Short and Distort Manipulations in the Cryptocurrency Market: Case Study, Patterns and Detection

Xun Sun[1], Xi Xiao[1], Wentao Xiao[1], Bin Zhang[2], Guangwu Hu[3](✉), and Tian Wang[4]

[1] Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
{x-sun21,xwt20}@mails.tsinghua.edu.cn, xiaox@sz.tsinghua.edu.cn
[2] Department of New Networks, Peng Cheng National Laboratory, Shenzhen, China
bin.zhang@pcl.ac.cn
[3] School of Computer Science, Shenzhen Institute of Information Technology, Shenzhen, China
hugw@sziit.edu.cn
[4] BNU-UIC Institute of Artificial Intelligence and Future Networks, Beijing Normal University (BNU Zhuhai), Zhuhai, China
tianwang@bnu.edu.cn

Abstract. Recently, with the development of blockchain, cryptocurrencies such as bitcoin are gaining popularity. However, at the same time, there are some problems such as lack of laws and supervision, leading to an endless stream of frauds and market manipulations, which seriously impacts the interests of investors and the development of the economic society. This paper studies Short and Distort, a new manipulation as opposed to Pump and Dump in the cryptocurrency market for the first time. By using the method of case study, Short and Distort is explained. Further, we mine three important patterns with ordinary least squares regression. That is, cryptocurrencies always show price reversal, abnormal trading volume, and widening of bid-ask spreads. Moreover, a model is constructed to detect Short and Distort. Thorough experiments show that our model is superior to the other four methods.

Keywords: Blockchain · Cryptocurrency · Fraud scheme · Short and Distort · Detection model

1 Introduction

Blockchain is a technical solution for collectively maintaining a reliable database in a decentralized manner. Since Satoshi Nakamoto put forward the concept of blockchain in 2008 [1], blockchain has attracted wide attention in the world. Bitcoin, as a cryptocurrency, is the first generation of applications of blockchain. The price of bitcoin has continued to climb, reaching a market capitalization of $1.1 trillion in March, 2021 [2]. However, while blockchain and cryptocurrency are booming, problems such as lack of regulations are gradually exposed in the field, which seriously infringe on the interests of investors [3, 4].

Y. Lai et al. (Eds.): ICA3PP 2021, LNCS 13157, pp. 494–508, 2022.
https://doi.org/10.1007/978-3-030-95391-1_31

Many studies [5, 6] have shown that cryptocurrencies are manipulated by a variety of complex schemes. Pump and Dump (P&D) is a well-known fraud scheme that manipulators spread false or misleading positive information to drive up the price of a stock in order to sell the stock at a higher price. Once the manipulators sell their overvalued shares, the price falls and the innocent investors suffer a loss. Then, is there such a fraud scheme opposite to P&D, which means the manipulators operate to depress the price, rather than driving up the price?

Few researches have studied this kind of opposite fraud scheme in the stock market [7, 8], naming it as Short and Distort (S&D). In the S&D scheme, manipulators taking a short position spread negative rumors to drive down the stock price. When the price is depressed, manipulators buy back to gain profits. The whole process with the P&D process happens to be dual. The details of comparisons between them are shown in Table 1.

Table 1. Comparisons between Pump and Dump (P&D) and Short and Distort (S&D)

Items	Pump and Dump (P&D)	Short and Distort (S&D)
Purpose of manipulation	Drive up prices	Depress prices
Type of spreading information	Positive and misleading	Negative and misleading
Way to profit	Buy low first, then sell high	Sell high first, then buy low
Target of manipulation	Small market capitalization	Medium or large market capitalization
Price trend	Shaped like "Λ"	Shaped like "V"
Regulation in stock market	Strict	Difficult
Familiarity	Well-known	Less well-known

In the stock market, P&D is widely known to people, and it is relatively easier to detect and regulate it than S&D in the stock market. In the past, there have been many cases where manipulators have been punished by the U.S. Securities and Exchange Commission (SEC) because of P&D manipulations. However, S&D also occurs frequently in stock market, but it is difficult to identify and regulate S&D schemes because manipulators operate it so covertly. Thus, cases of SEC Penalties for S&D have been far less than for P&D in the past 20 years [7]. People are also strange to this scheme with few scholars' researches.

What about the emerging cryptocurrency market? Cryptocurrencies are easy to be manipulated due to lack of regulations. With the popularity of cryptocurrencies, plenty of people come into this market, and many scholars pay attention to this area and focus on the fraud schemes, such as P&D in the cryptocurrency market. However, to the best of our knowledge, there are no scholars study S&D schemes in the cryptocurrency market yet. We first study it, trying to warn investors to stay out of the market when the manipulation comes in and provide a method for the supervision of regulators. The main contributions are as follows:

- We study Short and Distort in the cryptocurrency market for the first time. A case study is used to explain its whole process.
- Three important patterns are minded for people to understand the characteristics of Short and Distort.
- We first construct a model to detect Short and Distort in the market. Exhaustive experiments show its superiority. Further, a new dataset of Short and Distort is collected and published.[1]

2 Related Works

In this section, we review the researches on Pump and Dump schemes in the stock and cryptocurrency markets, Short and Distort schemes in the stock market, and some other scams in the cryptocurrency market.

