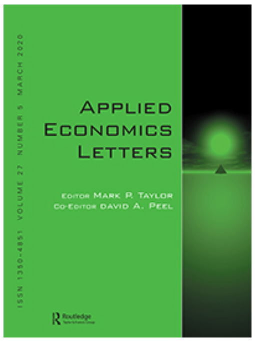

Applied Economics Letters

ISSN: 1350-4851 (Print) 1466-4291 (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

Sentiment disagreement and bitcoin price fluctuations: a psycholinguistic approach

Yongkil Ahn & Dongyeon Kim

APPLIED ECONOMICS LETTERS
2020, VOL. 27, NO. 5, 412–416
https://doi.org/10.1080/13504851.2019.1619013

ARTICLE

Sentiment disagreement and bitcoin price fluctuations: a psycholinguistic approach

Yongkil Ahn[a] and Dongyeon Kim[b]

[a]College of Management, National Taiwan Normal University, Taipei, Taiwan; [b]College of Business, Korea Advanced Institute of Science and Technology, Seoul, Korea

ABSTRACT
We investigate the extent to which Bitcoin price fluctuations are associated with investors' sentiment disagreement. We employ three textual sentiment analysis techniques: 1) a Python library offered by the Computational Linguistics and Psycholinguistics Research Center; 2) Loughran and McDonald's (2011) dictionary; and 3) semantic orientation by the point-wise mutual information method. The results show that investors' attention and sentiment disagreement induce extremely high volatility and jumps in Bitcoin prices. These findings complement existing studies on how investors' sentiment manifests in asset prices.

KEYWORDS
Bitcoin; computational linguistics; attention; disagreement

JEL CLASSIFICATION
G12; G14

I. Introduction

Innovative technologies tend to show an initial price surge and a subsequent sudden fall in the presence of extremely high volatility and jumps (e.g. Malkiel 1999). Bitcoin has also created a typical bubble-like pattern. The realised return volatility of Bitcoin prices was 107% in 2017 and 94% in 2018, approximately five times the realised volatility of the S&P 500.[1] From the literature, excessive Bitcoin price fluctuation can be attributed to irrational market participants (e.g. Shiller 2000), synchronization problems (e.g. Abreu and Brunnermeier 2003), or high uncertainty regarding future productivity (e.g. Pástor and Veronesi 2009).

Motivated by Bitcoin's market behaviour, trading volume, and price volatility in 2017 and 2018, this paper investigates the extent to which observed excessive Bitcoin price fluctuations are associated with investor sentiment. Following the burgeoning literature on online communication,[2] we employ textual sentiment analysis at high frequency using text threads on a Bitcoin bulletin board and explore the pricing implications of heterogeneous opinions among investors.[3] The fundamental difference between media articles and online posts is that while the former captures the more objective judgments, the latter may include the information content of the former in addition to the behavioural characteristics of investors.

We find that Bitcoin returns, trading volume, volatility, and signed jump variations are highly associated with investors' sentiment disagreement. As discussed by Hong and Stein (2007), disagreement may be induced by a large catalogue of variables: limited attention, heterogeneous priors, and unsmooth information flows. However, they all imply that sentiment differences among investors are linked to the extraordinary price fluctuations in the cryptocurrency market.

In evaluating the impact of investors' sentiment on asset prices, many studies focus on low-frequency implications and attribute the association to cyclical variation in the macroeconomy. By contrast, we combine a simple learning approach with opinion mining techniques to study high-

CONTACT Yongkil Ahn yongkil.ahn@gmail.com College of Management, National Taiwan Normal University, No. 31, Shida Road, Daan District, Taipei 10645, Taiwan

[1]We calculate the realised return volatility of Bitcoin using Bitstamp's closing prices. Bitstamp is ranked first among the eligible exchanges in terms of trading volume.

[2]See Das and Chen (2007), Goh, Heng, and Lin (2013), Ma, Sun, and Kekre (2015), Zhang, Bhattacharyya, and Ram (2016), and Adamopoulos, Ghose, and Todri (2018).

