

How Wash Traders Exploit Market Conditions in Cryptocurrency Markets*

Hunter Ng
Baruch College
hunterboonhian.ng@baruch.cuny.edu

15th October 2024

Abstract

Wash trading, the practice of simultaneously placing buy and sell orders for the same asset to inflate trading volume, has been prevalent in cryptocurrency markets. This paper investigates whether wash traders in Bitcoin act deliberately to exploit market conditions and identifies the characteristics of such manipulative behavior. Using a unique dataset of 18 million transactions from Mt. Gox, once the largest Bitcoin exchange, I find that wash trading intensifies when legitimate trading volume is low and diminishes when it is high, indicating strategic timing to maximize impact in less liquid markets. The activity also exhibits spillover effects across platforms and decreases when trading volumes in other asset classes like stocks or gold rise, suggesting sensitivity to broader market dynamics. Additionally, wash traders exploit periods of heightened media attention and online rumors to amplify their influence, causing rapid but short-lived spikes in legitimate trading volume. Using an exogenous demand shock associated with illicit online marketplaces, I find that wash trading responds to contemporaneous events affecting Bitcoin demand. These results advance the understanding of manipulative practices in digital currency markets and have significant implications for regulators aiming to detect and prevent wash trading.

Keywords: Wash Trading, Cryptocurrencies, Bitcoin, Market Manipulation, High-Frequency Trading, Trading Volume, Information Asymmetry, Financial Markets, Exchange Activity, On-Chain Transactions, Strategic Trading, Market Regulation, LTSM

JEL Classifications: G12, G13, G18, G23, D82, K22

*I am grateful to Lin Peng for helpful comments and feedback.

1. Introduction

Wash trades are trades where a market participant places a buy order at a specific price for a product, and simultaneously places a sell order for the same product, knowing that the two orders will match. This artificially inflates the volume of trading around a product, which is prosecuted by Section 17(a) of the 1933 Securities Act. Traditionally, financial products such as stocks and commodities have been regulated by global regulatory bodies using wash-trading surveillance models (CFTC, 2024).

Wash-trading in digital assets is a relatively new phenomenon with the inception of Bitcoin in 2009. Digital assets are split into two categories by the Securities and Exchange Commission (SEC). A digital asset can be classified as a security if it is intended to be used the same way as a stock, bond, or other investment assets. Alternatively, it can be exempt from registration if it is purely used as a payment channel, such as Uber Cash. If it is a security, it comes under SEC's regulation and wash-trading is illegal.

In this article, I want to examine the question of whether cryptocurrency trades are strategic and in turn, identify characteristics of wash trading. This is of extreme importance to regulators in characterizing and prosecuting wash trading of digital assets.^{1,2} For example, in a detailed amended complaint filed by SEC on April 18, 2024 against BitTorrent Foundation Ltd. and others, alleged wash-trading activity between two accounts were tracked and used in court proceedings (Securities and Exchange Commission v. Sun, 2023).

The results of this study find that the volume of wash trades increase when volume of non-wash trades is low, and conversely, wash-trades go down when non-wash trades increase. Wash trading's effects on one platform also spread to other trading platform. When other classes of assets such as stocks or gold are trading more, wash trade's volume reduces. Importantly, this article also finds that wash traders capitalize on rumors and information asymmetry such that when media interest on cryptocurrencies is high, wash trades cause a much more rapid spike in non-wash trade volume but this effect subsides quickly. Lastly, using an exogenous quasi-natural shock in demand, I show that wash trade's effectiveness is limited by contemporaneous events which affect the underlying cryptocurrency's demand.

¹SEC filed 13 charges against Binance Founder and one of the charges alleges that Sigma Chain, an entity controlled by Binance Founder Changpeng Zhao engaged in wash trading to inflate the platform's trading volume.

²SEC also took action against crypto entrepreneur Justin Sun and associates for fraudulently manipulating the secondary market for crypto asset securities Tronix (TRX) through extensive wash trading. The case also involved notable celebrities such as Lindsay Lohan, Jake Paul and others.

To explore the questions, I rely on internal transaction logs from Mt.Gox, the largest bitcoin exchange by volume at one point in time. These leaked data were released publicly following the exchange's closure in 2014. Gandal et al. (2018) show that bots used by Mt. Gox increased wash trading volumes, leading to a rise in the Bitcoin/USD exchange rate. They specifically identify the trading bots "Willy" and "Markus" and demonstrate that the increased volume resulted in widespread ripple effects to other exchanges. Subsequent analyses suggest that other wash trades during the Mt. Gox period, apart from the Mt. Gox bots, are characterized by lower prior trading volumes. In this paper, I focus on the leading and lagging factors of wash trades to better characterize them.

My article is related to three strands of literature. The first is the literature of how manipulative traders gain financial advantages. Models by Glosten and Milgrom (1985) and Allen and Gale (1992) discuss uninformed manipulators disguising themselves as informed traders to extract rents, with varying frequencies of trading among informed traders (Kyle, 1985; Holden and Subrahmanyam, 1992). Huddart (2003) introduces the concept of dissimulation in informed trading models, where insiders add random noise to obscure their private information. Huberman and Stanzl (2004) write that markets are viable only in the absence of quasi-arbitrage and this also characterizes the permanent price impacts of manipulative trading as linear.

Secondly, this article is related to the legal literature on stock market manipulation and its regulation. Fox et al.'s, (2018) paper shows that market manipulation's definition in today's legal framework is unclear and must be tackled from a legal and microeconomic theory angle. Kyle and Viswanathan (2008)'s definition of illegal manipulation is when a violator "pursues a scheme that undermines economic efficiency both by making prices less accurate as signals for efficient resource allocation and by making markets less liquid for risk transfer". In classifying any form of trading as manipulative, unverified empirical reports are admissible in different courts in securities class actions (Schneider, 2024), which emphasizes the importance of characterizing manipulative trading. In SEC v. Sun, 2023, the defendants' counter argues that wash trades alleged by SEC constitute 0.0039% and 0.0017% of the total dollar value of the cryptocurrency. It further adds that if the alleged trades shown by SEC were wash trades, then "any high-frequency trader would also be complicit in wash trading on Wall Street". This statement shows that unlike traditional stocks, cryptocurrency wash trades cannot be classified simply by volume, but instead, research can focus on other strategic factors of wash-traders³.

³Characterizing different aspects of wash-trading can also help in legal investigations, which is one of the motivations behind this paper. From the defence team's statement in Securities and Exchange Commission v. Sun, No. 1:23-cv-02433 (S.D.N.Y. Mar. 22, 2023), defense lawyers argued that wash-trading of the cryptocurrency is only a minor fraction of the total market capitalization and cited it as grounds to dismiss wash-trading charges.

The third strand is the literature on cryptocurrency wash trading. Studies evaluate statistical methods to determine actual wash trading activity (Aloosh and Li, 2024). Cong et al. (2023) use proprietary data from centralized exchanges to infer that more than 70% of trading volume is wash trading, and their models have aided federal investigations. Kumar et al. (2022) demonstrate spillover effects in crypto markets, suggesting that wash trading can impact not just a single cryptocurrency but potentially other digital assets as well.

This paper adds to the literature with three key contributions. First, it finds that wash trades in cryptocurrency are negatively correlated with liquidity in other asset classes. Wash traders reduce their activity when other markets are more liquid, possibly due to fears that increased liquidity elsewhere might dilute the impact of their manipulative activities. Second, wash traders exploit online rumors to amplify the effects of their wash trading, but these effects dissipate more rapidly when media attention is high. This provides empirical support for models of investor information processing costs (Blankespoor et al., 2022). Third, using an exogenous event known as Pot Day⁴, the study shows that wash trades respond to contemporaneous market events affecting demand for the underlying currency. This suggests that wash trading can be characterized not only through statistical distributions but also through strategic factors responsive to market conditions.

The rest of the paper is structured as follows. Section 2 discusses the the dataset, existing literature, and develops hypotheses for empirical study. Section 3 shows the main empirical results. Section 4 concludes.

2. Background and Literature Review

In early 2014, a data leak from Mt. Gox provided access to approximately 18 million buy and sell transactions from April 2011 to November 2013. This dataset differs from proprietary exchange data or Bitcoin blockchain transactions because it consists of back-end files recorded by the exchange itself. While blockchain transactions are publicly available, they often do not account for fiat-to-crypto transactions. Many exchanges also do not record internal trades on the public blockchain, instead managing trades among internal wallets and ensuring adequate crypto assets for deposits and withdrawals. Proprietary data released by exchanges may be filtered, potentially omitting undesirable transactions before

⁴Pot Day is a widely studied event which happened on April 20, 2012. At that time, Silk Road, which was the single largest online black market in the Dark Web, advertized this day massively, where many top sellers would offer large promotional sales on cannabis. This lead to many sellers entering the marketplace two weeks prior to the operation. It is also widely documented that during this period, Bitcoin was the primary currency of transaction in the Dark Web.

sharing with third-party providers. Notably, the leaked Mt. Gox data includes a separate file for April 2013, which is speculated to contain altered transactions intended for auditors and investigators.

The data includes transaction ID, amount, time, currency, transaction fees and other peripheral information. Each Mt. Gox user has a unique ID number, allowing transactions to be linked to specific actors. I supplement this dataset with additional variables from various sources. A list of these variables is provided in Table 1.

Characterizing cryptocurrency wash trading is a growing area of research. Gandal et al. (2018) find evidence of wash trading in the Mt. Gox dataset, identifying two bots—“Markus” and “Willy”—that executed suspicious wash trades from 2013 onwards. Aloosh and Li (2024) identify wash trading from June 2011 to May 2013 characterized by the same buyer and seller trade IDs. Chen et al. (2019) use the ratio of a digital currency exchange’s media popularity to its off-chain transactions as a proxy to estimate the severity of wash trading. Cong et al. (2023) identify evidence of wash-trading using proprietary data from several exchanges through novel statistical methods.

This article correlates strategic factors with verified wash trade activity. While empirical methods can identify wash trading using statistical techniques, pinpointing wash trading activity is challenging. Whether bad actors engage in strategic wash trading remains an open question that could enhance current methods of identifying wash trades.

2.1 Cryptocurrency wash trading is different from traditional wash trading

In models of trade manipulation (Allen and Gale, 1992; Aggarwal and Wu, 2006), a pooling equilibrium exists between informed traders and manipulative traders. Manipulators execute buy orders to drive prices up and then sell off, often spreading rumors to amplify the effect—a significant portion of which occurs online. Aggarwal and Wu (2006) review SEC cases and find that 54% of manipulators both buy and sell stocks to profit, while about 13% corner the supply to inflate prices. They also find that these manipulations often involve penny stocks, which require less capital and have noisier information environments.

While the market for traditional assets is well-studied in the trade manipulation literature, relatively less is known about cryptocurrencies, which do not represent any claims to fiat currency or ownership, and do not represent a tangible hold on physical goods. Gandal et al. (2018) find that 843 different cryptocurrencies with large volume jumps due to manipulative activities have their prices increased

significantly. This article uses the Mt. Gox dataset to explore strategic factors affecting the dynamics of wash trades.

2.2 Strategic wash trading against on-chain transactions, market transactions and other assets

From earlier, theory suggests that wash traders disguise as informed traders and thus, manipulate other investors into buying the cryptocurrency. As Bitcoin is also a speculative tool, wash traders may observe price movements in traditional commodities like gold, which is often considered as a hedge against inflation, to adjust their strategies. By monitoring correlations between assets, traders may artificially inflate cryptocurrency trading volumes to create misleading market activity based on perceived price relationships.

For instance, if gold prices are rising, a wash trader might simultaneously buy and sell Bitcoin to mimic this trend, creating the illusion of increased investor interest in both assets. This exploits behavioral biases, leading other market participants to perceive the cryptocurrency's price movement as a response to gold's trend.

In this paper, I use high-frequency data of assets to study if there is a correlation between these two factors. High-frequency data captures price movements at a minute-by-minute level and allows me to detect anomalies that may be invisible in lower-frequency data. This granularity enables the identification of the exact timing and extent of wash trades.