2.1 Pump and Dump Schemes in the Stock and Cryptocurrency Market

Many scholars have studied P&D schemes in the stock market in the early days. In 2006, Aggarwal et al. [9] showed that stocks with low trading volumes were more likely to be manipulated. Massoud et al. [10] found that OTC companies usually carried out stock promotion. Huang et al. [11] showed the manipulations had important effects on market efficiency.

With the rapid growth of cryptocurrencies, P&D schemes in the traditional stock market are also used in the emerging field. Kamps et al. [12] first studied P&D in the cryptocurrency market and proposed a standard to define these activities. Hamrick et al. [13] and Li et al. [14] analyzed P&D events and the new characteristics of P&D. Dhawan et al. [15] studied the reasons for people's participation. Recently, scholars begin to propose some methods to detect P&D. Victor et al. [16], Morgia et al. [17], Mirtaheri et al. [18], Mansourifar et al. [19], and Nilsen et al. [20] quantified and detected P&D scams in the real market. Chen et al. [21] proposed an Apriori algorithm to detect user groups involving in P&D. Xu et al. [22] tried to predict the target tokens of P&D.

2.2 Short and Distort Schemes in the Stock Market

In the traditional stock market, some researchers have studied S&D schemes. Weiner et al. [7] argued that the supervision of S&D should be on the same level as P&D. In 2019, they [8] also found that it was a challenge for companies to pursue claims in S&D. Mitts [23] showed that S&D scams cause a large amount of mispricing.

Activist short selling, in which short-sellers publicly talk down stocks to benefit their short positions, is similar to S&D in concept. Thus, works about activist short selling are also helpful for our research. Zhao [24] argued that companies that are grossly overvalued or uncertain are often the targets. Ogoodniks et al. [25] found activists with a good reputation were more likely to make profits in the short term. Kartapanisde et al. [26] found that activist short selling had a temporary negative impact on the capital market.

[1] https://github.com/DataCodeHub/short-and-distort.

2.3 Other Fraud Schemes in the Cryptocurrency Market

There are also many other fraud schemes in the cryptocurrency market due to lack of regulation. Gandal et al. [5] identified the suspicious trading activities of two “bots”, Markus and Willy, on the Mt. Gox exchange. Bartoletti et al. [6] applied data mining techniques to detect Bitcoin addresses related to Ponzi schemes. Nizzoli et al. [27] studied online cryptocurrency manipulations on social media.

3 A Case Study of the Short and Distort Scheme

On the night of March 7, 2018, social media users posted that the Binance exchange had been hacked, and that users’ cryptocurrencies in the exchange account had been sold by the hackers via Bitcoin (BTC) or Viacoin (VIA, a small market capitalization cryptocurrency). The hacker’s behavior directly led to a sharp rise in the price of VIA in a short time.

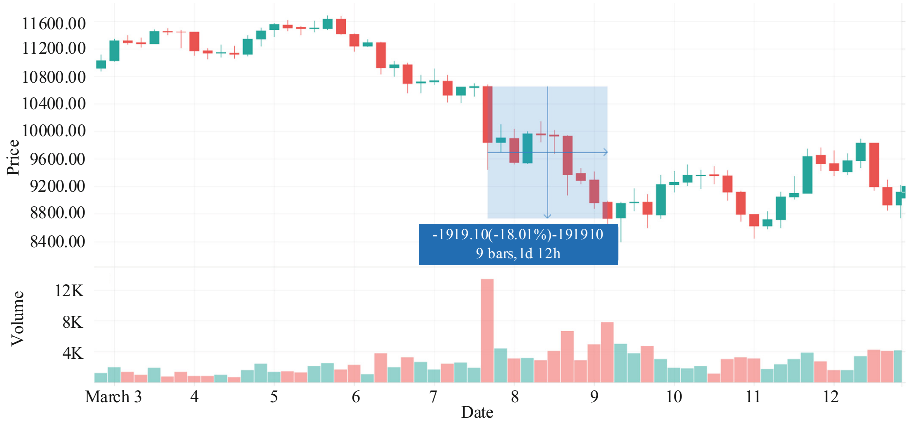

Fig. 1. The candlestick chart of the price and the bar chart of trading volumes around March 8, 2018 [28]. Each candlestick or bar represents 4 h. The top part of figure shows that the price dropped significantly after hacking (at about 20:00 on March 7). As indicated in the blue box in the figure, the price decreased by 18.01% only within 1 day and 12 h (9 bars). The bottom of figure shows that the volume increased sharply at about 20:00 on March 7, peaking over 12K. These two patterns are the signs of S&D scheme as we describe in Sect. 4.

