Economics Letters 203 (2021) 109822

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Retaliation in Bitcoin networks

Laura Lepomäki, Juho Kanniainen [*], Henri Hansen

Computing Sciences/Tampere University, Finland

ARTICLE INFO

Article history:
Received 9 February 2021
Received in revised form 15 March 2021
Accepted 16 March 2021
Available online 23 March 2021

JEL classification:
A1
G1
G2

Keywords:
Bitcoin
Peer-review networks

ABSTRACT

Due to counterparty risks, some Bitcoin trading platforms allow users to rate the level of trust they have in others. We examine users' feedback behavior on two Bitcoin trading platforms and provide statistically strong evidence that the feedback behavior of Bitcoin users is dependent on how they are rated themselves, that is, they retaliate. In addition, user's reputation is strongly and positively associated with the scores they deliver, and there is a certain persistence in the scores a user gives to others. We find that peers deliver negative feedback relatively quickly to users with bad reputation. Moreover, well-reputed users withhold negative feedback longer and give positive feedback faster than users with bad reputation.

1. Introduction

In contrast to a fiat currency, cryptocurrency creation is transparently realized via a computer algorithm, and transactions between so-called wallets are public information. However, due to users' anonymity, counterparty risk is a factor to consider. To provide information about possible counterparty risks, exchanges, where Bitcoin users rate the level of trust they have in other users (Moore and Christin, 2013), have been established. In recent years, cryptocurrencies such as Bitcoin have been a prominent subject of academic research. The existing research has focused, among others, on the inefficiency of Bitcoin (Urquhart, 2016; Nadarajah and Chu, 2017; Bariviera, 2017), Bitcoin volatility (Katsiampa, 2017; Chu et al., 2017; Lahmiri et al., 2018; Phillip et al., 2019), and price discovery and arbitrage (Brandvold et al., 2015; Brauneis and Mestel, 2018; Baur and Dimpfl, 2019; Makarov and Schoar, 2020) as well as other important topics. Moreover, the existing literature studies who the users of Bitcoin are (Bohr and Bashir, 2014; Yelowitz and Wilson, 2015), users' intentions and motivations for using Bitcoin (Glaser et al., 2014; Sas and Khairuddin, 2017), as well as models for user adoption (Athey et al., 2016). However, there seems to be very little research on peer review between bitcoin traders, a topic considered in this paper.

This article contributes the literature on cryptocurrencies from a completely new aspect by analyzing trader behavior in two peer-review platforms: Bitcoin-OTC (OTC for short) and Bitcoin-Alpha (Alpha for short). In both platforms, the users rate the trustworthiness of other users on a scale of −10 to +10 (0 excluded). A rating of −10 should be given to fraudsters while +10 means that a user trusts the person implicitly (see Kumar et al., 2016). Given the prominence of cryptocurrencies, and especially Bitcoin, in modern financial world, it is surprising how scantly the existing literature has analyzed user-to-user trust networks. Two most important papers are (Kumar et al., 2016, 2018), which consider the edge weight prediction and the detection of fraudulent users employing this data. However, to the best of our knowledge, there is no research on feedback behavior in peer-review networks in terms of how users give feedback and how they are evaluated by others.

We analyze the determinants of scores given by users to others. The hypothesis is that people can be motivated to retaliate if they think that they are poorly rated, or, conversely, deliver higher scores to users that have given them a higher rating. In addition to analyzing paired ratings between users, we ask if a user's reputation, which is earned by recent scores from other users, can affect their way of rating others. For example, one might hypothesize that users who have been rated as fraudsters would tend to rate others more harshly. We consider the impact of received reputation on the scores the users deliver to others. Moreover, we include the average of the scores that a user has given to others in our regression model, to control for potential persistence in the feedback a user gives to others.

Secondly, we analyze the time it takes until positive and negative feedback are next received or given. We determine whether these times are distributed differently between users with different reputation. Particularly, we postulate that low-reputed users deliver and receive feedback in a different way compared to well-reputed ones, which we statistically aim to verify.

