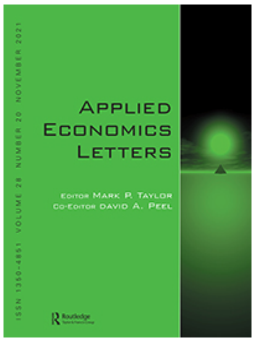


Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

# Twitter & bitcoin: are the most influential accounts really influential?

Serda Selin Öztürk & M. Emre Bilgiç

**To cite this article:** Serda Selin Öztürk & M. Emre Bilgiç (2021): Twitter & bitcoin: are the most influential accounts really influential?, Applied Economics Letters, DOI: 10.1080/13504851.2021.1904104

**To link to this article:** https://doi.org/10.1080/13504851.2021.1904104

Published online: 13 Apr 2021.

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2021.1904104

Routledge
Taylor & Francis Group

ARTICLE

Check for updates

# Twitter & bitcoin: are the most influential accounts really influential?

Serda Selin Öztürk[a] and M. Emre Bilgiç[b]

[a]Economics, Istanbul Bilgi University, Istanbul, Turkey; [b]Economics, Istanbul Technical University, Istanbul, Turkey

**ABSTRACT**

There is a vast amount of information flow in social media about bitcoin, which may affect investors' decisions. This article investigates whether tweets may affect returns or trade volume changes of bitcoin and, more importantly, whether some Twitter accounts are more influential than other Twitter accounts. We conduct two separate analyses based first on all Twitter accounts and then on the most influential 50 Twitter accounts, which have been selected as such by Unitedtraders. We use the number of positive, negative and neutral tweets by Valence Aware Dictionary and Sentiment Reasoner (VADER) in a logistic model to analyse if tweets have any valuable information about the change in both return and trade volume of bitcoin. Our results indicate that tweets can be used to predict bitcoin returns. Notably, the most influential accounts are the drivers of returns, but all Twitter accounts simply introduce some noise in volatility. This result indicates that following only these 50 most influential accounts may provide the information needed for investors.

**KEYWORDS**

Bitcoin; sentiment analysis; VADER; Twitter; logit model

## I. Introduction

Bitcoin has grabbed both investors' and researchers' attention since it was first introduced by Nakamoto in (2008). Although it has some favourable effects such as hedging, diversifying and safe haven capabilities (Dyhrberg 2016; Bouri et al., 2016); for researchers, it is more tricky to understand the forces that drive the price of bitcoin. Researchers try to capture the dynamics behind the sharp movements in bitcoin, such as in 2017, which makes bitcoin riskier for investors.

In terms of the sentiment aspect of bitcoin research, analysing tweets has been one of the popular tools. Although existing results such as Mai et al. (2015), Colianni, Rosales, and Signorotti (2015), Kinderis, Bezbradica, and Crane (2018), Lamon, Nielsen, and Redondo (2017), Stenqvist and Lönnö (2017), Matta, Lunesu, and Marchesi (2015), Abraham et al. (2018) and Shen, Urquhart, and Wang (2019) indicate sentimental effect of tweets on bitcoin price and/or volume, investors should be careful about the Twitter accounts they are following since some accounts may be misleading, and some accounts may be more influential. Therefore, in this article, we investigate whether some Twitter accounts are more influential than others regarding their effect on bitcoin returns and volume. No study in the literature differentiates the effects of Twitter accounts on bitcoin or any other cryptocurrency yet. In the first part of our analysis, we estimate our models using tweets from all Twitter accounts. Then, we only use tweets that belong to 50 accounts, chosen as the most influential Twitter accounts in terms of cryptocurrencies by Intratraders in 2018 (the most influential accounts hereafter). Furthermore, we use the number of positive, negative and neutral tweets by Valence Aware Dictionary and Sentiment Reasoner (VADER) at day t-1 separately to examine if their effects differ. Both differentiating accounts' effects and the effects of positive, negative and neutral tweets are essential in terms of investor decision-making as a way of identifying the valuable part of the total flow of information on social media.

The rest of the article is organized as follows: Section 2 presents the data and the methodology.

**CONTACT** Serda Selin Öztürk serda.ozturk@bilgi.edu.tr İstanbul Bilgi University, Eski Silahtaraga Elektrik Santrali Kazim Karabekir Cad. No:2/13, Eyup, Istanbul, Turkey

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

Section 3 discusses the results, and the last section concludes.