The hacker’s action caused panic in the market and uninformed traders joined in the selling activity, which led to prices of cryptocurrencies on all major exchanges fell sharply. Figure 1 shows the price and trading volume of BTC/USD on the Bitstamp exchange around March 8, 2018. It indicates that in just one and a half days after the hacking, the price of bitcoin dropped as much as 18.01%, while the trading volumes increased significantly.

However, hackers didn’t end here. They had expected the exchange to stop losses by banning withdrawals. Thus, they just caused bearish sentiment to the market and profited from the orders which they held in advance.

The hacker's behavior in this case shows the characteristics of S&D schemes. First, hackers took short positions in cryptocurrencies on various exchanges in advance. Later, they brought panic to the market by attacking the Binance exchange. In the end, a sharp drop in prices and a surge in trading volumes occurred due to the selling of the uninformed users, and then the hackers were profiting from their short position.

4 Three Patterns of Short and Distorts Schemes

In this section, we conduct a quantitative study through event study methodology and Ordinary Least Squares (OLS) regression analysis to learn some patterns of S&D schemes. First, data collection is introduced.

4.1 Data Collection

According to works of Bouoiyour et al. [29], we collected the data of the bitcoin market from August 21, 2017 to July 30, 2021, including price, volume, bid-ask spreads, bulls, bears, hash, average time between transactions, google trend, gold price, crypto fear & greed index and cryptocurrency index. Data sources are shown in Table 2.

Table 2. Data source.

Variable	Meaning	Time granularity	Source
P	Prices	2 h	https://app.intotheblock.com/
V	Trading volume	2 h	http://data.bitcoinity.org/markets/volume/
S	Bid-ask spreads	1 day	http://data.bitcoinity.org/markets/spread
Bu	Bulls	2 h	https://www.tradingview.com/chart/
Be	Bears	2 h	https://www.tradingview.com/chart/
H	Hash	1 day	https://app.intotheblock.com/coin/BTC
ABT	Average time between transactions	1 day	https://app.intotheblock.com/coin/BTC
GT	Google trend	1 day	https://trends.google.com/trends/explore
GP	Gold price	1 day	https://fred.stlouisfed.org/
CFGI	Crypto fear & greed Index	1 day	https://alternative.me/crypto/fear-and-greed-index/
CRIX	cryptocurrency Index	1 day	http://data.thecrix.de/

Bid-ask spreads are the gap between bid and offer prices. Bulls and Bears represent the prices of the orders of BTCUSDLONGS and BTCUSDSHORTS, which indicates the power of the bulls and the bears, respectively. Hash, which reflects the computational difficulty of bitcoin miners, is used to represent the technical drivers that affect bitcoin's returns. Average time between transactions refers to the time between two bitcoin transactions, which can reflect the activity of the bitcoin market. Google trend denotes the frequency of searching the "bitcoin" keyword in google searches around the world, which can indicate people's interest in bitcoin over a period. Gold price is the dollar price of one ounce of gold. Some scholars believe that bitcoin can replace gold as a hedging tool [30]. Thus, to a certain extent, the returns of bitcoin are related to the gold price. Crypto fear & greed index is used to reflect the panic or greed level of the whole cryptocurrency market. Cryptocurrency Index is used to reflect the returns of the cryptocurrency market portfolio [31]. The bitcoin returns that are different from the market portfolio are considered as abnormal returns, which is necessary for event study methodology in the following section.

Furthermore, we collected 472 "short ideas" (i.e., posts of S&D schemes) about bitcoin on the TradingView website[2]. TradingView is a well-known social network platform for trading and investment in cryptocurrencies. A wide range of investors pay close attention to the opinion posts on the website. Therefore, it is one of the "best ways" for S&D manipulators to spread negative information on TradingView.

In the following, we examine whether any patterns occurred in the bitcoin market around the publication time of these "short ideas".

4.2 Pattern I: Price Reversal

Some scholars have studied that when the stock market experiences S&D manipulation, the stock prices first decline and then reverse [25], as shown in Fig. 2.

Fig. 2. Diagram of price reversal.

The horizontal axis of Fig. 2 represents the time. "0" refers to the time when S&D manipulation starts or the bad news begin to be published, "1" represents the time of one time step after the publication, "−1" refers to the time of one time step before the publication, and so on. The vertical axis denotes returns.

S&D schemes often exhibit the pattern shown in Fig. 2 [23, 32]. Prices begin to fall before the publication, which may be due to the manipulator making relevant trades in advance or the recent downward tendency of the market (such as "bear market"). In this market, the S&D manipulators observe for a few days and decide to carry out a manipulation. At time step 0, manipulators begin to spread false bad news on the

2 https://tradingview.com/.