H1: Bitcoin's wash trading on one exchange leads trading in other platforms

H2: Bitcoin's wash trading decreases when liquidity in other asset classes increase

2.3 Strategic wash trading through spreading rumors and dark web illicit activities

Aggarwal and Wu (2005) also show that the spread of rumors can explain why uninformed, noisy traders converge on the volume surge and cause the cryptocurrency to rally. In 2011 to 2013, Bitcoin was nascent and Bitcoin trade mainly comprised two uses. Firstly, Bitcoin was used as a currency to perform transactions on the dark web (Christin, 2012; Foley et al, 2019). Many studies show that Bitcoin was favored as a transaction platform for illicit activities due to its anonymous nature and lack of regulation. As much as 76 billion dollars of illegal activity per year involve bitcoin in later years, while Christin (2012) documents that almost 20% of all Bitcoin transactions were through the Silk Road, a large dark web marketplace for goods and services.

Second, Bitcoin served as a speculative instrument. The speculative nature of Bitcoin is also well-documented, with the meteoric rise in price from \$0.01 in 2010 to a high of \$73,737 in March 2024. Schilling and Uhlig (2019) show that a two currency economy with fiat and bitcoin can reach equilibrium with bitcoin speculation. Both speculation and currency usage drive the price of bitcoin. As such, press and Internet coverage and rumors could drive speculation, which affects the price of cryptocurrencies. Wash traders could then potentially use this channel to enhance their wash trades. Thus, I test whether a weaker information environment is exploited by wash traders to strategically increase their volume of wash trades.

The Dark Web Silk Road was an online black market, notorious for facilitating the anonymous trade of illegal goods and services. Launched in 2011 by Ross Ulbricht, the platform used the anonymity of the Tor network and Bitcoin transactions to evade law enforcement. At its peak, Silk Road became a haven for illicit activities, predominantly drug trafficking, but also including weapons sales, counterfeit documents, and hacking services. It disrupted traditional drug distribution channels, introduced sophisticated encryption and payment systems, and caused regulators difficulties in policing cyberspace. Silk Road's eventual shutdown in 2013 by the FBI. Because Silk Road transactions were heavily financed by Bitcoin, events on Silk Road affected Bitcoin's demand. One such event is Pot Day on April 20th, 2012 (Christin et al. 2012). I exploit the exogenous shock to show causal evidence that wash trades are strategically increased in the time period leading up to Pot Day.

I state my next two hypotheses in the alternate form below.

H3: Bitcoin's wash trading manipulative effect is higher when media attention is high

H4: Bitcoin's wash trading increases during exogenous demand events

3. Data and Results

The Mt.Gox dataset consists of two folders of "*backoffice*" and "*trades*". "*backoffice*" consists of some peripheral information related to the leak while "*trades*" contain the leaked trade files. I follow the standard de-duplication process in past literature documented in Gandal et al. (2018) and Aloosh and Li (2024), with the details documented in Appendix A1. The extracted variables, including the other variables to be used in the paper, are documented in Table 1.

Based on the variables listed in Table 1, I compiled the descriptive statistics for selected key variables in Table 2.

3.1 Do crypto wash trades' volume or value matter?

To answer the question of whether BTC wash trade volumes matter, I first identify the total circulating supply of Bitcoins that is in the market. New Bitcoins are minted everyday and is expected to stop once the total supply reaches 21 million BTC in year 2140. Market capitalization is an important variable because it has been cited as a point in legal defense in *Securities and Exchange Commission v. Sun, No. 1:23-cv-02433 (S.D.N.Y. Mar. 22, 2023)*, where defense lawyers argued that wash-trading of the cryptocurrency is only a minor fraction of the total market capitalization and cited it as grounds to dismiss wash-trading charges. I gather the total circulating supply of Bitcoins from *bitcoinexplorer.com* on several days, and use linear interpolation to compute a daily count of total circulating supply. In untabulated results, the average wash-trade is $6.5 \times 10^{-5}\%$ of the total market cap. This is much lower than the 0.0039% cited in the aforementioned SEC case. Thus, I first show that capitalization of documented wash trades in the MtGox data is insignificant.

Next, to determine how wash-traders decide when to wash-trade, I first turn to Classification and Regression Trees (CART) models. CART offers a non-parametric approach to predict outcomes based on a series of decision rules derived from the MtGox data. By constructing a binary decision tree, CART models segment the data into subsets that are increasingly homogeneous with respect to the occurrence of wash trades. This segmentation helps to identify critical thresholds and patterns in the data that influence wash trading behavior.

CART also excels in handling complex, non-linear relationships within the data, making it well-suited for financial markets where interactions between variables can be intricate and unpredictable. Consequently, this helps in developing strategies to monitor and potentially mitigate such manipulative trading activities.

I use four popular methods of CART analysis to perform my tests. I use 1) standard CART algorithm, 2) random forest, an ensemble learning method that uses multiple decision trees to improve model accuracy and reduce overfitting 3) gradient boosted tree, where decision trees are built sequentially and each tree corrects the errors of the previous one. I implement XGBoost (Extreme Gradient Boosting) for this, which incorporates advanced features such as regularization to prevent overfitting, parallel processing for faster computation, and handling of missing values. Lastly, I use 4) AdaBoost, an ensemble learning method that focuses on improving the accuracy of weak classifiers by iteratively adjusting the weights of misclassified data points.

I use 5 variables in my CART analysis - 1) the non-wash trade volume (in 30-minute intervals), 2) the wash trade volume (in 30 minute intervals), 3) the total trade volume (in 30-minute intervals), 4) liquidity of BTC market in Mtgox via the Amihud measure and 5) realized volatility of BTC in Mtgox. Before performing CART, I also check that the 5 timeseries are stationary and contain no time trends via the Augmented Dickey Fuller test. In my CART models, I use an arbitrary train-test ratio of 7:3, depth-size of 3, and I set the next period's wash trade volume as the target. In Table 3, I take time lags of $t-1$, $t-2$ and $t-24$. In extended analysis in the Appendix, I take time lags of $t-1$, $t-2$, $t-3$, $t-4$, $t-5$, $t-6$, and $t-24$. For robustness, I include a random placebo to check. Any variables that have lower importance than a random placebo means that for that model, they are disregarded.

To strengthen my analysis, I include two other deep learning models. Firstly, I use Gated Recurrent Units (GRU), a type of recurrent neural network (RNN), to predict the extent of wash trading. GRUs are specifically engineered to manage sequential data by utilizing gating mechanisms that adeptly retain long-term dependencies and control information flow through the network. This feature makes GRUs exceptionally effective in processing time-series data.

Secondly, I use Long Short-Term Memory (LSTM) networks, a sophisticated variant of recurrent neural networks (RNNs). LSTMs are designed to address the vanishing gradient problem, which often plagues traditional RNNs. This problem affects the network's ability to learn and remember long-term dependencies in sequential data. The architecture of LSTM networks includes a unique design that enables them to retain information over extended time periods, making them particularly well-suited for time-series prediction and sequence modeling tasks. For both the deep-learning models, I use the same train-test ratio of 7:3 and 20 epochs with batch size of 32.

A visualization of Table 2 is provided in Appendix A6. As total trade volume may not be robust from wash trade volume, I prioritize the other variables. From Table 3 Columns (3),(4), I can determine that volatility and non-wash volume affects wash-trades, immediately after total trade volume. They are also more important than the placebo, which provides evidence that under decision models, wash traders consider the volume of non-wash trade before making wash trades. In columns (5),(6), the deep learning models instead place the placebo very low in importance, instead showing that non-wash volume in $t - 2$ and $t - 3$ affect the decision to do wash-trading the most. Again, this confirms my hypothesis that wash trades are strategic in nature and are based on volume of non-wash trades.

Next, to determine the conditions that are correlated with wash-trading, I next perform a vector autoregression (VAR) analysis framework. I consider the VAR model below

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t$$

where y_t is a 5x1 vector of endogenous variables where 1) the wash trading volume (BTC units), 2) the non-wash trading volume (BTC units), 3) rate of BTC to USD (per BTC), 4) Amihud measure of liquidity, 5) realized volatility of Bitcoin return. The time series are calculating in intervals of 30-minutes. I confirm that the time series are stationary and contain no time trends, as I reject the unit root hypothesis using Augmented Dickey-Fuller test.

For robustness, I first conduct a Granger Causality test between the variables. The Granger Causality test helps to determine whether one time series can predict another time series. The results show that wash trading volume Granger-causes non-wash trading volume, the BTC to USD exchange rate, the Amihud liquidity measure, and the realized volatility of Bitcoin returns. This causality suggests that fluctuations in wash trading volume can be used to predict future movements in these variables.

Next, I conduct a Johansen cointegration test that identifies the presence of long-term equilibrium relationships between the multiple time series variables. To do this, I estimate a vector error correction model (VECM) derived from the VAR model and the test evaluates the null hypothesis of no cointegration against the alternative of cointegration among the variables. These results are presented in Table 4.

From Table 4, Panel A, the results of the Johansen cointegration test indicate the presence of at least one cointegrating relationship at the 5% significance level. This finding suggests that there exists a stable, long-term equilibrium relationship among the wash trading volume, non-wash trading volume, total volume, the Amihud measure of liquidity, and the realized volatility of Bitcoin returns. The existence of cointegration implies that despite short-term deviations, the variables tend to move together in the long run.

Panel B shows that non-wash and volatility could be causing wash trade activity. Liquidity does not significance, which is in line with the results in Table 3. Panel C, Column (2) then shows that when non-wash trades increase, in $t + 1$ and $t + 2$, the values are negative, suggesting that wash trades decrease when non-wash trades increase. This provides evidence that wash trades go down once non-wash trades increase.

3.2 Can on-chain and cross-market data predict wash-trading?

Wash traders may base their illicit activity off transactions in other exchanges and on-chain transactions. In Chen, Lin and Wu (2021), they use the ratio between an exchange's % of overall market transactions and an exchange's % of overall media popularity as a gauge of potential suspicious activity. Chen, Lin

and Wu (2021)'s method depends on the presence of data of all exchanges, and calculating the relative ratios to determine if wash-trading takes place. It is also dependent on assuming that all transactions must go through the publicly listed hot and cold wallet addresses that have been revealed at some point through Proof-of-Funds disclosure by the exchange. Not all wallet addresses must be disclosed and they should not be, due to security reasons and privacy concerns. Exchanges only typically disclose a subset of wallets to show proof, not an exhaustive disclosure of all funds. Furthermore, Mt Gox's wallet addresses were not publicly available during this period⁵.

Thus, in this section, I want to understand whether volume in other public exchanges and on-chain can serve as leading or lagging factors of wash-trades. My underlying argument is that if there are legitimate transactions based on insider information, on-chain and other exchange transactions should mirror exchange transactions. Similarly, wash trades in one exchange may also bring about legitimate trading activity in on-chain and other exchanges. Understanding the extent and pattern to which these trends correlate is important for characterizing leading and lagging factors of wash-trades.

I first identify on-chain transactions from block 145,500 to block 182,600. On-chain transactions are public information that is stored in nodes across the entire Bitcoin blockchain. There are 9,267,803 transactions in the blocks. These are immutable once recorded, serving as proof of the transactions and also one of the defining feature of blockchain where every past transactions is publicly visible. The blocks represent a time frame from 16th September 2011 to 1st June 2012⁶.

A snippet of the data format is shown in Appendix A3. Afterwards, I collect the transaction volume into 30 minutes, similar to the MtGox dataset. I then split the MtGox dataset into 4 quartiles based on the wash trade volume per day, accumulated from the 48 30-minute intervals in a day. There are a total of 259 days in the data frame, and a visualization of which quartile each day fell into. Next, I perform a linear regression of the $onchain_t$ against $nonwash_t$ split into the 4 quartiles, where $t = 30$ mins. I also perform a two-step Engle-Granger cointegration test between $onchain_t$ and $nonwash_t$. The optimal number of lags is determined by the Akaike Information Criterion (AIC). The results of the regression and cointegration are in Table 3 Panel A and Panel B respectively.

Next, I also identify a list of 30 other exchanges that linked their APIs to *bitcoincharts.com*, building a comprehensive, daily dataset of total BTC (USD) market activity from 26th June 2011 to 20th May

⁵Arkham transactions, the most comprehensive, major cryptocurrency analytics provider's data for MtGox's publicly listed wallet addresses only go back to 2014.