[3]Textual sentiment analysis can be substantially noisy (e.g. Loughran and McDonald (2016)). Thus, we follow Turney and Littman (2003) and apply a learning algorithm to online sentiment classification.

frequency sentiments on a bulletin board system. The techniques used in this paper can be easily applied to any asset, and the results complement existing studies on the anatomy of cryptocurrency markets (e.g. Yelowitz and Wilson 2015).

Peer opinions appear to play a role in the financial decision-making process. For example, Chen et al. (2014) examine the role of peer-based investment advice from Seeking Alpha[4]; Dougal et al. (2012) analyse the predictive power of Wall Street Journal columnists' opinion on stock returns; and Sprenger et al. (2014) demonstrate that Twitter messages contain valuable information on target companies. In line with the previous studies, we utilise methods from computational linguistics and offer supporting empirical evidence regarding the influence of the information content of micro-blogging bulletin boards on the cryptocurrency market.

The remainder of the paper is organised as follows. Section 2 briefly summarises how we mine the opinions on a Bitcoin bulletin board and defines the variables of interest. Section 3 discusses the results. Section 4 concludes the paper.

II. Empirical design

Textual sentiment analysis

Bitcointalk.org serves as a major source of investor sentiment data. We analyse a sub-bulletin board, entitled 'Bitcoin Discussion', where market participants voice their opinions on Bitcoin without any restrictions. As of 31 December 2018, there were approximately two million posts on the bulletin board. We crawled 1,026,804 posts published between 1 January 2017 and 31 December 2018.

To minimise potential measurement errors, we utilise three different methods to define our sentiment dictionary.[5] A polarity score that ranges from −1 to 1 can then be computed for each word in the dictionary. The Computational Linguistics and Psycholinguistics (CLiPS) Research Center offers a Python library that enables us to compute polarity scores for 2,896 predefined words, 917 positive and 1,018 negative.[6] This dictionary, Dictionary 1, is based on common situations, that is, not in the context of finance. We cannot exclude the possibility that some words might have different meanings in financial markets. Thus, we follow Loughran and McDonald (2011), a seminal article on textual analysis in the field of finance, to define 354 positive words and 2,355 negative words, Dictionary 2.[7] There might be some additional words that have specific implications in the context of cryptocurrencies. To identify such words, we employ the semantic orientation from the point-wise mutual information (SO-PMI) method proposed by Turney and Littman (2003) to define a third dictionary, Dictionary 3:

$$\mathrm{SOPIM}(w) = \sum_{pw \in PW} \ln \frac{Prob.[w \cap pw]}{Prob.[w]Prob.[pw]} - \sum_{nw \in NW} \ln \frac{Prob.[w \cap nw]}{Prob.[w]Prob.[nw]} \quad (1)$$

where pw (nw) is a set of positive (negative) words defined in Dictionaries 1 and 2. The polarity value of any word, denoted by w, can be calculated in the context of pw and nw.

Using the above three dictionaries, we first find all words that have polarity scores in each post. We compute the polarity score of each sentence in the post by summing the word-level polarity scores. If a sentence contains any negation, such as 'no' or 'not', we negate the sentence's polarity score and divide the score by two. For example, if the polarity score of the word 'bad' is −0.6, the score of 'That is bad' is also −0.6; however, the polarity score of 'That is not bad' is 0.3. Finally, the polarity score of each post is derived by adding up the polarity scores of all sentences in the post. We require threshold values that create a wedge among negative, neutral, and positive posts. Those posts that have a polarity score above 0.2 are defined as positive, and those with a polarity score below −0.1 are set to be negative. We randomly select a number of posts to train the whole sample. The chosen thresholds classify 81.6% of the whole sample correctly.

[4] https://seekingalpha.com/.
[5] Kearney and Liu (2014) and Loughran and McDonald (2016) offer a survey on textual sentiment analysis techniques in finance.
[6] Refer to https://www.clips.uantwerpen.be.
[7] For more details, we direct readers to Bill McDonald's website, www3.nd.edu/~mcdonald.