Internet, causing a panic. Then, the uninformed investors sell their assets with prices falling further, and the manipulators buy the assets at lower prices (commonly known as "picking the bottom"). From time step 2 to time step 5, people begin to realize that the news is false, and someone is manipulating the market. Thus, the contrast of the powers of the bulls and the bears gradually reverses the prices towards the true values.

After defining Cumulative Abnormal Returns (*CAR*) in event study methodology [33], we conduct the following OLS regression to test price reversal pattern in the bitcoin market during S&D period.

$$\begin{aligned} CAR_{10,50} = \beta_0 + \beta_1 CAR_{-20,9} + \beta_2 CAR_{-20,-19} + \beta_3 CBB_{-20,9} \\ + \beta_4 CBB_{-20,-19} + \varepsilon \end{aligned} \tag{1}$$

where $CAR_{10,50}$ represents Cumulative Abnormal Returns from 10 × 2 h to 50 × 2 h after the publication of bad news (Note that the time interval is 2 h). $CBB_{-20,9}$ refers to the Cumulative changes of the force of the Bulls and the Bears from 20 × 2 h before and 9 × 2 h after the publication (*C* of *CBB* is short for Cumulative, and *BB* is short for Bulls and Bears). Others follow the same definition. ε is the error term.

The most important coefficient is β_1, which reflects the influence of $CAR_{-20,9}$ on $CAR_{10,50}$. β_1 is expected to be negative because the more negative $CAR_{-20,9}$ is, the more positive $CAR_{10,50}$ is, which means the more prices fall, the more prices reverse in the following days. According to column 2 of Table 3, the value of β_1 is −0.5963, that is, every 1% drop in the returns results in a cumulative reversal of 0.5963% in the following time. T-test shows that the coefficient of $CAR_{-20,9}$ is significant to a degree of 1%, which indicates that the cryptocurrency market with the S&D manipulation is like the stock market. Both of them have the pattern of the extreme price reversal.

4.3 Pattern II: Abnormal Trading Volume

According to the case study in Sect. 3, when the market is manipulated by S&D schemes, the trading volume will surge, significantly higher than that at other times. We test this pattern by the following OLS regression.

First, we define Abnormal Volume (*AV*) to represent the abnormal trading volume of bitcoin:

$$AV = V/EV \tag{2}$$

where *V* refers to volume, and *EV* is expected volume, i.e., the average volume in the estimation window [33].

Then, the following OLS regression is performed:

$$\begin{aligned} CAV_{-20,9} = \beta_0 + \beta_1 CAR_{-20,9} + \beta_2 CAV_{-20,-19} + \beta_3 CBB_{-20,9} \\ + \beta_4 CBB_{-20,-19} + \varepsilon \end{aligned} \tag{3}$$

where $CAV_{-20,9}$ is the cumulative abnormal volume from 20 × 2 h before and 9 × 2 h after the publication of bad news; others have the same meaning as Eq. (1).

According to Sect. 4.3, the negative value of $CAR_{-20,9}$ represents an obvious decline in the returns of bitcoin, which can be used as a sign of S&D schemes. Therefore, the coefficient β_1 reflects the abnormal trading volume caused by S&D schemes. As column 3 in Table 3, when the bitcoin returns fall by 1%, the abnormal trading volume increases by 1.4992%. T-test is significant at the level of 1%, which shows that when the bitcoin market suffers from S&D manipulation, there will be a significant surge in trading volume.

Table 3. Results of OLS regressions of S&D schemes in the bitcoin market.

Independent variables	Dependent variables		Independent variables	Dependent variables
	$CAR_{10,50}$	$CAV_{-20,9}$		$\Delta S2$
$CAR_{-20,9}$	−0.5963*** (−14.062)	1.4992*** (3.407)	$CAR_{-1,1}$	−0.5416*** (−7.913)
$CAR_{-20,-19}$	0.1133 (0.302)	----	H	0.0230 (0.441)
$CAV_{-20,-19}$	----	0.8172*** (51.334)	GP	0.0069 (0.126)
$CBB_{-20,9}$	−0.0019 (−0.590)	0.0333 (0.955)	ABT	0.0262 (0.609)
$CBB_{-20,-19}$	0.0378 (0.769)	−0.5531 (−1.040)	GT	0.0785 (1.572)
β_0	−0.0155	−2.5752	$CFGI$	−0.0398 (−1.134)
			β_0	−0.6070

t statistics in parentheses. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

4.4 Pattern III: Widening of Bid-Ask Spreads

It is indicated in [23] that when informed trading occurs in the stock market, the market makers usually increase the bid-ask spread to cover the potential losses in the future. The manipulator of S&D schemes has already bought or sold certain assets before the publication of bad news, and this constitutes informed trading [23, 34]. Besides, the market makers are only passive receivers of the bad news because they can't control the behavior of the manipulator. The publication of the bad news is bound to bring some uncertainty to the market. Therefore, from the perspective of the market makers, they can only expand bid-ask spreads to make up for some potential losses in the future. In the following, we investigate whether the pattern of widening of bid-ask spreads exists in the cryptocurrency market.