⁶As the program constantly calls the API from blockchain.info and is limited by a timeout function, some blocks may not have been copied due to network error or API error. Retrieving the on-chain data for this time frame took a week of non-stop API calls to complete.

2013. *Bitcoincharts.com* is a resource that is often used to measure total BTC market activity in this time period and the data is high-quality. In Appendix A3, I plot a distribution of MtGox activity against the total BTC-USD market activity at that time. I achieve a mean of 83% which is comparable to past literature. Many exchanges during this time were short-lived, and completeness of the entire market activity may lead to slight differences in Mt Gox's transactions as a % of the entire market activity.

Since the market activity volume is a daily count, I use a regression instead as I am unable to identify precise time trends in a day, unlike the on-chain data. I present the linear regression of the $onchain_t$ against $nonwash_t$ split into the 4 quartiles, where $t = 30$ mins, in Table 3 Panel C.

Table 5 Panel A indicates that in a OLS regression, Quartile 3 non-wash trade volume correlates more closely with on-chain BTC volume. From the Table 5 Panel B, a lower p-value suggests stronger evidence against the null hypothesis of no cointegration, indicating a closer long-term relationship between non-wash trade and on-chain trade. The increasing p-values from Quartile 1 to 4 suggests that when wash-trading is at its peak in volume, it mirrors the on-chain transactions the most. This is counterintuitive but because on-chain transactions represent BTC transactions across the globe, and not just BTC-USD transactions. However, the linearly increasing p-value suggests that there is a strong trend here and I reason that when wash trades are high, it sets off more transactions across other platforms and wallets, thus, periods of high wash trade in the MtGox exchange may mirror on-chain transaction activity more.

From Table 5 Panel C, it shows that when wash-trade activity is high in Quartile 4, it also has the highest coefficient. This mirrors Panel A's and B's results that when wash trading is high, non-wash trade shows a much higher cointegration and correlation to the entire BTC-USD landscape. As explained earlier, this could be because when the market is not trading as much, trading activity on each platform becomes idiosyncratic to its own exchange specific customers but when wash trading is high, there is generally an entire contagion effect among them.

However, all 3 panels may suffer from missing omitted variables due to significance in the constant, and therefore, not be conclusive. For the on-chain results, this could be because off-chain and on-chain activity still differs very significantly, while for the market results, activity in one exchange may not necessarily propagate to other exchanges, else, all exchanges would be equally popular in trading volume but that is not the case as Mt Gox itself controlled about 70% of all BTC trades during the sample's time frame.

3.3 Are wash trades motivated by other assets?

The activity of other assets could also serve as leading and lagging factors of wash trades. My reasoning is that when the volatility of other assets decrease, wash-traders may use this to motivate their wash-trades in order to increase speculation. To determine whether BTC wash trades are correlated with the movement of other assets, I use high-frequency data of 1) SNP 500 index, 2) Gold Spot Price, 3) VIX Index, 4) EUR-USD Exchange.

I first convert the time of each exchange to UTC to match the Mt Gox data. This high-frequency data is in intervals of 1 minute, thus, I accumulate them into 30-minute intervals, and then calculate the Amihud Liquidity and Realized Volatility where available. As these markets are not 24 hours, whenever they are not trading, I impute zeroes to them. I perform a VAR of these variables against $wash_t$ and present the IRF results in Table 6.

In Table 6, we see that generally, when other assets increase in price or volume traded, $wash_t$ decreases. The trends are also very consistent throughout the time periods from $t + 1$ to $t + 10$. This is evidence that wash trades are not random and does not take place when other markets are trading more actively. This is potentially a new angle that investigators can consider since wash trades are meant to maximize BTC activity, it makes sense that wash trades should not be done when other assets are trading more. This is because BTC would itself be cannibalized by these other assets. Thus, wash trades can be taken to be strategic during this period.

3.4 Does media attention affect wash trades' effects?

In this section, I reason that due to a lack of a robust information environment around cryptocurrencies, wash trades can disproportionately affect trading much more than traditional stocks. This is because information structure surrounding traditional stocks and companies have been in existence for a long time and many stakeholders, including analysts, employees, regulations and other shareholders add to this rich information environment (Asquith, Mikhail and Au, 2005; Babenko and Sen, 2015). Thus, I hypothesize that wash trades are strategically executed when the information environment surrounding the underlying digital currency is weak and rumors are rife. To proxy for this, I use Google media trends to study this. I take a binary variable of whether the online interest score is above or below the median, and then perform a VAR in Table 7.

From Table 7, we see that when there is a huge media search popularity on Bitcoin, non-wash trades react much faster to wash trades from Column (6) in $t + 1$ compared to Column (2). However, the sum

for Columns (1),(2) also are more than Columns (5),(6). This could mean that the wash trades affect non-wash trades much faster when there is more media intensity but the effects also subside much more quickly compared to when there is less media attention. Overall, this means that wash traders could be capitalizing on rumors and information asymmetry, such that when media interest is high, to do wash trades that would have a much more rapid effect but also tends to subside faster.

3.5 Exogenous shock - Pot Day, 20th April 2012

Lastly, to determine if wash trading is strategically motivated by actual demand for Bitcoin, I use an exogenous shock. Pot Day, 20th April 2012 was a documented day of big sales surrounding cannabis. According to Christin (2012), there was large promotional sale on Silk Road (a popular Dark Web marketplace at that time) on this day, where there were a large number of sellers entering the marketplace two weeks before the actual day to sell cannabis. The majority of these trades took place using Bitcoin.

I hypothesize that wash traders may respond correspondingly to this exogenous shock and increase their wash trade activity before Pot Day to take advantage of the increased demand for BTC. There would also be a much bigger bigger response

From Table 8, by observing the sum of the %IRF, we see a big drop in the *sum* IRF of both $total_t$ and $nonwash_t$ from Columns (1),(2) versus Columns (3),(4), from $7.54e+04$ to $4.63e+04$. This shows that the trading activity responded less to wash trades in the direct aftermath of Pot Day. This provides strong evidence that wash trade's effectiveness is limited by contemporaneous events which affect BTC's basic demand. When non wash trade increased, Columns (4) and (8) also show that wash trade generally decreased, which is in line with predictions that when the market returns to an active state, wash traders focus on trading with legitimate people compared to performing wash trades. The bigger $-8.61e+03$ in Column (4) compared to the smaller $-6.94e+04$ in Column (8) also shows that wash traders reduce their wash trading less in the aftermath of Pot Day when non wash trades increase, showing that they did more wash trades to hype up the market before Pot Day and decreased afterwards. This shows that wash trade behavior was strategic during this period, as they correlated with contemporary events and kept up wash trades more during popular times of BTC demand.

4. Conclusion

The results in this paper show that wash-trading is indeed strategic and do not occur in a vacuum. I document certain strategic factors that may be of interest to federal investigators and other cryptocurrency stakeholders who may be interested in investigating wash-trading activity.

Firstly, lagged non-wash trade volume and volatility triggers more wash trading, showing that wash trading takes place strategically. The results are robust to a placebo test. Secondly, when wash-trading in an exchange is high, the exchange's activity mirrors the on-chain transactions and other markets' transactions. This is most probably because when trading is low, the on-chain transactions and activity in other markets are more idiosyncratic to their own customers and more general. When wash trading is high, the contagion effect causes every one's activity to be more similar.

Thirdly, wash-trading also takes place more when trading on other important asset classes are lower in volume. Fourthly, when online rumor interest is high, wash trades will cause the volume traded to go up fast but also recedes more quickly than if there is no media interest. In other words, when the information environment is weaker such as during lower google search interest time periods, other legitimate cryptocurrency transactions could be fueled by information asymmetry according to traditional wash trading theory and the response is slower but longer-lasting. Finally, when there are contemporary demand events that would likely trigger more customers to buy bitcoin, wash trading increases more to take advantage of this speculation.

One limitation of this article is that by using the Mt.Gox dataset, there is a tradeoff between verifiability and timeliness. Current wash trading activity may differ from the past during Mt.Gox's time due to the different macroeconomic factors surrounding BTC currently.

Future works can repeat the techniques in this article on newer, proprietary information from exchanges, other than the Mt.Gox dataset. While current research generally relies on general statistical trends such as Zipf's Law, Benford's Law and trade-size clustering to *prima facie* identify wash-trading on an exchange-level basis, finer measures such as CART and LSTM models can profile wash-trading if one has access to the exchange's transactions. This can help to combat wash-trading and also assist the legal scene in prosecuting wash traders.

References

- Aggarwal, R. K., & Wu, G. (2006). Stock Market Manipulations. *The Journal of Business*, 79(4), 1915–1953. <https://doi.org/10.1086/503652>
- Allen, F., & Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, 5(3), 503–529.
- Aloosh, A., & Li, J. (2024). Direct Evidence of Bitcoin Wash Trading | *Management Science*. *Management Science*.
- CFTC. (2024). CFTC Market Surveillance Program | CFTC. CFTC Market Surveillance Program. <https://www.cftc.gov/IndustryOversight/MarketSurveillance/CFTCMarketSurveillanceProgram/index.htm>
- Chen, W., Wu, J., Zheng, Z., Chen, C., & Zhou, Y. (2019). Market Manipulation of Bitcoin: Evidence from Mining the Mt. Gox Transaction Network. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 964–972. <https://doi.org/10.1109/INFOCOM.2019.8737364>
- Christin, N. (2013). Traveling the silk road: A measurement analysis of a large anonymous online marketplace. *Proceedings of the 22nd International Conference on World Wide Web*, 213–224. <https://doi.org/10.1145/2488388.2488408>
- Cong, L. W., Li, X., Tang, K., & Yang, Y. (2023). Crypto Wash Trading. *Management Science*, 69(11), 6427–6454. <https://doi.org/10.1287/mnsc.2021.02709>
- Foley, S., Karlsen, J. R., & Putniņš, T. J. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? *The Review of Financial Studies*, 32(5), 1798–1853. <https://doi.org/10.1093/rfs/hhz015>
- Fox, M. B., Glosten, L. R., & Rauterberg, G. V. (2018). Stock market manipulation and its regulation. *Yale J. on Reg.*, 35, 67.
- Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100. [https://doi.org/10.1016/0304-405X\(85\)90044-3](https://doi.org/10.1016/0304-405X(85)90044-3)
- Gandal, N., Hamrick, J., Moore, T., & Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem. *Journal of Monetary Economics*, 95(C), 86–96.
- Kyle, A. S., & Viswanathan, S. (2008). How to Define Illegal Price Manipulation. *American Economic Review*, 98(2), 274–279. <https://doi.org/10.1257/aer.98.2.274>

- Kumar, A., Iqbal, N., Mitra, S. K., Kristoufek, L., & Bouri, E. (2022). Connectedness among major cryptocurrencies in standard times and during the COVID-19 outbreak. *Journal of International Financial Markets, Institutions and Money*, 77, 101523. <https://doi.org/10.1016/j.intfin.2022.101523>
- Schilling, L., & Uhlig, H. (2019). Some simple bitcoin economics. *Journal of Monetary Economics*, 106, 16–26. <https://doi.org/10.1016/j.jmoneco.2019.07.002>
- Schneider, M. (2024). Coming up short: Using Short-seller Reports to Plead Loss Causation in Securities Class Actions. *Columbia Law Review*. <https://columbialawreview.org/content/coming-up-short-using-short-seller-reports-to-plead-loss-causation-in-securities-class-actions/>
- SEC. (2023). SEC.gov | SEC Files 13 Charges Against Binance Entities and Founder Changpeng Zhao. SEC Files 13 Charges Against Binance Entities and Founder Changpeng Zhao. <https://www.sec.gov/news/press-release/2023-101>
- Securities and Exchange Commission v. Sun, No. 1:23-cv-02433 (S.D.N.Y. Mar. 22, 2023).: Hearing on 59. <https://www.courtlistener.com/docket/67071330/securities-and-exchange-commission-v-sun/>
- Securities and Exchange Commission v. Sun, No. 1:23-cv-02433 (S.D.N.Y. Mar. 22, 2023).: Hearing on 62. <https://www.courtlistener.com/docket/67071330/securities-and-exchange-commission-v-sun/>

Tables

Table 1. Variable Definitions. This table presents definitions of the variables used in the paper.