[*] Corresponding author.
E-mail address: juho.kanniainen@tuni.fi (J. Kanniainen).

https://doi.org/10.1016/j.econlet.2021.109822

2. Data

Datasets used in this work are available online with links to the source of the data, i.e. the web pages of the actual trading platforms (BitcoinAlpha, 2019; BitcoinOTC, 2019). Although the link to Bitcoin Alpha trading platform is no longer valid, the datasets are considered accurate as they are used in previous publications (Kumar et al., 2016, 2018). Both datasets contain traders numbered by positive integers and integer trust ratings ranging from −10 up to 10, excluding 0. Rating value 10 represents the highest possible trust, while −10 means complete distrust. The time the ratings are given is also given. A new score from trader i to user j overwrites the previous score i has given to j, and the datasets contain the most recent scores. Therefore, trader i's score on user j appears either once or never in the data. Moreover, a trader cannot give a trust rating to himself/herself. Table 1 shows the dimensions of the datasets. As shown in the table, both the datasets include thousands of users. In both of the datasets, there are more users who have received scores than users who have rated others. Therefore, the average number of given scores among raters is bigger than the average number of received scores among receivers.

3. Methods and results

3.1. Explaining scores that users give to their peers

In this section, we analyze how the scores the user has received and given are associated to the scores the user gives. Particularly, we explain a score the user gives to their counterparty by

- the most recent score that the counterparty gave to the user,
- the average of recently received scores from other users, and
- the average of recently given scores to other users.

Intuitively, if a user has received good (bad) score from another user, they can respond by giving good (bad) score in return, a hypothesis we aim to test. A user can have good (bad) reputation because of high (low) recent scores given by other users, which will be a controlled variable. Also, a user can have a general persistence to give high (low) scores, which we also control (i.e. autocorrelation).

We formulate a regression model as follows:

$$S_{i,j}(t) = a + b_1 S_{j,i}(t-) + b_2 \bar{S}_{*,i}(t-) + b_3, \bar{S}_{i,*}(t-) + \varepsilon(t)$$

where $S_{i,j}(t)$ is a score given by user i to user j at time t, and $i, j = 1, 2, \ldots, N$; $S_{j,i}(t-)$ is the most recent score given by user j to user i; $\bar{S}_{*,i}(t-)$ is the average of scores received by user i from other users calculated from k earlier data points; and $\bar{S}_{i,*}(t-)$ is the average of scores delivered by user i to other users calculated from k most recent data points. In the main analysis, we use $k = 5$, that is, the explanatory variable is calculated from the five most recent observations if they are available. To address the multicollinearity of explanatory variables, we run the regressions for each explanatory variable separately and together. We include only those data rows for which there are observations for the dependent variable and all the explanatory variables. For that reason, the number of observation varies through the regression models with different explanatory variables.

Table 2 reports the results from the regression based on OTC data (Panel A) and Alpha data (Panel B). We find that all the four models yield consistent results in terms of the sign and significance of the coefficient of the explanatory variables. Particularly, the score the users has received from a counterparty is positively related to the score the user gives to the same counterparty, i.e., we do observe a tendency to retaliate. Moreover, the average values of unpaired, recently received and given scores are positively associated to the scores users give to their peers. The first indicates that a user's general reputation is positively associated with the scores the user gives, and the second that there is a certain persistence in the scores given by a user. Regarding the choice of k in the determination of the averages of received and given scores, the results are very robust and remain approximately the same if we vary k between 1 and 10 (results available upon question). When comparing the importance of the explanatory variables, in terms of R2, the first variable (scores received from the same counterparty) has a relatively dominating impact compared to the others (the unpaired average values of the recently received and given scores).

3.2. Relation of reputation to time to receive or give feedback

In the above, we analyzed the level of scores that the users give to each other. Here we continue our research by examining the relation of users' reputation to time it takes to receive and give positive versus negative feedback. Particularly, our hypotheses are the following:

H1: Users with low reputation receive bad scores fast compared with well-reputed users.
H2: Users with low reputation receive good scores slow compared with well-reputed users.
H3: Users with low reputation give bad scores fast compared with well-reputed users.
H4: Users with low reputation give good scores slow compared with well-reputed users.