Volatility and signed jumps

In addition to the usual fluctuation metrics such as volatility, we consider signed jumps (e.g. Barndorff-Nielsen, Kinnebrock, and Shephard 2008; Patton and Sheppard 2015). Let us consider a continuous martingale for log Bitcoin price, p_t, in the following form:

$$p_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_t + J_t \tag{2}$$

where W represents a standard Wiener process and J a jump process. It can be shown that the signed jump variation is

$$\Delta J^2 \xrightarrow{p} \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s > 0\} - \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s < 0\} \tag{3}$$

The first part of Equation (3), that is, $\sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s > 0\}$, denotes the variation only due to positive returns. Similarly, the second part, that is, $-\sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s < 0\}$, represents the variation only due to negative returns.

The Chicago Mercantile Exchange CF Bitcoin Reference Rate (BRR) and the Chicago Mercantile Exchange CF Bitcoin Real-Time Index (BRTI) were launched on 14 November 2016. Several bitcoin exchanges (i.e. 'eligible exchanges'), including Bitstamp, itBit, and Kraken, provide pricing data for BRR and BRTI. We use Bitstamp's tick data for our empirical analysis.

III. Empirical results

We look for evidence of Bitcoin returns, intra-day volatility, trading volume, and signed jumps by running the following regressions:

$$\text{TargetVariable}_t = \beta_0 + \beta_1 \times \text{Investor Sentiment}_t + \sum_m \gamma_m \text{Controls}_t^m + \epsilon_t \tag{4}$$

We consider two sentiment-related variables: 1) attention and 2) disagreement. Attention, or the total number of posts per day, represents the notice that investors take of Bitcoin. Disagreement, defined as the standard deviation of all polarity scores per day, is the degree of sentiment differences. The controls include the S&P 500 index returns, the VIX index, and the spread between Moody's seasoned Baa corporate bond and the 10-year treasury constant maturity.[8] Table 1 shows the summary statistics for the key variables.

Sentiment data are not free from cyclical variation. We thus detrend the attention and disagreement variables using the Hodrick-Prescott filter. Table 2 shows the results based on the SO-PMI method, which considers both the CLiPS technique and the Loughran and McDonald (2011) approach. Panel A (B) summarises the results from data on business (non-business) days with (without) control variables.[9] The results ascertain that both attention and disagreement are significantly associated with Bitcoin intra-day volatility, trading volume, positive jumps, and negative jumps, especially on business days. The results are in accordance with Harris and

Table 1. Summary statistics. This table shows summary statistics for the key variables. The sample is from 1 January 2017 to 31 December 2018. Time is daily. Return and intra-day volatility are expressed as percentiles. Trading volume is scaled by 1000. N denotes the number of observations.

Variable	N	Mean	Std Dev	Min	25th	Median	75th	Max
Attention	730	1406.58	710.77	95	785	1270	1978	3101
Disagreement	730	0.2075	0.0097	0.1802	0.2006	0.2087	0.2148	0.2307
Return	730	0.2934	4.6961	−16.1804	−1.7317	0.3003	2.5699	26.9351
Intra-day volatility	730	0.0623	0.0280	0.0165	0421	0.0565	0.0811	0.2242
Trading Volume	730	11.8234	7.9101	0.8392	6.5836	10.1946	15.1090	70.9613
Up Jumps	730	0.0066	0.0108	0.0001	0.0013	0.0031	0.0070	0.1193
Down Jumps	730	0.0065	0.0105	0.0001	0.0015	0.0031	0.0067	0.1188

[8]The control variables are obtained from the Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.
[9]Bitcoin is traded on a 24–7 basis whereas usual financial securities are traded only on business days. We are grateful to an anonymous referee for suggesting this exercise. The control variables are not available on non-business days.

Table 2. Regression results. Panel A (B) summarises the results from data on business (non-business) days with (without) control variables. Controls include the S&P 500 index returns, the VIX index, and the spread between Moody's seasoned Baa corporate bond and the 10-year treasury constant maturity (BAA10Y). Control variables are not available on non-business days. The sample is from 1 January 2017 to 31 December 2018. Attention is scaled by 1000. We detrend the attention and disagreement variables using the Hodrick-Prescott filter to remove cyclical variations. Return and volatility are expressed as percentiles. Test statistics are between square brackets. [a], [b], and [c] represent significance at the 1%, 5%, and 10% levels, respectively.