## II. Data and methodology

### Data

#### Tweet collection and bitcoin data

In this study, we used the method developed by J. Henrique in Python to collect older tweets since Twitter API does not allow us to download data older than 7 days. The data collection required time and machine power; therefore, we got the help of cloud servers and R Studio Server to download the data faster. 'Bitcoin' and 'btc' keywords were used to collect the tweets, using 1 day and one keyword at a time with a daily maximum tweet amount limitation of 100 thousand.[1] At the end of the process, around 25 million tweets posted between 01.06.2017 and 25.06.2018 were collected. The data is cleaned from; the punctuation marks except for #$@.',? to exclude irrelevant tweets and prevent programmatic errors during the sentiment extraction. All the URLs, Non-Latin characters, the tweets which are tagged as non-English are cleared by the 'cldr' library (developed based on Chrome's language identification library) in R. Lastly; all the capital letters were converted to lowercase letters. At the end of the process, the total data count decreased to around 21 million for all accounts and 17,021 for the most influential accounts.

In terms of bitcoin price and volume data, Bitfinex exchange market data are selected since it has above-average transaction volume. The data are publicly available in daily series at investing.com and collected for the period given above.

#### Tweet sentiment evaluation

Among many methods to evaluate the sentiment of a text, VADER was chosen. VADER is a lexicon and rule-based sentiment analysis tool developed by Hutto and Gilbert (2014) to evaluate social media messages' sentiment and was proven that it gives promising results when evaluating tweet polarization. VADER provides four scores as the output that is neutral, positive, negative and compound. The scores are normalized; therefore, positive, negative and neutral text scores add up to be 1. Only the compound score is used in this study, which is a unidimensional measure of the sentiment. The score is calculated as the aggregation of each word's valence scores in lexicon using some rules and normalized as −1 to be extremely negative and +1 to be extremely positive. The tweets with an absolute compound score that is less than 0.05 are accepted as neutral as suggested by Hutto and Gilbert (2014). After each tweet is scored, we count the number of positive, negative and neutral tweets for each day.

### Methodology

Our methodology is based upon the logit model to estimate the effect of the Twitter sentiment on the Bitcoin returns and transaction volumes. Logarithmic changes that are used as dependent variables in the regressions are provided below:

$$Z_t = \begin{cases} 1\ \textit{if return or volume change} \\ \textit{at day t is greater than}\ 0 \\ 0\ \textit{otherwise} \end{cases}$$

The logit regression model is formulated as below:

$$\begin{aligned} Z_t = \alpha &+ \beta_1 PosCount_{t-1} + \beta_2 NegCount_{t-1} \\ &+ \beta_3 NeuCount_{t-1} + \beta_4 Return_{t-1} \\ &+ \beta_5 Volume_{t-1} + u_t \end{aligned}$$

$$PosCount_{t-1} : Countofpositivetweetsindayt - 1$$

$$NegCount_{t-1} : Countofnegativetweetsindayt - 1$$

$$NeuCount_{t-1} : Countofneutraltweetsindayt - 1$$

$$Return_{t-1} : Bitcoinreturnattimet - 1$$

$$Volume_{t-1} : Bitcointradevolumechangeatdayt - 1$$

$$u_t : iiddisturbanceterm$$

[1]We limit the number of tweets to 100 thousand to shorten the data collection process. We believe that this will not affect the estimation results since there were only a couple of days with more tweets than 100 thousand tweets.

Table 1. Results for return when number of positive, negative and neutral tweets in the previous day are used.

| | All accounts | The most influential accounts |
|---|---|---|
| PosCount(−1) | 2.367** | 0.391** |
| | (1.197) | (0.181) |
| NegCount(−1) | −1.914*** | −0.403*** |
| | (0.653) | (0.009) |
| NeuCount(−1) | −0.339 | −0.314 |
| | (1.263) | (0.110) |
| Return(−1) | −3.133 | −2.507 |
| | (1.934) | (1.942) |
| Volume(−1) | 0.180 | 0.273 |
| | (0.223) | (0.226) |
| C | 0.121 | 0.111 |
| | (0.104) | (0.105) |
| McFadden R-squared | 0.029 | 0.029 |
| Log likelihood | −261.893 | −265.004 |
| LR statistic | 15.569*** | 15.514*** |
| Prob (LR statistic) | 0.008 | 0.008 |

Notes: The numbers in the brackets are standard errors. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 2. Results for volume when number of positive, negative and neutral tweets in the previous day are used.

| | All accounts | The most influential accounts |
|---|---|---|
| PosCount(−1) | −3.163** | 0.149 |
| | (1.308) | (0.185) |
| NegCount(−1) | 3.314*** | 0.083 |
| | (0.728) | (0.157) |
| NeuCount(−1) | 0.4490 | 0.308 |
| | (1.374) | (0.203) |
| Return(−1) | −3.363 | −3.351 |
| | (2.205) | (2.126) |
| Volume(−1) | 1.399*** | 1.423*** |
| | (0.263) | (0.261) |
| C | −0.210* | −0.183 |
| | (0.112) | (0.108) |
| McFadden R-squared | 0.129 | 0.082 |
| Log likelihood | −232.179 | −242.628 |
| LR statistic | 68.969*** | 43.182*** |
| Prob (LR statistic) | 0.000 | 0.000 |