Since we could only find the daily data of some variables such as spreads, we define the dependent and independent variables and test Pattern III with the time granularity of 1 day.

After defining ΔS_2 as the change of the bid-ask spreads of bitcoin from day 0 to day 2 according to the work of Mitts [23], the OLS regression is performed:

$$\Delta S_2 = \beta_0 + \beta_1 CAR_{-1,1} + \beta_2 H + \beta_3 GP + \beta_4 ABT + \beta_5 GT + \beta_6 CFGI + \varepsilon \tag{4}$$

where ΔS_2 represents the change of bid-ask spreads. $CAR_{-1,1}$ refers to the cumulative abnormal returns from the day before to the day after the publication of bad news (Note that the time interval is 1 day). Other variables' meanings can be found in Table 2.

The results of OLS regression are shown in column 4 of Table 3. β_1, the coefficient of $CAR_{-1,1}$, is equal to -0.5416, which indicates that every 1% drop in bitcoin returns increases the bid-ask spreads by 0.5416% two days later. T-test is significant at the 1% level, which strongly proves that the emergence of S&D manipulation in the bitcoin market leads to the widening of bid-ask spread, and shows that S&D manipulation exists in the bitcoin market to a certain extent.

5 Detection of Short and Distort Schemes

In this section, the random forest model is employed to detect S&D schemes in the real market with the patterns obtained in Sect. 3 and Sect. 4, compared with other four detection algorithms.

5.1 Feature Engineering

Section 3 and Sect. 4 study the characteristics of S&D schemes in the cryptocurrency market, showing that markets manipulated by S&D schemes often have the abnormal price and trading volume. Thus, it is possible to judge whether there exist S&D manipulations through the movement of market indicators such as price and volume, which should be also considered in feature engineering.

There are mainly the following considerations in feature selection. First, S&D schemes are phenomena of the cryptocurrency market over a period, rather than at a certain point. Therefore, although the label is marked on a certain day, it is difficult to detect S&D only with the data of that time point. Instead, the market movement characteristics of prices and trading volumes within a time window must be used.

Second, according to the above research, price and trading volume contain much useful information, and thus are important for S&D schemes. It is necessary to fully explore the information contained therein. We focus on processing features by the raw data of price and trading volume, such as average value, maximum, minimum, and standard deviation within the time window. Table 4 shows all the features in the detection model, which are extracted from the time windows.

Table 4. Features used in the detection model.

Feature	Description
P_avg	The average price
P_max	The maximum price
P_min	The minimum price
P_poc	The percentage of price change, defined as the maximum price minus the minimum price divided by the average price
P_std	The standard deviation of the prices
V_avg	The average volume
V_max	The maximum volume
V_min	The minimum volume
V_poc	The percentage of volume change, defined as the maximum volume minus the minimum volume divided by the average volume
V_std	The standard deviation of the volumes

5.2 Detection Model

Sample. The raw data of time series used in the detection model are consistent with regression analysis in Sect. 4. See Sect. 4.1 for details. We set the time window as 20 h (10 time steps). After sliding time window processing and feature engineering, there are 20,046 samples in total, including 3,856 positive cases ($S\&D = 1$) and 16,190 negative cases ($S\&D = 0$).

Model. In the cryptocurrency market, to detect whether S&D appears (`TRUE`) or not (`FALSE`), the detection model is as follows:

$$S\&D = DA(feature1, feature2, \ldots) \tag{5}$$

First, features ($feature1, feature2, \ldots$) stated in Sect. 5.1 are input into the detection algorithm DA. Then, the algorithm DA outputs the result. A binary variable $S\&D$ is equal to 1 (`TRUE`) when S&D exists in the market, otherwise 0 (`FALSE`).

The detection algorithm DA we employ here is Random Forest (RF), which is a superior ensemble learning algorithm. It uses decision trees as individual learners, and finally outputs the result by integrating the voting of different decision trees. RF is simple and easy to implement with great advantages in dealing with multivariate and overfitting, and usually converges to a lower generalization error with the increase of the number of individual learners.

Since we are the first to study and detect S&D in the cryptocurrency market, there are no previous works to compare. We can only select the other four classical and commonly used algorithms as DA in the model for comparison, which are Naive Bayes Classifier (NBC), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Back Propagation (BP).

5.3 Evaluation

Model Evaluation. First, we calculate the confusion matrices, which gives the true and false results of binary classification [35]. According to confusion matrices, some metrics can be further defined to measure the performance of the models, which is commonly used to evaluate the performance of binary classification, as listed in the first column of Table 5.

Table 5. Performance measures.