Variables	Definition	Source
$supply_t$	Circulating Supply of BTC since start of BTC blockchain (every day), based on time interval t	blockchain.com
$onchain_t$	Volume of BTC transacted off-chain in all BTC markets, based on time interval t	bitcoincharts.com
$market_t$	Volume of BTC transacted off-chain in all BTC markets, without MtGox, based on time interval t	bitcoincharts.com
Other Variables		
$wash_t$	Total volume of wash trades in MtGox data (every 30 mins)	MtGox Data Dump
$nonwash_t$	Total volume of non-wash trades in MtGox data (every 30 mins)	MtGox Data Dump
$total_t$	Total volume of trades in MtGox data (every 30 mins)	MtGox Data Dump
liq_t	Amihud measure of liquidity in MtGox data (every 30 mins)	MtGox Data Dump
vol_t	Realized volatility in MtGox data, (every 30 mins)	MtGox Data Dump
$eur-close_t$	% change in spot EUR-USD exchange rate	Bloomberg
$eur-liq_t$	% change in spot EUR-USD liquidity, based on Amihud measure	Bloomberg
$eur-vol_t$	% change in spot EUR-USD realized volatility	Bloomberg
$eur-tick_t$	% change in spot EUR-USD tick count	Bloomberg
$gold-close_t$	% change in Gold AU spot price	Bloomberg
$gold-liq_t$	% change in Gold AU liquidity, based on Amihud measure	Bloomberg
$gold-vol_t$	% change in Gold AU realized volatility	Bloomberg
$gold-tick_t$	% change in Gold AU tick count	Bloomberg
$snp-close_t$	% change in SPY (SPDR S&P 500)	Bloomberg
$snp-liq_t$	% change in SPY liquidity, based on Amihud measure	Bloomberg
$snp-vol_t$	% change in SPY realized volatility	Bloomberg
$snp-volume_t$	% change in SPY volume	Bloomberg
$vix-close_t$	% change in VIX (CBOE Volatility Index)	Bloomberg

Table 1. Variable Definitions. This table presents definitions of the variables used in the paper.

Variables	Definition	Source
<i>google</i>	Weekly average interest in Bitcoin on Google	Google API
<i>potday</i>	Binary variable for whether the time period is before or after Pot Day, 20th April 2012	

Table 2. Descriptive Statistics

	Count	Mean	Std Dev	Min	25%	50%	75%	Max
<i>onchain_t</i>	9.268e+06	57.6	2.53e+03	0	0.05	0.715	8.09	5.000e+05
<i>nonwash_t</i>	3.32e+04	1.22e+03	2.5e+03	0	240	516	1.14e+03	5.45e+04
<i>wash_t</i>	3.32e+04	18.3	404	0	0	0	0.06	3.44e+04
<i>total_t</i>	3.32e+04	1.24e+03	2.59e+03	0.02	241	519	1.15e+03	5.89e+04
<i>liq_t</i>	3.32e+04	0.00156	0.0497	0	2.63e-06	8.71e-06	3.96e-05	6.16
<i>vol_t</i>	3.32e+04	834	1.288e+05	0	0.0659	1.17	6.48	2.342e+07
<i>market_t</i>	695	1.1e+04	1.46e+04	1.36e+03	5.02e+03	7.84e+03	1.11e+04	1.693e+05
<i>supply_t</i>	1.09e+03	9.094e+06	2.150e+06	5.180e+06	7.390e+06	9.386e+06	1.099e+07	1.219e+07
<i>gold-close_t</i>	3.33e+04	1.16e+03	763	0	0	1.61e+03	1.69e+03	1.92e+03
<i>gold-tick_t</i>	3.33e+04	2.74e+03	3.02e+03	0	0	1.92e+03	4.41e+03	2.58e+04
<i>gold-vol_t</i>	3.33e+04	0.000906	0.00105	0	0	0.000697	0.00129	0.0171
<i>gold-liq_t</i>	3.33e+04	0.000157	0.00301	0	0	1.1e-05	3.05e-05	0.306
<i>eur-close_t</i>	3.33e+04	0.95	0.593	0	0	1.29	1.32	1.46
<i>eur-tick_t</i>	3.33e+04	1.35e+04	1.15e+04	0	0	1.27e+04	2.19e+04	1.059e+05
<i>eur-vol_t</i>	3.33e+04	0.000654	0.001	0	0	0.000529	0.000889	0.0302
<i>eur-liq_t</i>	3.33e+04	2.49e-05	0.00158	0	0	1.35e-06	3.37e-06	0.226
<i>snp-close_t</i>	3.32e+04	22	44.3	0	0	0	0	137
<i>snp-volum_t</i>	3.32e+04	2.545e+06	6.873e+06	0	0	0	0	1.179e+08
<i>snp-vol_t</i>	3.32e+04	0.000408	0.00126	0	0	0	0	0.0317
<i>snp-liq_t</i>	3.32e+04	2.53e-09	6.77e-09	0	0	0	0	8.45e-08
<i>vix-close_t</i>	3.32e+04	0.00205	0.861	-16.7	0	0	0	25.1
<i>google</i>	100	79.8	9.23	43.1	76.2	81.9	86	94.6

Table 3. How do wash traders decide when to wash-trade? This panel presents coefficients from VAR examining lagged variables of liq_t , vol_t , $nonwash_t$, $total_t$ and their impact on $wash_t$. I use an arbitrary lag of up to $t - 4$. The coefficients here do not refer directly to the variables in Table 1 but the weighted feature importance of each of them in the model. The higher the value, the more important the variable is. A placebo has been placed and as a guideline, any variable with importance lower than the placebo is not important as it shows that a randomly generated placebo has higher importance than it. Time period is from 26th June 2011 to 20th May 2013. The ranking of each variable in the corresponding model are shown in parentheses. Variable definitions are detailed in Table 1.

Feature	(1)	(2)	(3)	(4)	(5)	(6)
Importance	CART	Random For- est	AdaBoost	XGBoost	GRU	LSTM
liq_{t-1}	6.28e+07 (7)	0 (11.5)	0.0147 (16)	0.0165 (14)	2.95e-06 (8)	1.81e-08 (9)
liq_{t-2}	8.46e+07 (4)	0 (11.5)	0.0271 (9)	0.0113 (16)	-1.11e-05 (14)	1.31e-08 (10)
liq_{t-3}	7.57e+07 (6)	0 (11.5)	0.0207 (13)	0.0267 (12)	2.91e-06 (9)	4.46e-09 (11)
liq_{t-4}	0 (15.5)	0 (11.5)	0.0217 (12)	0.0165 (15)	2.04e-06 (11)	-6.12e-10 (12)
vol_{t-1}	1.45e+08 (2)	0 (11.5)	0.041 (7)	0.0971 (4)	4.89e-06 (5)	0.00016 (1)
vol_{t-2}	0 (15.5)	0 (11.5)	0.022 (11)	0.034 (10)	4.4e-06 (6)	8.44e-05 (2)
vol_{t-3}	2.5e+07 (11)	0.473 (1)	0.13 (4)	0.0325 (11)	3.06e-06 (7)	5.73e-05 (3)
vol_{t-4}	4.65e+06 (12)	0 (11.5)	0.0205 (14)	0.0451 (6)	2.75e-06 (10)	2.86e-05 (4)
$nonwash_{t-1}$	4.04e+07 (8)	0.0679 (3)	0.122 (5)	0.118 (3)	-2.23e-05 (15)	1.06e-07 (8)
$nonwash_{t-2}$	2.88e+07 (9)	0.00613 (4)	0.028 (8)	0.0195 (13)	9.61e-06 (4)	2.94e-07 (7)
$nonwash_{t-3}$	8.22e+07 (5)	0 (11.5)	0.0239 (10)	0.0349 (9)	-7.46e-06 (13)	9.67e-06 (5)
$nonwash_{t-4}$	0 (15.5)	0 (11.5)	0.0161 (15)	0.0113 (17)	-2.06e-06 (12)	6.35e-06 (6)
$total_{t-1}$	2.57e+07 (10)	0.00542 (5)	0.151 (2)	0.119 (2)	0.000139 (1)	-4.77e-07 (13)
$total_{t-2}$	1.46e+08 (1)	0 (11.5)	0.156 (1)	0.24 (1)	7.35e-05 (2)	-9.04e-07 (14)
$total_{t-3}$	1.32e+08	0.448	0.143	0.0918	-0.000156	-1.32e-05

	(3)	(2)	(3)	(5)	(16)	(17)
$total_{t-4}$	0	0	0.0136	0.0431	2.88e-05	-1.01e-05
	(15.5)	(11.5)	(17)	(8)	(3)	(16)
placebo	3.79e+05	0	0.0481	0.0432	-0.000254	-3.71e-06
	(13)	(11.5)	(6)	(7)	(17)	(15)
n	3.32e+04	3.32e+04	3.32e+04	3.32e+04	3.32e+04	3.32e+04

Table 4 - Panel A. How do wash traders decide when to wash-trade? This panel presents coefficients from a **Johansen test** examining variables of liq_t , vol_t , $nonwash_t$, $total_t$ and $wash_t$. Time period is from 26th June 2011 to 20th May 2013. Variable definitions are detailed in Table 1.

Hypothesized of Cointegrating Equations	No.	Trace Statistic	0.95 Critical Value	0.95 Critical Value Max-Eigen	Pass/Fail (95% Sig- nificance)
At most 0		4.4e+04	60	30	Pass
At most 1		3e+04	40	24	Pass
At most 2		1.6e+04	24	18	Pass
At most 3		5.3e+03	12	11	Pass
At most 4		-2.1e+02	4.1	4.1	Fail

Panel B. How do wash traders decide when to wash-trade? This panel presents coefficients from a **Granger Causality Panel** examining variables of liq_t , vol_t , $nonwash_t$, $total_t$ and $wash_t$. Time period is from 26th June 2011 to 20th May 2013. Variable definitions are detailed in Table 1.

Independent Variable	Dependent Variable	Lag	F-test p-value	Result
$wash_t$	$nonwash_t$	1	0.018	Pass
$wash_t$	$nonwash_t$	2	5.6e-09	Pass
$wash_t$	$total_t$	1	3.4e-32	Pass
$wash_t$	$total_t$	2	1.5e-47	Pass
$wash_t$	liq_t	1	0.91	Fail
$wash_t$	liq_t	2	0.97	Fail
$wash_t$	vol_t	1	0.97	Fail
$wash_t$	vol_t	2	4.3e-107	Pass

Panel C. How do wash traders decide when to wash-trade? This panel presents coefficients from a VAR following the model in Table 2 with up to $t - 4$ lags and its **Impulse Response Function** examining variables of liq_t , vol_t , $nonwash_t$, $total_t$ and $wash_t$. I use arbitrary forward time steps of up to $t + 10$. Time period is from 26th June 2011 to 20th May 2013. Variable definitions are detailed in Table 1.