The intuition behind the hypotheses is in line with the results of Section 3.1 in the above. Regarding H1, if a user already has bad reputation, then a counterparty has a lower threshold to give negative feedback immediately. That is, the counterparty can think that as others' have already rated the user as being disreputable, they should not hesitate to deliver their negative scores right away. Correspondingly, when it comes to receiving positive feedback (H2), one could expect that well-reputed users are rewarded without delay. The same logic applies to how a user, whose reputation we measure, delivers feedback to others (H3 and H4). Particularly, we expect that a user with bad reputation tends to retaliate' by showing his negative feedback to others without delay (H3). Correspondingly, well-reputed users do not dawdle to thank others (H4).

To test these hypotheses, we apply the following procedure:

(i) Calculate user i's reputation at a point she receives a score from another user. Similarly to the analysis in Section 3.1, reputation is calculated from the $k = 5$ most recent scores (including the newest score). More formally, given that user i's jth received score is denoted by $x(i, j)$, $j = 1, 2, \ldots, M_i$, her reputation at this point is $r(i, j) = \frac{1}{k}\sum_{l=0}^{k-1} x(i, j-l)$.

(ii) Measure time distance from a point at which the user receives a score to a point at which the user receives a score next time, denoted by $\Delta t_r(i, j)$ for user i's jth score. Correspondingly, measure time distance to the point at which the user gives a score next time, denoted by $\Delta t_g(i, j)$ for user i's jth score.[1]

(iii) Observe the levels of next received and given scores. The data show that feedback given by the users is clustered in time. For that reason, we decide to measure the level

[1] The datasets contain the most recent scores. This, however, should not bias the tests because we are not primarily interested in the absolute values of the waiting times, but if waiting times are *differently distributed* between users with high and low feature values.

Table 1
The first column denotes the dataset in question and the next three columns show the number of users who have rated others, the number of users who have received scores, and the number of scores, respectively. 'Time Range' column shows the first and the last time-stamp in the data. The average number of the received scores among receivers, μ_{in}, and given scores among raters, μ_{out}, as well as their standard deviation ($\sigma_{\text{in}}, \sigma_{\text{out}}$ resp.) rounded to two decimals are shown in the last columns. There are users who have only received ratings and not given any, as well as users who have rated others without receiving any scores. Therefore, the number of receivers and raters differ, and μ_{in} and μ_{out} are not the same.

Dataset	Raters	Receivers	Scores	Time Range	μ_{in}	σ_{in}	μ_{out}	σ_{out}
Bitcoin OTC	4814	5858	35592	2010-11-08–2016-01-25	6.08	17.71	7.39	23.1
Bitcoin Alpha	3286	3754	24186	2010-11-08–2016-01-22	6.44	16.46	7.36	19.45

Table 2
Results of the linear regression to explain the scores the users give to their peers. We report t-statistics in parentheses and statistical significance is indicated as follows: *with p < 0.1, **with p < 0.01, ***with p < 0.001.

Parameter	Model 1	Model 2	Model 3	Model 4
	Panel A: OTC			
Intercept	0.30	−0.13	0.17	−0.49
	(13.13)***	−(6.03)***	(8.05)***	−(16.62)***
Received score from the same counterparty	0.79			0.66
	(95.83)***			(71.94)***
Average of received scores		0.80		0.47
		(106.90)***		(42.78)***
Average of given scores			0.72	0.14
			(98.16)***	(13.34)***
N	14100	25206	25206	10005
R2	0.39	0.31	0.28	0.52
	Panel B: Alpha			
Intercept	0.36	0.32	1.04	−0.96
	(4.50)***	(11.24)***	(34.56)***	−(9.45)***
Received score from the same counterparty	0.69			0.60
	(30.13)***			(27.13)***
Average of received scores		0.67		0.76
		(59.47)***		(26.99)***
Average of given scores			0.27	0.13
			(20.33)***	(3.96)***
N	2739	15974	15974	2182
R2	0.25	0.18	0.03	0.43

of the next scores by calculating the average over all the received/given scores that take place in one hour window starting from the next received/given score. For user i's jth score, the average values of the next received and given score values (that appear within one hour) are denoted by $s_r(i,j)$ and $s_g(i,j)$, respectively.