Metric	Formula	NBC	SVM	KNN	BP	RF
Error	$(FP + FN) / N$	0.4343	0.1519	0.2246	0.1756	**0.0821**
Accuracy	$(TP + TN) / N$	0.5657	0.8481	0.7754	0.8244	**0.9179**
FPR	$FP / (TN + FP)$	0.4198	0.4133	**0.2125**	0.3716	0.2190
Specificity	$TN / (TN + FP)$	0.5802	0.5867	**0.7875**	0.6284	0.7810
Precision	$TP / (TP + FP)$	0.8499	0.9030	0.9389	0.9083	**0.9483**
Recall	$TP / (TP + FN)$	0.5622	0.9099	0.7726	0.8708	**0.9503**
$F1$-score	$2*TP / (N + TP - TN)$	0.7343	0.9065	0.8476	0.8892	**0.9493**

$N = TP + FN + FP + TN$ represents the total number of test samples, and the deeper the background's color is, the larger the value is. Red indicates the best performance, while the lower Error and *FPR*, and the higher the others, the better.

Error represents the ratio of misclassified samples to the total samples, while Accuracy is just the opposite. Row 2 and row 3 in Table 5 show that RF has the best performance, with an Accuracy of 91.79% and an Error of 8.21%.

Recall (i.e., True Positive Rate (*TPR*) or Sensitivity) is the ratio of the number of samples correctly classified as S&D to the total number of true S&D samples. Recall of RF is the highest (95.03%), followed by SVM (90.99%). BP and KNN are in the middle (lower than 90%), while NBC is the lowest (56.22%).

False Positive Rate (FPR) is the ratio of the number of samples classified as S&D to the total number of true non-S&D samples. FPR of KNN is the lowest (21.25%), followed by RF (21.90%). Both SVM and BP are around 40%, while NBC is the highest (41.98%), indicating that NBC misclassified non-S&D as S&D in many samples. Although FPR of RF is not better than that of KNN, they are similar. What's more, Recall of RF is much higher than that of KNN.

Specificity is the rate of correctly classifying negative cases. In the view of Specificity, KNN is the highest (78.75%), NBC is the lowest (only 58.02%), and the other three models are in the middle (about 60%).

Precision represents how many samples predicted as S&D are true. For Precision, RF is the highest, reaching 94.83%, while the other models are a bit lower than RF.

Moreover, the commonly used evaluation metric is $F1$-score, which is the balance of Recall and Precision [35]. As shown in the last row of Table 5, F1-score of RF is the highest at 94.93%, followed by SVM at 90.65%, and the other three models are all lower than 90%.

AUC is the Area Under Receiver Operating Characteristic Curve (ROC), and the larger values mean the better performances. Figure 3 plots the ROC curves. As can be seen from the legend below in the right, AUC of RF is the largest (0.96), followed by KNN (0.85) and BP (0.83). AUCs of SVM (0.68) and NBC (0.57) are much lower than them.

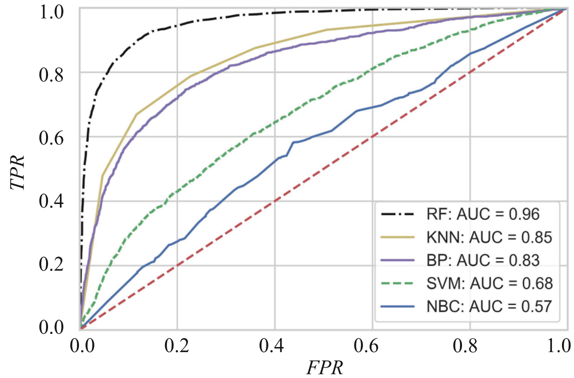

Fig. 3. ROC curves and AUC.

To sum up, as shown in Table 6, from all aspects, RF, the detection algorithm we select, has the best performance compared with the other four methods. BP and SVM also show great performances. KNN follows them, while NBC, as a primary classification algorithm, is not satisfactory at all.

Table 6. Comparisons of performances.

Metric	Rank
Error and accuracy	RF > SVM, BP > KNN > NBC
TPR and *FPR*	RF > SVM > BP, KNN > NBC
Sensitivity and specificity	RF > KNN > BP, SVM > NBC
Precision, recall and $F1$-score	RF > SVM, BP > KNN > NBC
ROC and AUC	RF > KNN, BP > SVM > NBC

Feature Evaluation. Feature Importance Score [35] obtained by rf can be used to evaluate the role of features when detecting S&D in the market. Figure 4 shows the importance scores of 10 features based on the Gini coefficient.

In Fig. 4, the top two features are Minimum Price (*P_min*) and Maximum Price (*P_max*), which suggests that they are more important in detection. This may be because when S&D schemes appear in the market, the prices decline sharply (as described in Pattern I). The minimum and maximum prices catch this tendency, resulting in higher feature importance scores.

In addition, we can also find that in RF detection model, the features of prices (or returns) are more important than those of volumes, such as *P_ avg* and its corresponding *V_ avg*. This may be because the price is a more native variable. People react to the market by quotes, while volume is only a consequence of people's trading behavior as a secondary variable. This suggests that people should pay more attention to price features when studying S&D schemes.