Lag	% IRF of $nonwash_t$ to $wash_t$	% IRF of $wash_t$ to $nonwash_t$	% IRF of $total_t$ to $wash_t$	% IRF of $wash_t$ to $total_t$
1	1.3e+04	-inf	6.7e+03	inf
2	1.7e+03	-9e+02	1e+03	93
3	-1.3e+05	2.8e+05	-2.1e+05	-1.2e+06

4	1.6e+03)	-4.3e+02	1.7e+03	2.5e+02
5	-1.3e+05	6.9e+05	-1.2e+05	-6.4e+05
6	2e+03	-60	2e+03	4.1e+02
7	-1.1e+05	-5.7e+06	-1.1e+05	-4.5e+05
8	2.3e+03	-59	1.8e+03	5.6e+02
9	-9.4e+04	-5.5e+06	-1.2e+05	-3.4e+05
10	2.3e+03	1.2e+02	2.3e+03	7.1e+02

Table 5 - Panel A. Do wash trades correlate with on-chain transactions? This panel presents coefficients from OLS regressions examining changes in non-wash trade volume to on-chain transaction volume. $nonwash_t$ and $onchain_t$ second-level data and accumulated to $t = 30$ minutes. Quartiles refer to whether the wash-trades are in the highest volume arranged to the lowest volume, with Quartile 4 being the highest trade volume and Quartile 1 being the lowest. Time period is from 16th September 2011 to 1st June 2012. p-values are shown in parentheses. Variable definitions are detailed in Table 1. *, **, *** represent significance at the 10%, 5% and 1% level.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
$nonwash_t$	0.447	-0.186	2.2***	-0.199
	0.475	0.68	5.27e-05	0.424
constant	2.4e+04***	2.16e+04***	2.3e+04***	1.95e+04***
	8.06e-120	7.12e-89	9.49e-46	9.07e-58
adj r^2	-0.000166	-0.000283	0.0052	-0.000122
n	2.96e+03	2.94e+03	2.94e+03	2.96e+03

Panel B. Do wash trades correlate with on-chain transactions? This panel presents coefficients from a two-step Engel-Granger cointegration examining changes in non-wash trade volume to on-chain transaction volume. $nonwash_t$ and $onchain_t$ seconds-level data and accumulated to $t = 30$ minutes. Quartiles refer to whether the wash-trades are in the highest volume arranged to the lowest volume, with Quartile 4 being the highest trade volume and Quartile 1 being the lowest. Time period is from 16th September 2011 to 1st June 2012.

Quartile	Cointegration p-value
Quartile 1	0.153
Quartile 2	0.0937
Quartile 3	0.0834
Quartile 4	0.0511*
n	2.94e+03

Panel C. Do wash trades correlate with other exchange transactions? This panel presents coefficients from OLS regressions examining changes in non-wash trade volume to market transaction volume. $nonwash_t$ and $market_t$ second-level data and accumulated to $t = 30$ minutes. Quartiles refer to whether the wash-trades are in the highest volume arranged to the lowest volume, with Quartile 4 being the highest trade volume and Quartile 1 being the lowest. Time period is from 26th June 2011 to 20th May 2013. p-values are shown in parentheses. Variable definitions are detailed in Table 1. *, **, *** represent significance at the 10%, 5% and 1% level.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
30-min-non-wash	0.0500***	0.1205***	0.1692***	0.2269***
	(0.0003)	(0.0000)	(0.0000)	(0.0000)

constant	5198.7580*** (0.0000)	2666.5921*** (0.0000)	1994.8933 (0.1138)	−4762.9151** (0.0209)
adj r^2	0.0685	0.3851	0.3397	0.5211
n	174	174	173	174

Table 6 - Panel A. Do wash trades correlate with other asset movements? This panel presents coefficients from IRFs derived from VAR examining changes in wash trade volume $wash_t$ to other variables based on EURO-USD spot, Gold spot, SNP500 index and VIX. The prefixes represent each of the 4 assets respectively, and the suffixes - liq_t represents % change in liquidity, vol_t represents % change in realized volatility, $close_t$ represents % change in price, $tick_t$ represents % change in tick count and $volume_t$ represents % change in volume. The variables are seconds-level and accumulated to $t = 30$ minutes. Time period is from 26th June 2011 to 20th May 2013. Variable definitions are detailed in Table 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lag	% IRF of $wash_t$ to $eur-vol_t$	% IRF of $wash_t$ to $eur-liq_t$	% IRF of $wash_t$ to $eur-tick_t$	% IRF of $wash_t$ to $eur-close_t$	% IRF of $wash_t$ to $gold-vol_t$	% IRF of $wash_t$ to $gold-liq_t$	% IRF of $wash_t$ to $gold-tick_t$	% IRF of $wash_t$ to $gold-close_t$
								t
1	-199	-133	-260	-157	-212	-32.3	-171	-247
2	-105	-546	-3.88	-109	-123	7.88	126	-170
3	-254	-261	-199	3.11e+03	-94	-163	-247	-143
4	-47.8	-180	-30.2	-154	3.2e+03	2.92e+03	-173	-120
5	-20.2	-162	-6.83	-155	-157	-61.3	-106	-683
6	-221	-46.9	-113	-467	-73	-410	-1.02e+03	-112
7	-186	-159	-204	33.3	-896	-105	-338	-609
8	-119	64.8	-766	-217	-98.4	-227	-196	-361
9	4.98e+03	-276	155	-16.4	-8e+03	-60.4	-43.8	-144
10	-85.5	-185	-158	-226	-153	-120	-133	-241
n	3.33e+04	3.33e+04	3.33e+04	3.33e+04	3.33e+04	3.33e+04	3.33e+04	3.33e+04

Panel B. Do wash trades correlate with other asset movements? This panel presents coefficients from IRFs derived from VAR examining changes in wash trade volume $wash_t$ to other variables based on EURO-USD spot, Gold spot, SNP500 index and VIX. The prefixes represent each of the 4 assets respectively, and the suffixes - liq_t represents % change in liquidity, vol_t represents % change in realized volatility, $close_t$ represents % change in price, $tick_t$ represents % change in tick count and $volume_t$ represents % change in volume. The variables are seconds-level and accumulated to $t = 30$ minutes. Time period is from 26th June 2011 to 20th May 2013. Variable definitions are detailed in Table 1.

	(1)	(2)	(3)	(4)	(5)
Lag	% IRF of $wash_t$ to $snp-vol_t$	% IRF of $wash_t$ to $snp-liq_t$	% IRF of $wash_t$ to $snp-volume_t$	% IRF of $wash_t$ to $snp-close_t$	% IRF of $wash_t$ to $vix-close_t$
1	-345	-225	-217	-176	-428

2	-90.1	-72.8	-116	32.5	-105
3	50.3	-191	-207	-317	-4.77e+03
4	-58.1	-210	-114	-251	-167
5	-217	-14.4	-918	-135	-123
6	-754	-264	-21.4	-99.7	-24.9
7	-189	-152	-647	2.97e+04	-1.02e+03
8	-53.2	-12.7	-149	-76.7	-127
9	133	-741	-233	-375	-257
10	-143	-177	-176	-197	-136
n	3.32e+04	3.32e+04	3.32e+04	3.32e+04	3.32e+04

Table 7. Does media attention affect the effect of wash trades on market activity? This panel presents coefficients from IRFs derived from VAR examining changes in wash trade volume $wash_t$, non-wash trade volume $nonwash_t$, total trade volume $total_t$, liquidity liq_t , volatility vol_t . The variables are seconds-level and accumulated to $t = 30$ minutes. They are then accumulated to weekly intervals. Weeks that do not have stationary data trends are dropped, thus, there are $n = 83$ points. *google* is a weekly datapoint and the median media popularity is based on whether the *google* for that week is higher or lower than the median. Columns (1)-(4) are for those below median media popularity, columns (5)-(8) are for those above median media popularity. Time period is from 26th June 2011 to 20th May 2013. Future time periods of $t + 1$ to $t + 10$ are shown and the sum is shown in the 11th row. Variable definitions are detailed in Table 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% IRF of $total_t$ to $wash_t$	% IRF of $nonwash_t$ to $wash_t$	% IRF of $wash_t$ to $total_t$	% IRF of $wash_t$ to $nonwash_t$	% IRF of $total_t$ to $wash_t$	% IRF of $nonwash_t$ to $wash_t$	% IRF of $wash_t$ to $total_t$	% IRF of $wash_t$ to $nonwash_t$
	Below median media popularity				Above median media popularity			
1	-2.550e+05	-2.464e+05			1.268e+05	2.566e+05		
2	3.31e+03	4.16e+03	55.2	-9.18e+03	2.79e+04	-193	-647	946
3	3.324e+05	1.112e+06	1.792e+06	-2.96e+04	-1.736e+05	3.24e+04	2.183e+05	-1.43e+04
4	-148	156	259	-9.05e+03	-3.09e+04	64.9	-586	-1.21e+04
5	2.845e+05	3.006e+05	-1.009e+06	4.72e+03	-5.81e+04	-4.76e+04	1.440e+05	-8.36e+03
6	7.44e+03	8.65e+03	-90.8	-689	-692	488	-1.13e+03	2.22e+03
7	2.501e+05	3.74e+03	-4.23e+04	4.38e+03	6.38e+03	9.4e+04	3.39e+04	-8.66e+03
8	1.32e+03	1.55e+03	-129	-6.13e+03	343	-668	-600	-3.57e+03
9	-2.584e+05	-8.84e+04	-7.67e+04	2.45e+03	4.08e+04	2.4e+04	7.8e+03	-5.8e+03
10	889	235	-577	-2.91e+03	-968	4.44e+03	-1.19e+03	2.26e+03
sum	3.664e+05	1.096e+06	6.630e+05	-4.6e+04	-6.2e+04	3.634e+05	3.999e+05	-4.73e+04
n	83	83	83	83	83	83	83	83

Table 8. Does media attention affect the effect of wash trades on market activity? This panel presents coefficients from IRFs derived from VAR examining changes in wash trade volume $wash_t$, non-wash trade volume $nonwash_t$, total trade volume $total_t$, liquidity liq_t , volatility vol_t . The variables are seconds-level and accumulated to $t = 30$ minutes. They are then accumulated to weekly intervals. Weeks that do not have stationary data trends are dropped, thus, there are $n = 83$ points. *google* is weekly data and the median media popularity is based on whether the *google* for that week is higher or lower than the median. Columns (1)-(4) are for those below median media popularity, columns (5)-(8) are for those above median media popularity. Time period is from 26th June 2011 to 20th May 2013. Future time periods of $t + 1$ to $t + 10$ are shown and the sum is shown in the 11th row. Variable definitions are detailed in Table 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% IRF of $total_t$ to $wash_t$	% IRF of $nonwash_t$ to $wash_t$	% IRF of $wash_t$ to $total_t$	% IRF of $wash_t$ to $nonwash_t$	% IRF of $total_t$ to $wash_t$	% IRF of $nonwash_t$ to $wash_t$	% IRF of $wash_t$ to $total_t$	% IRF of $wash_t$ to $nonwash_t$
	—2 weeks before			Pot Day	—2 weeks after			Pot Day
1	6.48e+04	7.79e+04			1.3e+04	1.32e+04		
2	-212	-197	-219	-143	-1.13e+03	-1.21e+03	-194	-9.34e+03
3	2.03e+04	4.24e+03	-6.98e+03	-2.7e+04	1.13e+04	1.14e+04	1.090e+05	874
4	-1.84e+03	-8.97e+03	4.03e+03	2.9e+03	-1.12e+03	-1.33e+03	-289	-1.26e+04
5	-957	-760	152	501	9.97e+03	1.05e+04	6.43e+04	795
6	1.01e+04	-7.46e+03	1.44e+04	6.27e+03	-1.3e+03	-1.43e+03	-365	-1.57e+04
7	-780	458	650	865	8.79e+03	9.66e+03	5.23e+04	698
8	-1.22e+03	1.24e+03	3.74e+03	3.52e+03	-1.53e+03	-1.43e+03	-461	-1.57e+04
9	-658	8.79e+03	2.92e+03	2.67e+03	7.89e+03	8.63e+03	3.4e+04	564
10	-125	147	1.84e+03	1.84e+03	-1.65e+03	-1.58e+03	-589	-1.9e+04
<i>sum</i>	8.94e+04	7.54e+04	2.05e+04	-8.61e+03	4.42e+04	4.63e+04	2.577e+05	-6.94e+04
<i>n</i>	672	672	672	672	670	670	670	670

Appendix

A1. Data Processing of Mt.Gox Dataset

I first take the original files from the leaked datadump and take each half of a transaction with the same transaction ID into one row. I then de-duplicate them by identifying unique rows based on unique combined primary keys of ['Buyer', 'Seller', 'Bitcoins', 'Money', 'Date']. There are many duplicate transactions such as these, which have the same "Bitcoins" and "Money" amount with the same timestamp.