(iv) Do (i–iii) for each user at each point she receives a score.

(v) To test hypotheses H1 and H2, for a given i and j, exclude all the observations ($r(i,j)$, $\Delta t_r(i,j)$, $s_r(i,j)$) if one of them cannot be determined.

(vi) Split the data rows into four categories based on the median values for reputation $\{r(i,j); i = 1 \dots N, j = 1 \dots M_i\}$ and the value of the next scores received $\{s_r(i,j); i = 1 \dots N, j = 1 \dots M_i\}$. In that way, we have data for four user categories: low reputation, low scores to be received ($\mathcal{LL}$); high reputation, low scores ($\mathcal{HL}$); low reputation, high scores ($\mathcal{LH}$); high reputation, high scores ($\mathcal{HH}$).

(vii) Perform one-side two-sample t-test on time-distances, Δt_r, between categories $\mathcal{LL}$ and $\mathcal{HL}$ against alternative hypothesis H1. Similarly, perform t-test between $\mathcal{LH}$ and $\mathcal{HH}$ against alternative hypothesis H2.

(viii) To test hypotheses H3 and H4, use $\Delta t_g(i,j)$ and $s_g(i,j)$ instead of $\Delta t_r(i,j)$ and $s_r(i,j)$ in steps (vi–vii).

Above, we used median values to assign data rows into different categories. Alternatively, one could split the dataset around zero-values. However, not scores nor reputation (calculated from scores) are symmetrically distributed around zero. Fig. 1 shows the histograms of users' reputation for OTC and Alpha platforms. Because of this asymmetry, median values are used instead of classifying users with respect to negative vs positive reputation.

Table 3 reports the results. Firstly, we find strong empirical evidence for H1. That is, users with bad reputation get negative feedback significantly faster compared with well-reputed users. This means that negative feedback is delivered promptly to people that have bad reputation, while peers do not hurry to deliver negative feedback to well-reputed people. The results are statistically highly significant and robust with respect to platform (OTC vs Alpha) and on how data is split (medians vs quartiles). Secondly, H2 lacks of empirical evidence, indicating that users' reputation is not statistically related to time it takes to receive good scores. This introduces an asymmetric relation: reputation matters only if a user receives negative feedback from the peers.

Thirdly, we provide evidence for H3, confirming that users with low reputation are quick to give negative feedback to others. This result is consistent otherwise, but not with the data split with quartiles for Alpha platform. In this regard, we observe that users' own behavior is similar to the way they are treated by their peers (that is, H1 and H3 are consistent). Fourthly, regarding H4, we confirm that users with high reputation are relatively eager to provide positive feedback to their peers. Also this result is quite robust with an exception of data from Alpha platform with the use of median for the determination of the cut-off point. Interestingly, while reputation is not related on receiving good scores (H2), it affects the way how users delivers good scores (H4).

To summarize, the results indicate that the schedule of negative feedback is affected by the reputation of both the target and source users. At the same time, there is an asymmetry in that bad reputed users receive *negative* feedback without delay, while reputation is not related to the way users receive