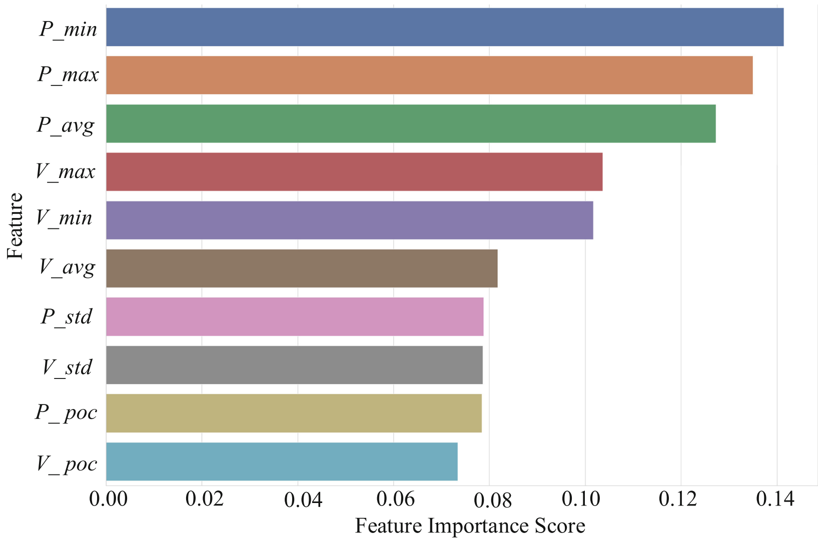

Fig. 4. Feature importance score obtained by RF.

6 Conclusion

This paper studies a new manipulation, Short and Distort in the cryptocurrency market. Through a typical case study, we state the process of the fraud. Using event study methodology and OLS regression, we verify three patterns of Short and Distort, which are price reversal, abnormal trading volume, and widening of bid-ask spreads. Detection of Short and Distort is conducted by the random forest algorithm, which gets the best performance comparing with the other four methods. Our detection model can effectively remind investors away from the market when the fraud scheme comes and provide a method for regulators.

Acknowledgement. This work was supported in part by the National Key Research and Development Program of China (2018YFB1800204), the National Natural Science Foundation of China (61972219), the Research and Development Program of Shenzhen (JCYJ20190813174403598, SGDX20190918101201696), the National Key Research and Development Program of China (2018YFB1800601), the Overseas Research Cooperation Fund of Tsinghua Shenzhen International Graduate School (HW2021013), the Peng Cheng National Laboratory Project (Grant No. PCL2021A02), and Guangdong Major Project of Basic and Applied Basic Research (2019B030302002).

References

1. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. Decentral. Bus. Rev. 21260 (2008)
2. Cryptocurrency Monitoring Website. http://coinmarketcap.com. Accessed 21 Mar 2021
3. Haigh, T., Breitinger, F., Baggili, I.: If I had a million cryptos: cryptowallet application analysis and a trojan proof-of-concept. In: International Conference on Digital Forensics and Cyber Crime (ICDF2C) 2018, LNICST, vol. 259, pp. 45–65. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05487-8
4. MacRae, J., Franqueira, V.N.: On locky ransomware, al capone and brexit. In: International Conference on Digital Forensics and Cyber Crime (ICDF2C) 2017, LNICST, vol. 216, pp. 33–45. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73697-6
5. Gandal, N., Hamrick, J.T., Moore, T., Oberman, T.: Price manipulation in the bitcoin ecosystem. J. Monet. Econ. **95**, 86–96 (2018)
6. Bartoletti, M., Pes, B., Serusi, S.: Data mining for detecting bitcoin ponzi schemes. In: 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), pp. 75–84, IEEE, Zug, Switzerland (2018)
7. Weiner, P.M., Weber, R.D., Hsu, K.: The growing menace of "short and distort" campaigns. Thomson Reuters Exp. Anal. (2017)
8. Weiner, P.M., Totino, E.D., Goodman, A.: SEC issues warning to analysts profiting from "short and distort" schemes, opens the door for civil claims. J. Invest. Compl. (2019)
9. Aggarwal, R.K., Wu, G.: Stock market manipulations. J. Bus. **79**(4), 1915–1953 (2006)
10. Massoud, N., Ullah, S., Scholnick, B.: Does it help firms to secretly pay for stock promoters? J. Financ. Stab. **26**, 45–61 (2016)
11. Huang, Y.C., Cheng, Y.J.: Stock manipulation and its effects: pump and dump versus stabilization. Rev. Quant. Financ. Acc. **44**(4), 791–815 (2013). https://doi.org/10.1007/s11156-013-0419-z
12. Kamps, J., Kleinberg, B.: To the moon: defining and detecting cryptocurrency pump-and-dumps. Crime Sci. **7**(1), 1–18 (2018). https://doi.org/10.1186/s40163-018-0093-5
13. Hamrick, J.T., et al.: The economics of cryptocurrency pump and dump schemes. https://papers.ssrn.com/sol3/pa-pers.cfm?abstract_id=3310307 (2018)
14. Li, T., Shin, D., Wang, B.: Cryptocurrency pump-and-dump schemes. https://papers.ssrn.c-om/sol3/papers.cfm?abstract_id=3267041 (2020)
15. Dhawan, A., Putnins, T.J.: A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets. In: Pump-and-Dump Manipulation in Cryptocurrency Markets. https://pape-rs.ssrn.com/sol3/papers.cfm?abstract_id=3670714 (2020)
16. Victor, F., Hagemann, T.: Cryptocurrency pump and dump schemes: quantification and detection. In: International Conference on Data Mining Workshops (ICDMW) 2019, pp. 244–251. IEEE, Beijing, China (2019)