User_Id	Trade_Id	Date	Japan	Currency	Bitcoins	Money
176214	2650732688407216	2011-12-31 21:19:04	NJP	USD	6.00	28.12392
176215	2650732688897936	2011-12-31 21:19:04	NJP	USD	6.00	28.12392
176216	2650732689364758	2011-12-31 21:19:04	NJP	USD	6.00	28.12392

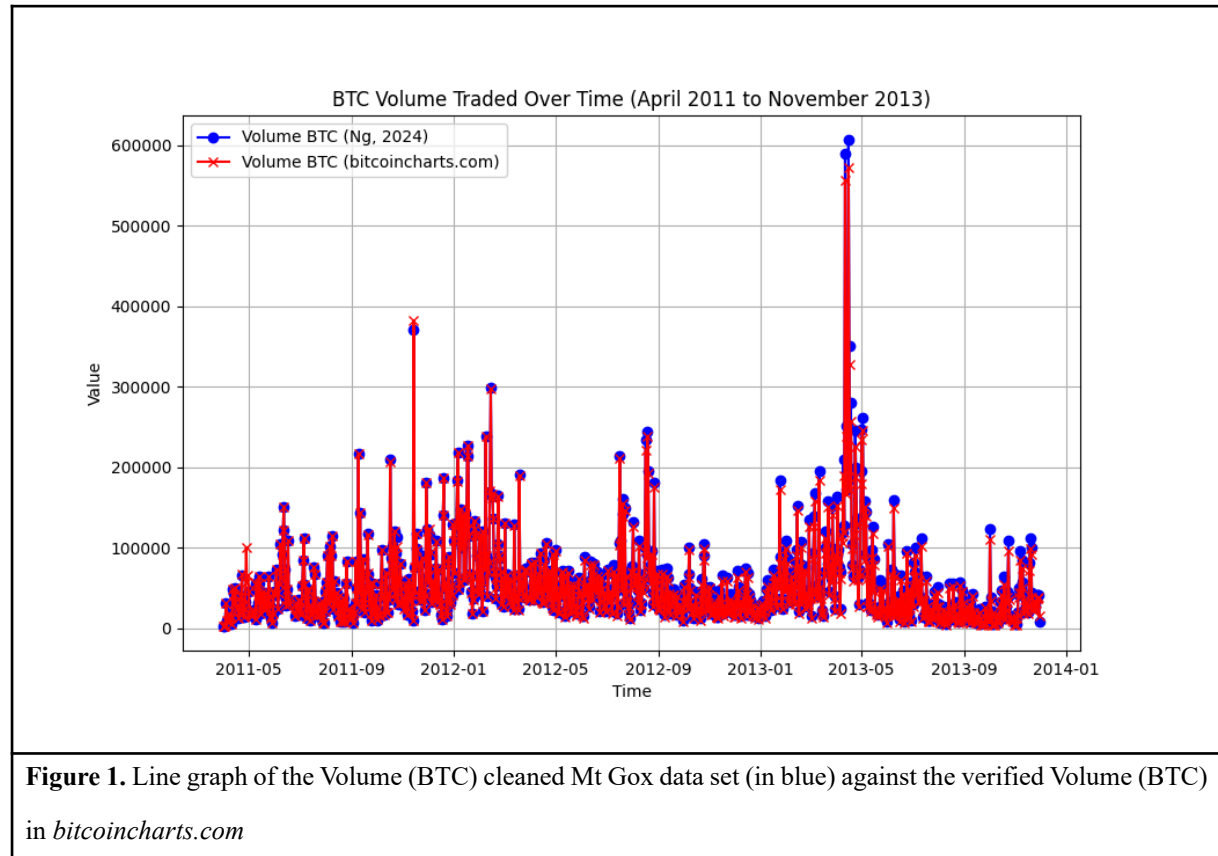
I start with 16748680 of unpaired transactions. I disregard files in 2013 which are broken down into days such as [2013-03-12_mtgox_japan.xlsx] as these transactions are captured in the monthly files. There are 32 files in total. 9 months in 2011, 12 months in 2012 and 11 months in 2013. I pair the transactions and then de-duplicate.

Similar to Aloosh and Li (2023), I identify that wash trades take place from June 26th 2011 to May 20, 2013. This is a different time period from Gandal et al. (2018), where they focus from February 2013 to November 2013 for two suspected trading bots - Markus and Willy.

I arrive at 7,741,721 de-duplicated transactions for the final dataset.

A2. Verification of Mt Gox Data using External Data Source

After the de-duplication process and combining of buy-sell transactions, I checked the Volume of BTC traded each day against a verified data source - *bitcoincharts.com*. The blue points



A3. On-chain Collection of BTC Transactions

The on-chain activity is collected from <https://blockchain.info>. I collect the transactions every 100 blocks, and save it as a JSON format. After this, I write an algorithm to collect the transaction volume of BTC. The data snippet shows that the following fields are collected from on-chain [timestamp, transaction_id, address, type]

A	B	C	D	E
Timestamp	Transaction ID	Address	Type	Amount (BTC)
2011-10-06 11:55:30	42101c19a153051c9f0e2afb107561afe724f47ae8fce00383a8f62eb9d801cf	N/A	input	0
2011-10-06 11:55:30	42101c19a153051c9f0e2afb107561afe724f47ae8fce00383a8f62eb9d801cf	15wNVu5eEynhJmitToi	output	50.0385001
2011-10-06 11:55:30	d4afede15f815de5de9ed8c86a50b8907dcb2a9b01530eb1150367115a8154e7	1M9Pfd6R8ecm4ksBoH	input	24
2011-10-06 11:55:30	d4afede15f815de5de9ed8c86a50b8907dcb2a9b01530eb1150367115a8154e7	198jCSHAquMDMK8Czy	output	4
2011-10-06 11:55:30	d4afede15f815de5de9ed8c86a50b8907dcb2a9b01530eb1150367115a8154e7	1JpmMhCUUSUB1kcr1	output	20
2011-10-06 11:55:30	a60818f3f85eeba2e618995aa4b16e241904a084866c1bd20d09f44e1b9de55d	1K8mkyFF79eJmdooZiV	input	144.16
2011-10-06 11:55:30	a60818f3f85eeba2e618995aa4b16e241904a084866c1bd20d09f44e1b9de55d	1Mxy7THEHTA89vU1Zf	output	110.18
2011-10-06 11:55:30	a60818f3f85eeba2e618995aa4b16e241904a084866c1bd20d09f44e1b9de55d	1BSPSBQ1SLftPQzwQ6	output	33.98
2011-10-06 11:55:30	c8b920175f1149f4584ca8adfe1cc91acb55f17f9a0da51a2546ba5de1b91645	1KGqhA6DeFv7u2UWp	input	49.46108639
2011-10-06 11:55:30	c8b920175f1149f4584ca8adfe1cc91acb55f17f9a0da51a2546ba5de1b91645	18zf1H6ZnWNdxXpSt5C	output	45.00388716
2011-10-06 11:55:30	c8b920175f1149f4584ca8adfe1cc91acb55f17f9a0da51a2546ba5de1b91645	12hgsAQW1hR5UQAG	output	4.45719923
2011-10-06 11:55:30	8f9052442d843e18a2ffaec8cde95f4f14649a6808007331f04f5c4c197e9021	1Dnbqj1y3LybLoeG1De	input	2.6
2011-10-06 11:55:30	8f9052442d843e18a2ffaec8cde95f4f14649a6808007331f04f5c4c197e9021	1MgonWT1bS3jFTYDXV	input	0.0102
2011-10-06 11:55:30	8f9052442d843e18a2ffaec8cde95f4f14649a6808007331f04f5c4c197e9021	15bnDJUUmsoPukr1Xz	output	0.0131
2011-10-06 11:55:30	8f9052442d843e18a2ffaec8cde95f4f14649a6808007331f04f5c4c197e9021	1P61fSnjEZP4t75gn7Pb	output	2.5971
2011-10-06 11:55:30	9337d4d2481f6da6e4379a44c1622c416da26225dc914ceeb707ee896fc5935d	1LEp7L4fyvx8RtQXJTvz	input	4.5286
2011-10-06 11:55:30	9337d4d2481f6da6e4379a44c1622c416da26225dc914ceeb707ee896fc5935d	16GcStMf9qBvLSugeRS	input	0.0202724
2011-10-06 11:55:30	9337d4d2481f6da6e4379a44c1622c416da26225dc914ceeb707ee896fc5935d	1McqcKpos3ga6tiHG5q	output	0.0141784
2011-10-06 11:55:30	9337d4d2481f6da6e4379a44c1622c416da26225dc914ceeb707ee896fc5935d	18EVYrWCJLDHf6AQ5C	output	4.534694
2011-10-06 11:55:30	8f502e6be62b93025e3a30844f12603b6b566f61eb5a0aa8679da81309ce7f96	1HGVQM2StPjne4auKz	input	20
2011-10-06 11:55:30	8f502e6be62b93025e3a30844f12603b6b566f61eb5a0aa8679da81309ce7f96	19qNekBmDpo3n9AuK	output	20
2011-10-06 11:55:30	2143131c99baf6b5f20cc206cf1c1b1dfcc2b64640c8fc7c7a1edfb7870cae53	1MrBziDLAokqnBI41be	input	15
2011-10-06 11:55:30	2143131c99baf6b5f20cc206cf1c1b1dfcc2b64640c8fc7c7a1edfb7870cae53	1H5EMaBoM5WDWZo7	output	7.9
2011-10-06 11:55:30	2143131c99baf6b5f20cc206cf1c1b1dfcc2b64640c8fc7c7a1edfb7870cae53	1BFRM1tFgijPejmwHYy	output	7.1
2011-10-06 11:55:30	8632a8952166128a0b6ac2b65d2dd845190216e124cd9da5fe30ffbed4fd1fc1	1FcAEdtWo9S5AevPf	input	169.989874
2011-10-06 11:55:30	8632a8952166128a0b6ac2b65d2dd845190216e124cd9da5fe30ffbed4fd1fc1	1A6fdLYdw6tVskigAgje	output	119.9869359
2011-10-06 11:55:30	8632a8952166128a0b6ac2b65d2dd845190216e124cd9da5fe30ffbed4fd1fc1	1LjpkLYTyCpJBbdsiKavj	output	50.002938
2011-10-06 11:55:30	ab329cb14e36e2f4b1001b7e2972ac8c9fbfe71d453995097fddc274dd8a529d8	1EsN7CAHzBz5S1VMDM	input	74.521
2011-10-06 11:55:30	ab329cb14e36e2f4b1001b7e2972ac8c9fbfe71d453995097fddc274dd8a529d8	1B6M8t5fqDpDmLmLn	output	73.5805
2011-10-06 11:55:30	ab329cb14e36e2f4b1001b7e2972ac8c9fbfe71d453995097fddc274dd8a529d8	14zWrUshFQ2aKWP7xk	output	0.94
2011-10-06 11:55:30	19e5f50b06cec6546a9d4cd02c5f9b49197434ad11c092c0c1ca893b99a14ee6	1AUESAHGGbCzUphvSc	input	0.01024119
2011-10-06 11:55:30	19e5f50b06cec6546a9d4cd02c5f9b49197434ad11c092c0c1ca893b99a14ee6	1DxWx9XzPfAeR3ncKvM	input	0.01005458
2011-10-06 11:55:30	19e5f50b06cec6546a9d4cd02c5f9b49197434ad11c092c0c1ca893b99a14ee6	1LrsnemadPEhkPrzcFvF	input	4
2011-10-06 11:55:30	19e5f50b06cec6546a9d4cd02c5f9b49197434ad11c092c0c1ca893b99a14ee6	129SwTUDBKDFcEVAa7	output	0.01029577
2011-10-06 11:55:30	19e5f50b06cec6546a9d4cd02c5f9b49197434ad11c092c0c1ca893b99a14ee6	19wafKAqJtyCCPJhvuLf	output	4
2011-10-06 11:55:30	122905a30bdc13357d10d17b6dd422e1b6ba694ba00ffdfca652c3670f13cfbe	19eTFnQVnDnAVMJHt	input	43.3715

Figure 2. Data snippet of the on-chain transactions for BTC

A4. Comparison of Mt Gox Volume (BTC) against Volume (BTC) of 30 other exchanges during the time period from June 2011 to May 2013

I manually retrieved the Volume (BTC) traded on 30 BTC-USD exchanges that registered with bitcoincharts.com during this time period for their data to be collected. The list of exchanges is ['bitbox', 'bitcoin24', 'bitcoin2cash', 'bitcoin7', 'bitcurex', 'bitfloor', 'bitkonan', 'bitmarket.eu', 'bitme', 'bitstamp', 'btce', 'btcex', 'btctree', 'campbx', 'crypto-x-change', 'exchange-bitcoin', 'fbtc', 'imcex', 'intersango', 'itbit', 'justcoin', 'jutcoin', 'libertybit', 'local-bitcoin', 'mtgox', 'ruxum', 'therock-trading', 'tradehill1', 'tradehill2', 'vircurex', 'weexchange']. Some exchanges were relatively short-lived as they were shut down during the time period. The average % of Mt Gox's volume of the entire market was about 83.37%.