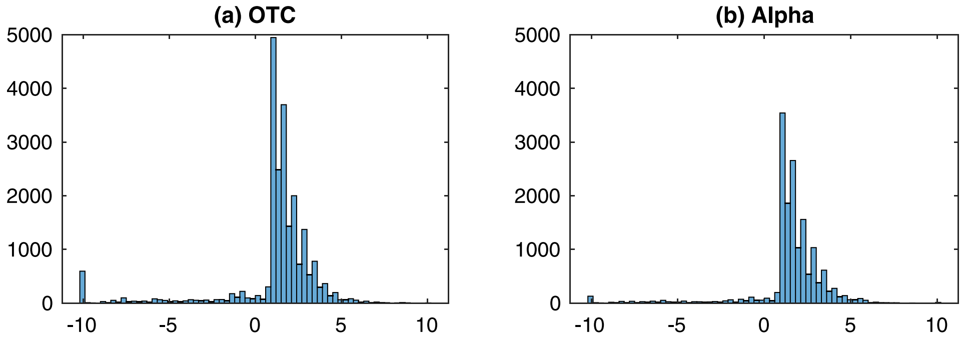

Fig. 1. Histogram of user's reputation in (a) OTC platform and (b) Alpha platform.

Table 3
Results on testing hypotheses H1–H4. Table reports the time-distances from a time-stamp where a user receives feedback to the time-stamp of the next event where the user either receives or gives scores. The results are reported in days. In Panel A, to test each hypothesis, the data rows are divided into four categories using median values on reputation and next scores. Then one-sided t-tests are run for the time distances to the next score that a user receives or gives. Panel B corresponds to Panel A except that the data rows are divided into four categories using the first and last quartiles rather than medians. Statistical significance is indicated as follows: *with $p < 0.1$, **with $p < 0.01$, ***with $p < 0.001$.

	Panel A: Median Cut-Off		Panel B: Quartile Cut-Off	
	OTC	Alpha	OTC	Alpha
H1: Receive bad scores				
Mean waiting time, low reputation	24.66	33.00	23.74	32.17
N, low reputation	7651	4512	5044	2769
Mean waiting time, high reputation	29.21	40.23	30.78	45.75
N, high reputation	4768	3267	2139	1445
p-val	(1.87E−03)***	(9.90E−04)***	(1.01E−03)***	(8.55E−05)***
H2: Receive good scores				
Mean waiting time, low reputation	18.14	21.27	18.29	24.14
N, low reputation	3559	2890	928	731
Mean waiting time, high reputation	18.45	21.39	18.79	21.89
N, high reputation	4623	3769	1540	1244
p-val	(0.604)	(0.53)	(0.602)	(0.18)
H3: Give bad scores				
Mean waiting time, low reputation	11.72	22.41	11.16	22.84
N, low reputation	7305	4232	4659	2443
Mean waiting time, high reputation	14.46	25.48	12.79	23.13
N, high reputation	4653	3289	2057	1454
p-val	(2.274E−03)***	(0.029)**	(0.097)*	(0.447)
H4: Give good scores				
Mean waiting time, low reputation	19.34	24.72	39.64	32.89
N, low reputation	3177	2590	702	610
Mean waiting time, high reputation	16.08	23.00	23.93	27.66
N, high reputation	4360	3402	1425	1172
p-val	(0.011)**	(0.16)	(4.55E−05)***	(0.08)*

positive feedback. The question of the cause for such a behavior, which is statistically significant and surprisingly strong and robust, remains.[2]

4. Conclusions

The above analysis shows that the way Bitcoin users rate each other are dependent on how they are rated themselves. Most importantly, we provide statistically strong evidence that users tend to give a high (low) score to a counterparty if they have received a high (low) score recently from the same counterparty. That is, in giving feedback through the rating system, a user tends to follow the receiving person's rating. Moreover, a user's reputation is strongly and positively associated with the scores the user delivers and, on the other hand, there is a certain persistence in the scores a user gives to others.

We also analyzed how users' reputation is associated to the time that it takes to give and get positive versus negative feedback. In this regard, the most important findings are that peers with bad reputation are delivered negative feedback relatively quickly by other users (H1) whereas reputation is not related to the way how users receive positive feedback (H2). Moreover, well-reputed users withhold negative feedback longer than users with bad reputation (H3) and consistently to this, users with bad reputation are not in hurry to deliver good scores to the peers (H4). Even though this paper considers data from Bitcoin platforms, the results may well generalize to other peer rating networks, such as the trust between Wikipedia editors based on the re-edits the editors do after each other (see Maniu et al., 2011), which we aim to investigate in our future research.