	Dictionaries 1 + 2 + 3				
Variables	Return	Volatility	Trading volume	Up jumps	Down jumps
Panel A: regression results on business days with control variables					
Attention	0.8009	0.0179[a]	6.2085[a]	0.0183[a]	0.0116[a]
	[0.53]	[3.10]	[2.59]	[3.32]	[3.33]
Disagreement	−38.9471	0.9303[a]	155.2448[b]	0.2752[a]	0.2604[a]
	[−0.87]	[3.93]	[2.01]	[2.85]	[2.77]
S&P500 return	37.5666	−0.3282[b]	−58.6824	−0.0528	−0.0484
	[1.18]	[−2.03]	[−0.87]	[−0.72]	[−0.65]
VIX	−0.0427	−0.0019[a]	0.0282	−0.0003[a]	−0.0003[b]
	[−0.68]	[−6.30]	[0.22]	[−2.46]	[−2.25]
BAA10Y	1.5104	0.0172[a]	−7.6568[b]	−0.0216[a]	−0.0221[a]
	[0.41]	[1.81]	[−2.13]	[−4.51]	[−4.58]
N	498	498	498	498	498
F-$132#statistic	1.06	17.57[a]	2.39[b]	5.99[a]	6.08[a]
R-$132#squared	1.65%	25.44%	5.05%	14.09%	14.55%
	Dictionary 1 + 2 + 3				
Variables	Return	Volatility	Trading volume	Up jumps	Down jumps
Panel B: regression results on non-business days without control variables					
Attention	−1.0954	−0.0192[b]	−2.1370	−0.0090[a]	−0.0092[a]
	[−0.60]	[−2.14]	[−1.12]	[−3.11]	[−3.23]
Disagreement	−38.6325	1.1308[a]	6.1980	0.0224	0.0174
	[−0.94]	[3.47]	[0.09]	[0.35]	[0.28]
N	228	228	228	228	228
F-$132#statistic	0.33	18.12[a]	0.47[a]	2.80[b]	3.02[b]
R-$132#squared	0.66%	19.18%	0.50%	9.45%	9.12%

Raviv (1993) and Hong and Stein (2003), who found that differences of opinion among traders induce bubble-like price fluctuations coupled with excessive trading volumes.

IV. Concluding remarks

As Malkiel (1999, 78) documents, 'What electronics was to the 1960s, biotechnology became to the 1980s..Valuation levels of biotechnology stocks reached levels previously unknown to investors .. From the mid-1980s to the late 1980s, most biotechnology stocks lost three-quarters of their market value'. Bitcoin also appears to show extremely high volatility and jumps in its price. We apply a novel textual sentiment analysis technique to study the excessive price fluctuation in the cryptocurrency market. The results show that extremely high volatility and jumps in Bitcoin prices are associated with investors' attention and sentiment disagreement.

The study has implications for the stock market. A positive correlation between aggregate stock market returns and proxies for sentiment is reported in the literatureBaker and Wurgler (2007), for example.[10] The sentiment waves in the stock market are discernible, but in fact, much remains to be done: (a) characterizing and measuring investor sentiment and (b) understanding the time-series variation in investors' sentiment. We believe that our new psycholinguistic approach to gauge investors' sentiment in the cryptocurrency market can be easily applied to the stock market.

Further studies are called for. Compared to the common quantitative methods, textual sentiment analysis tends to be substantially noisy. Therefore, the nuances of the text need to be further understood. Additionally, asset-pricing models incorporating sentiment disagreement on innovative assets would shed new light on the predictability of revolutionary technology prices on the financial market.

[10]Potential sentiment proxies in the stock market include investor surveys, investor mood, retail investor trades, mutual fund flows, closed-end fund discount, IPO first-day returns, IPO volume, equity issues over total new issues, and insider trading.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abreu, D., and M. K. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71 (1): 173–204. doi:10.1111/1468-0262.00393.