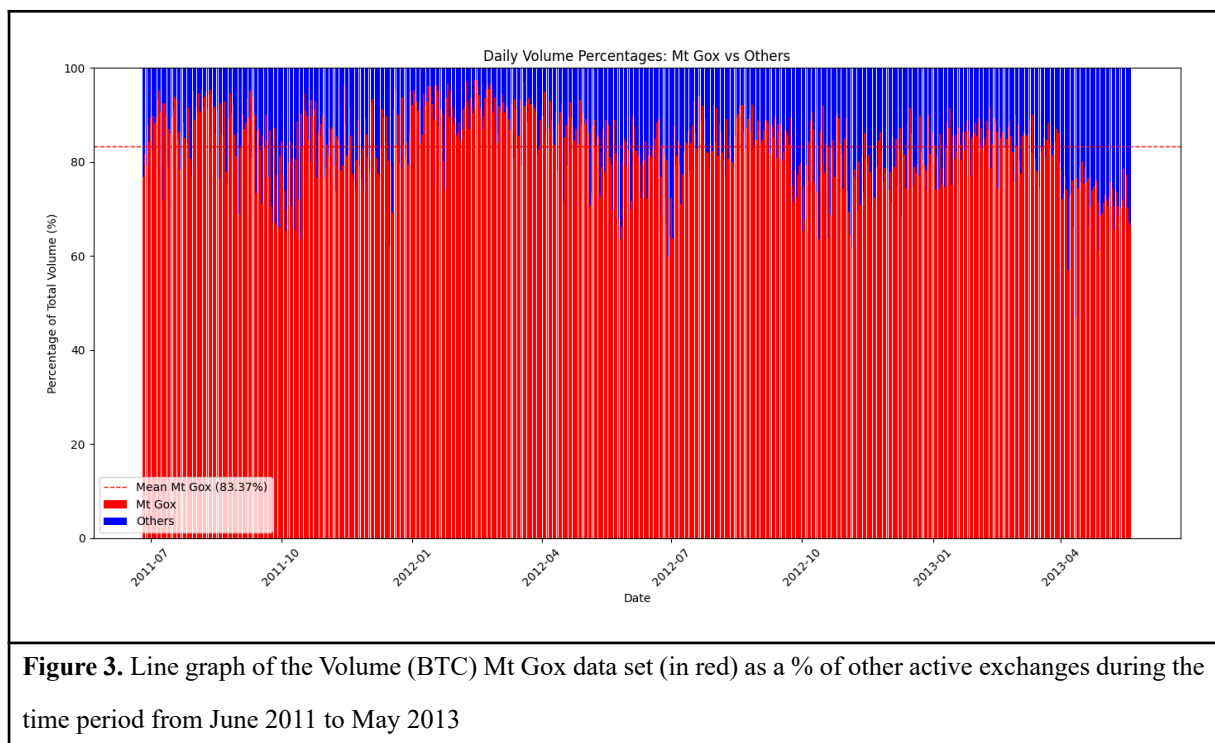


Figure 3. Line graph of the Volume (BTC) Mt Gox data set (in red) as a % of other active exchanges during the time period from June 2011 to May 2013

A5. Distribution of wash-trade volume

I compute the wash-trade volume each day and to ensure the robustness of the cointegration and regression Section 3.2, I plot the distribution of the wash-trade volume by splitting it into four quartiles.

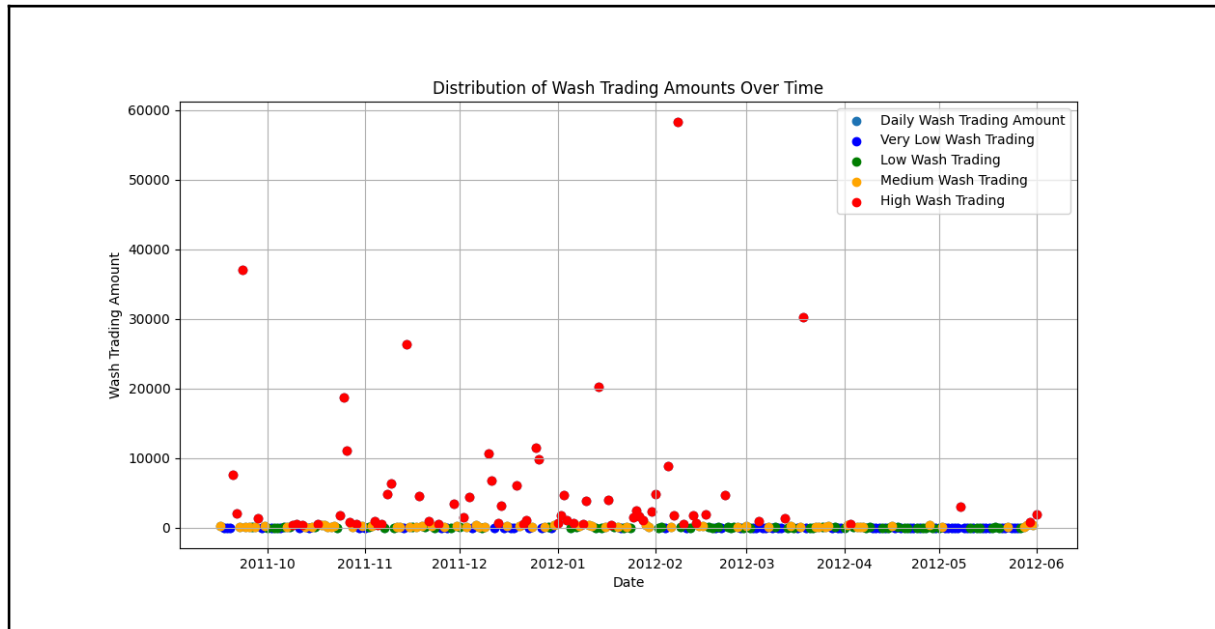


Figure 4A. Distribution of the wash-trade volume across the time period from September 2011 to June 2012. The points are evenly distributed across the entire time frame.

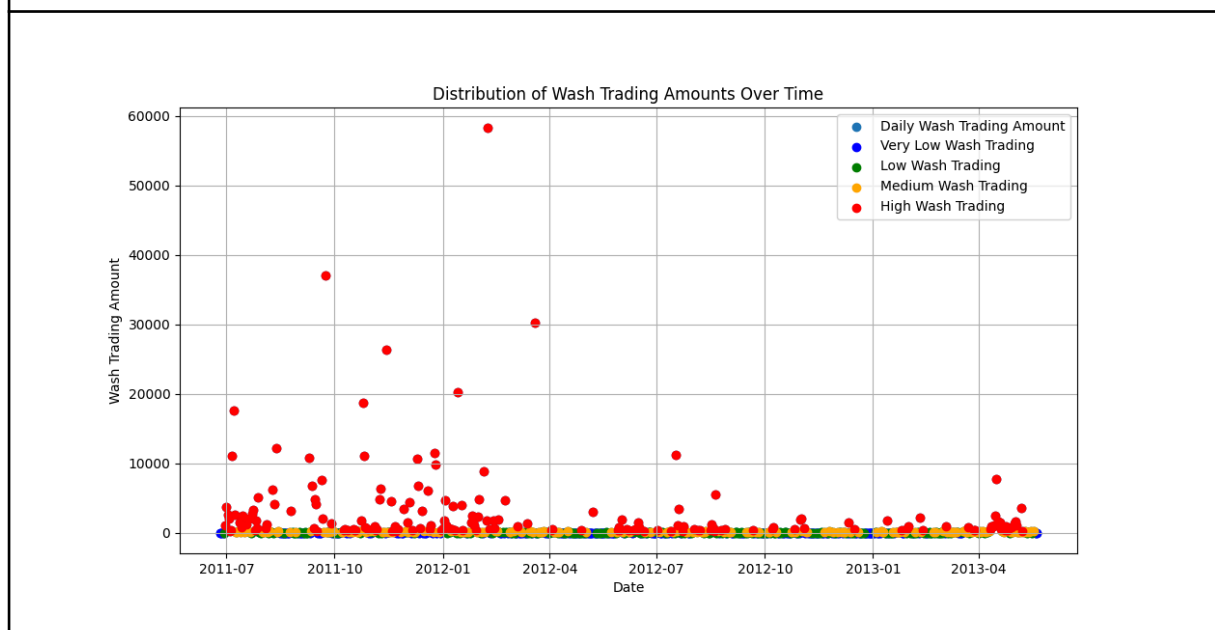
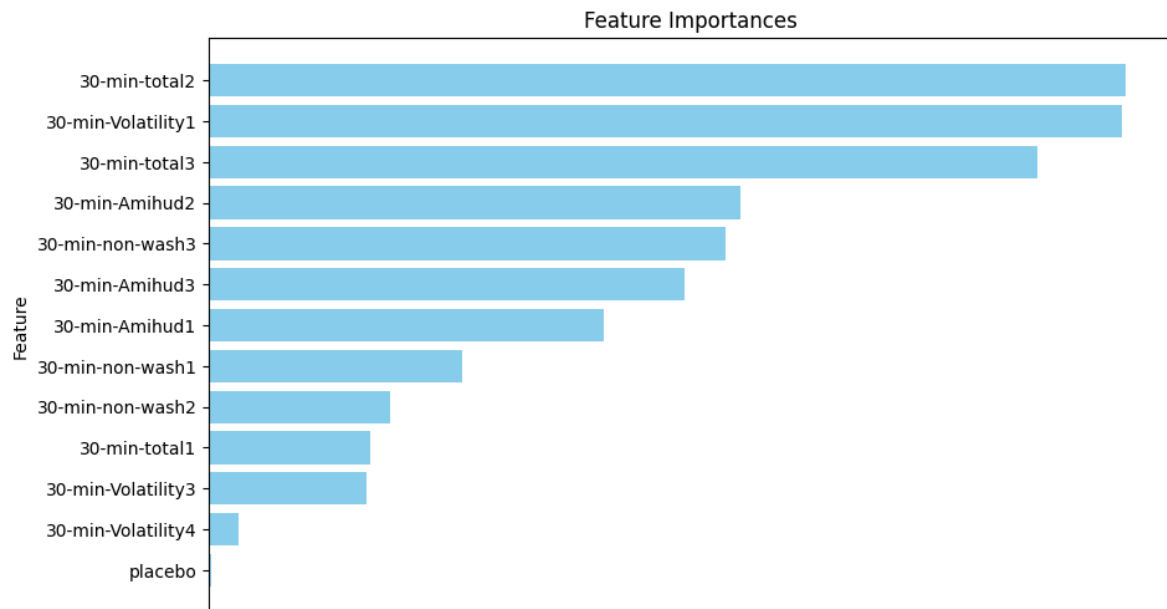


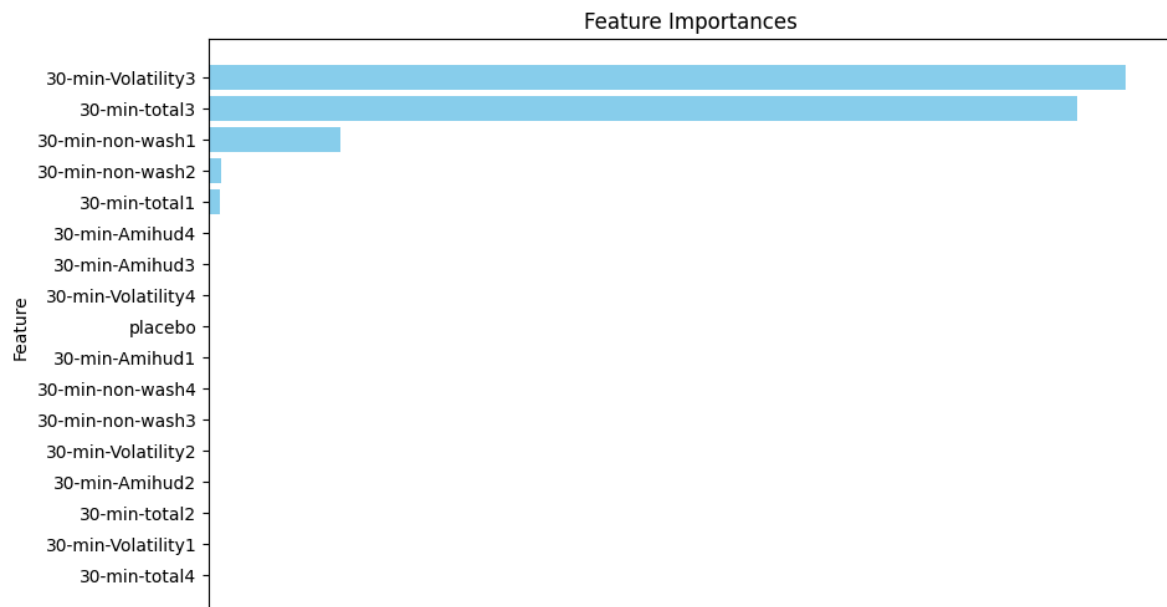
Figure 4B. Distribution of the wash-trade volume across the time period from June 2011 to May 2013. The points are evenly distributed across the entire time frame.