[2] We also observe that the number of the data rows vary across the four categories ($\mathcal{LL}$, $\mathcal{HL}$, $\mathcal{LH}$, $\mathcal{HH}$). This is because if a data point is at the median value, it is assigned to the 'low' category rather than excluded.

References

Athey, S., Parashkevov, I., Sarukkai, V., Xia, J., 2016. Bitcoin pricing, adoption, and usage: Theory and evidence.

Bariviera, A.F., 2017. The inefficiency of bitcoin revisited: A dynamic approach. Econom. Lett. 161, 1–4.

Baur, D.G., Dimpfl, T., 2019. Price discovery in bitcoin spot or futures? J. Futures Mark. 39 (7), 803–817.

BitcoinAlpha, 2019. Bitcoin alpha trust weighted signed network. https://snap.stanford.edu/data/soc-sign-bitcoin-alpha.html. (Accessed: 2019-08-09).

BitcoinOTC, 2019. Bitcoin OTC trust weighted signed network. https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html. (Accessed: 2019-08-09).

Bohr, J., Bashir, M., 2014. Who uses bitcoin? an exploration of the bitcoin community. In: 2014 Twelfth Annual International Conference on Privacy, Security and Trust. IEEE, pp. 94–101.

Brandvold, M., Molnár, P., Vagstad, K., Valstad, O.C.A., 2015. Price discovery on bitcoin exchanges. J. Int. Financ. Mark. Inst. Money 36, 18–35.

Brauneis, A., Mestel, R., 2018. Price discovery of cryptocurrencies: Bitcoin and beyond. Econom. Lett. 165, 58–61.

Chu, J., Chan, S., Nadarajah, S., Osterrieder, J., 2017. GARCH modelling of cryptocurrencies. J. Risk Financ. Manag. 10 (4), 17.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C., Siering, M., 2014. Bitcoin-asset or currency? revealing users' hidden intentions. Revealing Users' Hidden Intentions (April 15, 2014). ECIS.

Katsiampa, P., 2017. Volatility estimation for Bitcoin: A comparison of GARCH models. Econom. Lett. 158, 3–6.

Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., Subrahmanian, V., 2018. Rev2: Fraudulent user prediction in rating platforms. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 333–341.

Kumar, S., Spezzano, F., Subrahmanian, V., Faloutsos, C., 2016. Edge weight prediction in weighted signed networks. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE, pp. 221–230.

Lahmiri, S., Bekiros, S., Salvi, A., 2018. Long-range memory, distributional variation and randomness of bitcoin volatility. Chaos Solitons Fractals 107, 43–48.

Makarov, I., Schoar, A., 2020. Trading and arbitrage in cryptocurrency markets. J. Financ. Econom. 135 (2), 293–319.

Maniu, S., Cautis, B., Abdessalem, T., 2011. Building a signed network from interactions in wikipedia. In: Databases and Social Networks. pp. 19–24.

Moore, T., Christin, N., 2013. Beware the middleman: Empirical analysis of Bitcoin-exchange risk. In: International Conference on Financial Cryptography and Data Security. Springer, pp. 25–33.

Nadarajah, S., Chu, J., 2017. On the inefficiency of Bitcoin. Econom. Lett. 150, 6–9.

Phillip, A., Chan, J., Peiris, S., 2019. On long memory effects in the volatility measure of cryptocurrencies. Finance Res. Lett. 28, 95–100.

Sas, C., Khairuddin, I.E., 2017. Design for trust: An exploration of the challenges and opportunities of bitcoin users. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 6499–6510.

Urquhart, A., 2016. The inefficiency of Bitcoin. Econom. Lett. 148, 80–82.

Yelowitz, A., Wilson, M., 2015. Characteristics of Bitcoin users: an analysis of Google search data. Appl. Econ. Lett. 22 (13), 1030–1036.