Notes: The numbers in the brackets are standard errors. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

## III. Results and discussion

We represent the results for the return in Table 1. Our results are based on 419 data points in total for both return and volume data sets. The number of positive tweets has a significant and positive effect for all accounts and the most influential accounts. Based on this result, the probability of an increase in bitcoin's price gets higher when the number of positive tweets in the previous day increases. Furthermore, the number of negative tweets in the previous day has a significant and negative effect. Therefore, this implies that bitcoin return tends to decrease when the number of negative tweets increases at day t-1. The number of neutral tweets has no significant effect. Our results support Kaminski and Gloor (2014) 's results in terms of the effect of negative tweets on return, although they could not find any effect of positive tweets on return.

Moreover, when we take into account that we have much less number of tweets for the most influential accounts, both the significance of the coefficients for the most influential accounts and also no significant difference in the log-likelihood values of the models for all accounts and the most influential accounts suggest that the most influential accounts are the drivers of the change in return.

The results for volume change are represented in Table 2. Based on the results, numbers of positive and negative tweets at day t-1 have significant effects when all accounts are used, but they are not for the most influential accounts. Furthermore, they have opposite signs when we compare the results with return results. However, these results do not conflict since less volume does not necessarily mean higher return or higher volume does not necessarily imply lower return.

Moreover, our results again support Kaminski and Gloor (2014) findings, who found that negative tweets positively affect volume. Volume in the previous day is also significant for both all accounts and the most influential accounts. The significance of the number of positive and negative tweets also supports the result by Shen, Urquhart, and Wang (2019), suggesting that the number of tweets is a significant driver of next day trading volume.

Overall, our findings suggest that, on one hand, the most influential accounts are the ones that drive the returns since, despite having much less number of tweets, the change in the number of positive and negative tweets have significant effects on return. On the other hand, they do not affect the volume as it is only affected by the others' noise. In other words, the most influential accounts are influential for returns, but the others (the less influential ones) cause noise in the volatility, but they do not do much in explaining returns. Although the number of tweets is much more limited for the most influential accounts, the change in the number of positive and negative tweets significantly affects return. Therefore, these results suggest that it may be enough for investors only to follow these 50 most influential accounts regarding their bitcoin investment decisions.

## Acknowledgments

We are immensely grateful to Prof. Thanasis Stengos for his comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of this esteemed person.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Abraham, J., D. Higdon, J. Nelson, and J. Ibarra, 2018. "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis." SMU Data Science Review 1-3-1.

Bouri, E., G. Azzi, and A. H. Dyhrberg. 2016. "On the Return-volatility Relationship in the Bitcoin Market around the Price Crash of 2013." (Economics Discussion Papers, No. 2016-41). Kiel Institute for the World Economy (IfW), Kiel. https://www.econstor.eu/bitstream/10419/146870/1/869536516.pdf

Colianni, S., S. M. Rosales, and M. Signorotti, 2015. "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis." CS229 Project, 2015.

Dyhrberg, A. H. 2016. "Bitcoin, Gold and the Dollar - a Garch Volatility Analysis." *Finance Research Letters* 16: 85–92. doi:10.1016/j.frl.2015.10.008.

Hutto, C., and E. Gilbert, 2014. "Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." 8th International AAAI conference on weblogs and social media (IDWSM), Ann-Harbor, Michigan/USA.

Kaminski, J., and P. A. Gloor, 2014. "Nowcasting the Bitcoin Market with Twitter Signals." Arxiv 1406.7577.

Kinderis, M., M. Bezbradica, and M. Crane. 2018. "Bitcoin Currency Fluctuation." *Complexis*1*:* 31–41.

Lamon, C., E. Nielsen, and E. Redondo, 2017. "Cryptocurrency Price Prediction Using News and Social Media Sentiment."

Mai, F., Q. Bai, J. Shan, X. Wang, and R. H. Chiang. 2015. "From Bitcoin to Big Coin: The Impacts of Social Media on Bitcoin Performance." *SSRN Electronic Journal*, 1-46.

Matta, M., M. I. Lunesu, and M. Marchesi, 2015. "Bitcoin Spread Prediction Using Social And Web Search Media." UMAP Workshops 2015 Proceedings-Dublin/Ireland.

Nakamoto, S. 2008. A Peer-to-peer Electronic Cash System-White. paper-https://bitcoin.org/bitcoin

Shen, D., A. Urquhart, and P. Wang. 2019. "Does Twitter Predict Bitcoin?" *Economics Letters* 174 (C): 118–122. doi:10.1016/j.econlet.2018.11.007.

Stenqvist, E., and J. Lönnö. 2017. *Predicting Bitcoin Price Fluctuation with Twitter Sentiment Analysis*, 3–28. KTH Royal Institute of Technology School of Computer Science and Communication, Stockholm, Sweeden.