A6. Distribution of wash-trade volume

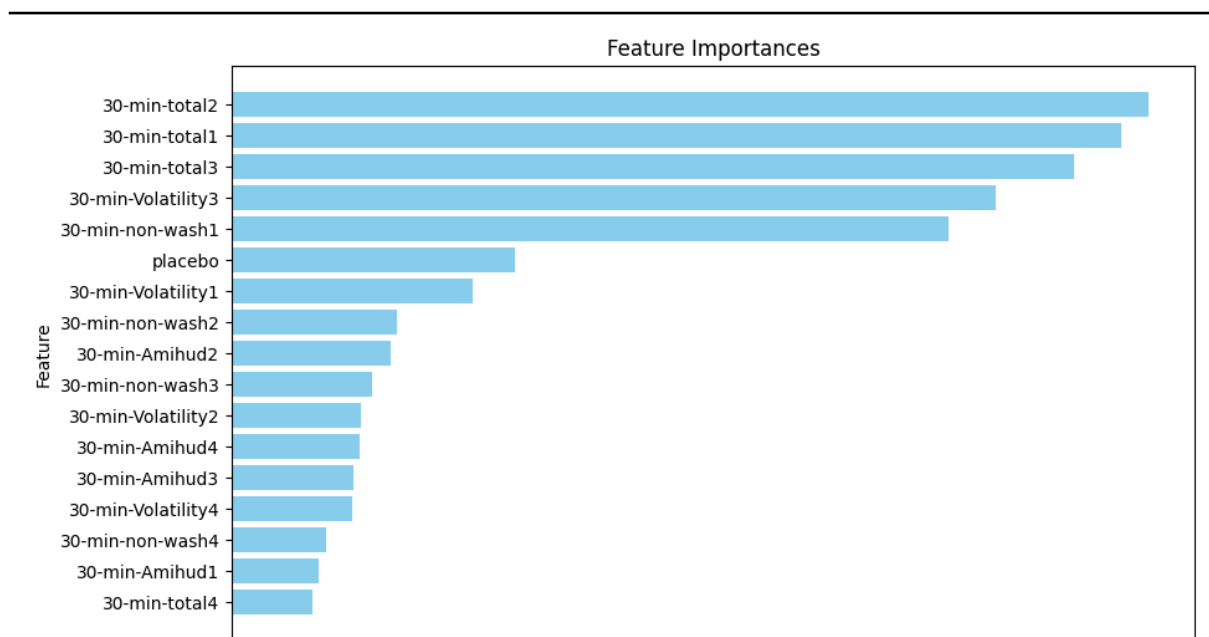
Figure 5. How do wash traders decide when to wash-trade? This panel presents a visualization of the results in Table 2. Time period is from 26th June 2011 to 20th May 2013. The ranking of each variable in the corresponding model are shown in parentheses. Variable definitions are detailed in Table 1.



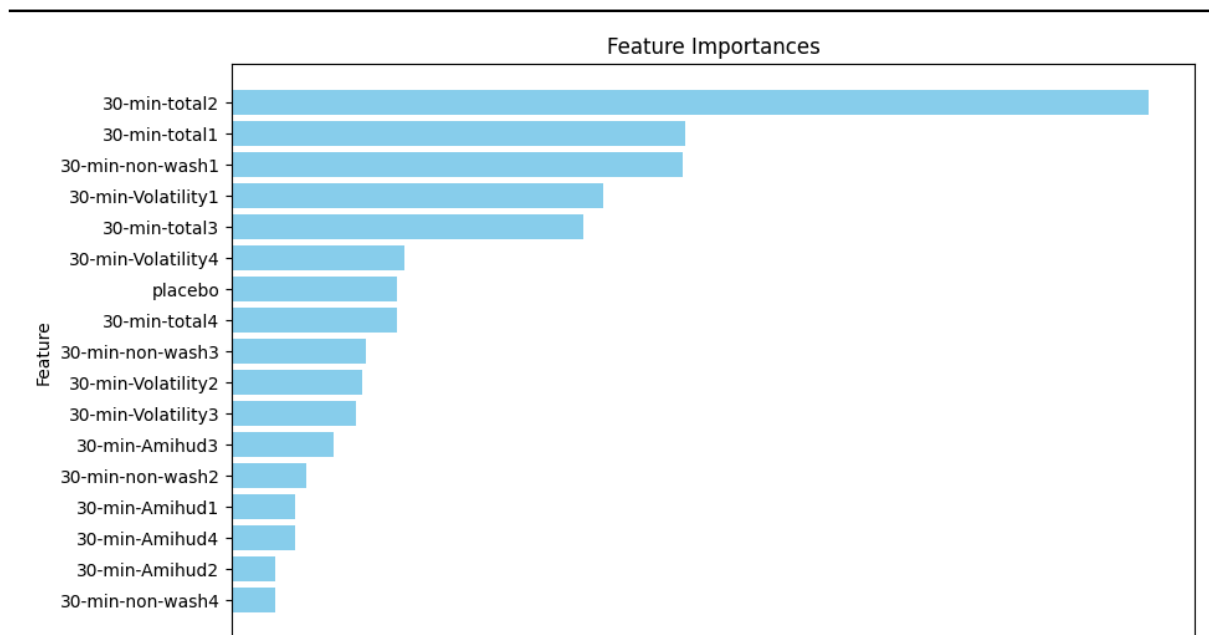
Xgboost Model



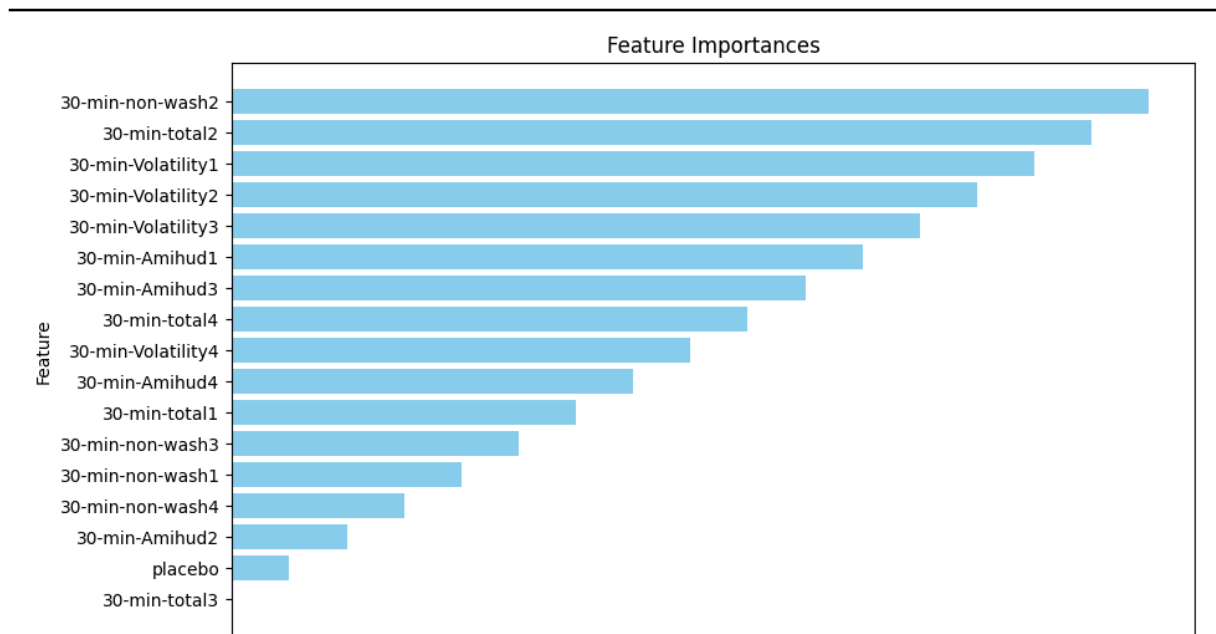
Standard CART



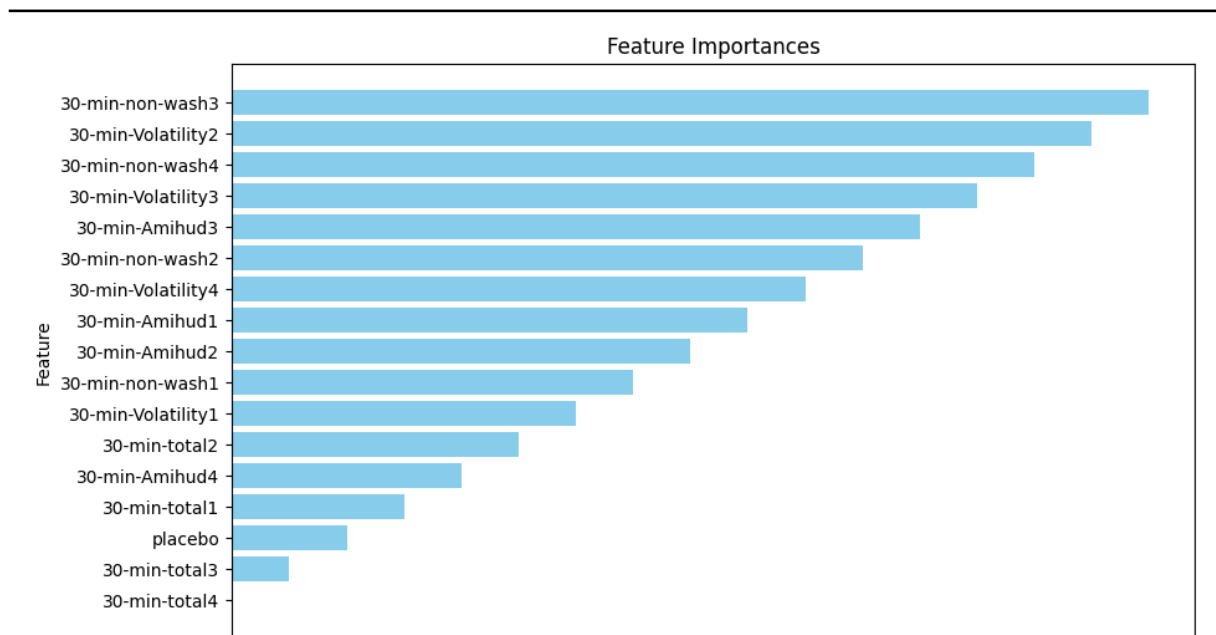
Random Forest



Adaboost



GRU Model



LSTM Model