17. La Morgia, M., Mei, A., Sassi, F., Stefa, J.: Pump and dumps in the bitcoin era: real time detection of cryptocurrency market manipulations. In: 29th International Conference on Computer Communications and Networks (ICCCN), pp. 1–9. IEEE, Honolulu, HI, USA (2020)
18. Mirtaheri, M., Abu-El-Haija, S., Morstatter, F., Steeg, G.V., Galstyan, A.: Identifying and analyzing cryptocurrency manipulations in social media. IEEE Trans. Comput. Soc. Syst. **8**(3), 607–617 (2021). https://doi.org/10.1109/TCSS.2021.3059286
19. Mansourifar, H., Chen, L., Shi, W.: Hybrid cryptocurrency pump and dump detection. arXiv 2003.06551, arXiv preprint arXiv:2003.06551 (2020)
20. Nilsen, A.I.: Limelight: real-time detection of pump-and-dump events on cryptocurrency exchanges using deep learning. https://munin.uit.no/bitstream/handle/10037/15733/thesis.pdf?sequence=2&isAllowed=y (2019)
21. Chen, W., Xu, Y., Zheng, Z., Zhou, Y., Yang, J.E., Bian, J.: Detecting "pump & dump schemes" on cryptocurrency market using an improved apriori algorithm. In: 2019IEEE International Conference on Service-Oriented System Engineering (SOSE), pp. 293–2935. IEEE, San Francisco, CA, USA (2019)
22. Xu, J., Livshits, B.: The anatomy of a cryptocurrency pump-and-dump scheme. In: 28th USENIX Security Symposium, pp. 1609–1625. USENIX Association, Santa Clara, CA, USA (2019)
23. Mitts, J.: Short and distort. J. Leg. Stud. **49**(2), 287–334 (2020)
24. Zhao, W.: Activist short-selling. https://www.proquest.com/openview/1559f45788a7fb12-5c3f2282c1e2a875/1?pq-origsite=gscholar&cbl=18750 (2017)
25. Ogorodniks, A., Sirbu, A.: Activist short selling campaigns: informed trading or market manipulation? https://www.sseriga.edu/sites/default/files/2020-05/1Paper_Ogorodniks_S-irbu.pdf (2018)
26. Kartapanis, A.: Activist short-sellers and accounting fraud allegations. https://repositories.lib.utexas.edu/bitstream/handle/2152/74990/KARTAPANIS-DISSERTATION-2019.pdf?-sequence=1&isAllowed=y (2019)
27. Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S., Tesconi, M., Ferrara, E.: Charting the landscape of online cryptocurrency manipulation. IEEE Access **8**, 113230–113245 (2020)
28. TradingView Website. https://www.tradingview.com/chart/. Accessed 31 Mar 2021
29. Bouoiyour, J., Selmi, R.: Coronavirus spreads and bitcoin's 2020 rally: is there a link? https://hal.archives-ouvertes.fr/hal-02493309/ (2020)
30. Shahzad, S.J.H., Bouri, E., Roubaud, D., Kristoufek, L., Lucey, B.: Is bitcoin a better safe-haven investment than gold and commodities? Int. Rev. Financ. Anal. **63**, 322–330 (2019)
31. Härdle, W.K., Trimborn, S.: Crix or evaluating blockchain based currencies. https://www.econstor.eu/handle/10419/122006 (2015)
32. Tetlock, P.C.: All the news that's fit to reprint: do investors react to stale information? Rev. Finan. Stud. **24**(5), 1481–1512 (2011)
33. MacKinlay, A.C.: Event studies in economics and finance. J. Econ. Literat. **35**(1), 13–39 (1997)
34. Feng, W., Wang, Y., Zhang, Z.: Informed trading in the bitcoin market. Financ. Res. Lett. **26**, 63–70 (2018)
35. Zhihua, Z.: Machine Learning. Tsinghua University Press, Beijing (2016)