Adamopoulos, P., A. Ghose, and V. Todri. 2018. "The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms." *Information Systems Research* 29: 612–640. doi:10.1287/isre.2017.0768.

Baker, M., and J. Wurgler. 2007. "Investor Sentiment in the Stock Market." *Journal of Economic Perspectives* 21 (2): 129–152. doi:10.1257/jep.21.2.129.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard. 2008. "Measuring Downside Risk—Realised Semivariance." *SSRN Electronic Journal*. doi:10.2139/ssrn.1262194.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media." *The Review of Financial Studies* 27 (5): 1367–1403. doi:10.1093/rfs/hhu001.

Das, S. R., and M. Y. Chen. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web." *Management Science* 53 (9): 1375–1388. doi:10.1287/mnsc.1070.0704.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons. 2012. "Journalists and the Stock Market." *The Review of Financial Studies* 25 (3): 639–679. doi:10.1093/rfs/hhr133.

Goh, K.-Y., C.-S. Heng, and Z. Lin. 2013. "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-And Marketer-Generated Content." *Information Systems Research* 24 (1): 88–107. doi:10.1287/isre.1120.0469.

Harris, M., and A. Raviv. 1993. "Differences of Opinion Make a Horse Race." *The Review of Financial Studies* 6 (3): 473–506. doi:10.1093/rfs/5.3.473.

Hong, H., and J. C. Stein. 2003. "Differences of Opinion, Short-Sales Constraints, and Market Crashes." *The Review of Financial Studies* 16 (2): 487–525. doi:10.1093/rfs/hhg006.

Hong, H., and J. C. Stein. 2007. "Disagreement and the Stock Market." *Journal of Economic Perspectives* 21 (2): 109–128. doi:10.1257/jep.21.2.109.

Kearney, C., and S. Liu. 2014. "Textual Sentiment in Finance: A Survey of Methods and Models." *International Review of Financial Analysis* 33: 171–185. doi:10.1016/j.irfa.2014.02.006.

Loughran, T., and B. McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35–65. doi:10.1111/j.1540-6261.2010.01625.x.

Loughran, T., and B. McDonald. 2016. "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research* 54 (4): 1187–1230. doi:10.1111/joar.2016.54.issue-4.

Ma, L., B. Sun, and S. Kekre. 2015. "The Squeaky Wheel Gets the Grease — An Empirical Analysis of Customer Voice and Firm Intervention on Twitter." *Marketing Science* 34 (5): 627–645. doi:10.1287/mksc.2015.0912.

Malkiel, B. G. 1999. *A Random Walk down Wall Street*. New York: WW Norton & Company.

Pástor, L., and P. Veronesi. 2009. "Technological Revolutions and Stock Prices." *American Economic Review* 99 (4): 1451–1483. doi:10.1257/aer.99.4.1451.

Patton, A. J., and K. Sheppard. 2015. "Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility." *Review of Economics and Statistics* 97 (3): 683–697. doi:10.1162/REST_a_00503.

Shiller, R. C. 2000. "Irrational Exuberance." *Philosophy and Public Policy Quarterly* 20 (1): 18–23.

Sprenger, T. O., A. Tumasjan, P. G. Sandner, and I. M. Welpe. 2014. "Tweets and Trades: The Information Content of Stock Microblogs." *European Financial Management* 20 (5): 926–957. doi:10.1111/j.1468-036X.2013.12007.x.

Turney, P. D., and M. L. Littman. 2003. "Measuring Praise and Criticism: Inference of Semantic Orientation from Association." *ACM Transactions on Information Systems (TOIS)* 21 (4): 315–346. doi:10.1145/944012.944013.

Yelowitz, A., and M. Wilson. 2015. "Characteristics of Bitcoin Users: An Analysis of Google Search Data." *Applied Economics Letters* 22 (13): 1030–1036. doi:10.1080/13504851.2014.995359.

Zhang, K., S. Bhattacharyya, and S. Ram. 2016. "Large-Scale Network Analysis for Online Social Brand Advertising." *Mis Quarterly* 40 (4). doi:10.25300/MISQ/2016/40.4.03.