Direct Evidence of Bitcoin Wash Trading*

Arash Aloosh† Jiasun Li‡

August 28, 2023

Abstract

We use the internal trading records of a major Bitcoin exchange leaked by hackers to detect and characterize wash trading — a type of market manipulation in which a single trader clears her own limit orders to "cook" transaction records. Our finding provides *direct* evidence for the widely-suspected "fake volume" allegation against cryptocurrency exchanges, which has so far only been backed by *indirect* estimation. We then use our direct evidence to evaluate various indirect techniques for detecting the presence of wash trades and find measures based on Benford's law, trade size clustering, lognormal distributions, and structural breaks to be useful, while ones based on power law tail distributions to give opposite conclusions. We also provide suggestions to effectively apply various indirect estimation techniques.

Keywords: bitcoin; cryptocurrency; exchanges; forensics; market manipulation; regulation.

*We thank Patrick Augustin, Bruno Biais (Editor), Will Cong, Neil Gandal, Giacomo Sbrana, Christine Parlour, Leif Anders Thorsrud, Andrew Urquhart, Baolian Wang, David Yermack, and three anonymous referees, as well as attendants at the NBER Summer Institute, the 3rd Cryptocurrency Research Conference, DC Junior conference, European Alternative Finance Research Conference, Cardiff Fintech Conference 2022, EFMA 2022, Rensselaer Polytechnic Institute, the digital finance chair of Université Paris 2 & Telecom Paris, EDHEC Business School, ICMA Centre at the University of Reading, Nankai Business School, Central University of Finance and Economics, the Renmin-HKBaptist-NTU Joint Digital-Economy Seminar Series, Sharif University of Technology, the Int'l Joint Conference on Theoretical Computer Science, and WFC for helpful comments. Jiasun Li acknowledges support by a U.S. Department of Homeland Security award #205187 through the Criminal Investigations and Network Analysis Center (CINA).

†arash.aloosh@devinci.fr. Léonard de Vinci Pôle Universitaire, Paris La Défense 92916, France.

‡jli29@gmu.edu. George Mason University, 4400 University Drive, MSN 1B8, Fairfax, VA 22030, USA.

There have been suspicions that many cryptocurrency exchanges manipulate the market by permitting or even engaging in so-called wash-trading, in which fake trading records are created by having the same trader clearing his or her own standing limit order(s).[1] Since exchanges only display trade and quote information rather than the trader identity behind each trade or order, market participants are unable to distinguish between a wash trade and a genuine trade. Wash trading hence inflates the volumes of cryptocurrency exchanges and misleads market participants about the actual exchange liquidity conditions. While wash trading is explicitly banned in all regulated stock, derivative, or commodity exchanges, similar regulations do not apply to many cryptocurrency exchanges. Such suspicions are so widely believed that many industry data providers now offer adjusted "real" trading volume of cryptocurrency exchanges (See for example Blockchain Transparency Institute).

It has immediate policy implications to understand whether such suspicions are indeed valid in reality, and if so, whether we can detect them using publicly available information. First, it helps us weigh in the debate on whether cryptocurrency exchanges should be subject to more strict regulations, given that some exchanges have complained about over- or mis-regulation (e.g. New York attorney general vs. Kraken),[2] leading to some exchanges leaving the US market for good (e.g. Poloniex following its acquisition by Justin Sun) while others opting for working closely with regulators (e.g. Gemini founded by the Winklevoss twins). Second, understanding manipulations on centralized cryptocurrency exchanges could help us make more fair comparisons between centralized cryptocurrency exchanges and the emerging applications of decentralized exchanges, given the various types of manipulations identified among those applications.[3] Finally, in its numerous rejections against various bitcoin ETF proposals, the SEC has referred to potential market fragility due to thin volume disguised by seemingly abundant "fake" volumes. A deeper understanding of the issue may further enlighten the regulatory decisions.

That said, so far all studies into potential wash trading and fake volume on exchanges have

[1] Some anecdotes: On July 22, 2020, Coindesk reports that Coinsquare will settle with the Ontario Securities Commission (OSC) over allegations that executives had employees fake trades to inflate the platform's volumes. On Aug 26, 2020, CoinDesk EU News Editor Daniel Palmer reports that Coinbit, South Korea's third-largest cryptocurrency exchange, appears to have been seized by police over allegations that it faked most of its trading volume.

[2] Summarized in Underwood (2018) and Kraken's Blog. As a side note, Kraken's CEO Jesse Powell contributed to rebuilding the Mt.Gox exchange following its 2011 hack, an event to be discussed in more detail later in the paper.

[3] See, for example, Daian, Goldfeder, Kell, Li, Zhao, Bentov, Breidenbach and Juels (2020).

been exclusively based on indirect estimations. For example, Ribes (2018) documents potential fake volume on exchanges using price impact analysis from large order execution experiments. In an influential presentation to the SEC, Bitwise (2019) claims that 95% of cryptocurrency exchange volumes are fake, even though the accuracy of this estimate has been subsequently questioned. In the academic literature, the first paper we are aware of that examines wash trading on several exchanges is Cong, Li, Tang and Yang (2020). Specifically, they examine 3 regulated and 26 unregulated crypto exchanges by comparing the statistical patterns in their respective trading logs during July 9$^{\text{th}}$ and November 3$^{\text{rd}}$, 2019. For each exchange within their sample, they evaluate the extent to which the trading records exhibit patterns in first-significant-digit distributions, size rounding, and tail distributions that are consistent with typical observations in financial markets and nature: Significant deviations from the benchmark are then interpreted as suggesting the presence of wash trading. Based on this method, they find no wash trading on regulated exchanges, but rampant wash trading on unregulated ones. To quantify the magnitude, the authors compare the percentage of rounded-size trades on regulated exchanges versus that on unregulated exchanges and estimate that 70% of the volumes on unregulated exchanges are due to wash trading. Amiram, Lyandres and Rabetti (2020) also use their indirect estimation techniques, including methods based on Benford's law, the distribution of short-horizon trading activities, and the presence of structural breaks in trading activities, to infer the presence of wash trading and uncover a positive relationship between wash trading and exchange competition.[4]

Despite these ample *indirect* estimates, without *direct* evidence from individual-level transaction data, it is impossible to verify the usefulness of various *indirect* estimation techniques. Since all the techniques proposed in the literature are based on statistical irregularities, they cannot directly point fingers at wash trading, as the captured anomalous transactions may be motivated by various strategies other than wash trading (e.g. stealth trading à la Alexander and Peterson (2007)). Indeed, Tibeaudou (2019) has presented counter-arguments against some indirect estimation techniques for identifying wash trading. Furthermore, even if indirect estimations do accurately quantify the magnitude of wash trading, they still cannot indisputably answer many follow-up ques-

[4] Section 3 will give more details about the various indirect estimation techniques adopted in the literature.

tions that the community may be interested in: For example, what are the characteristics of wash trades and wash trading perpetrators, and are the latter associated with the exchanges themselves? Separately, given that most of the fake volume allegation surfaced *circa* 2018, one may also ask whether wash trading is a recent phenomenon, or it has been with cryptocurrency exchanges since their very early days.

In this paper, we fill these gaps by providing direct evidence on wash trading. Our analysis uses the internal transaction logs from Mt.Gox, the largest bitcoin exchange by reported volume during our sample period. The internal data include complete transaction records with timestamps, price levels, sizes, and most importantly, trader identities (IDs) for both the buyer and seller of each trade. The data was first leaked by hackers to the bitcoin community upon the exchange's collapse in 2014, and have subsequently been studied by both academics (Gandal, Hamrick, Moore and Oberman (2018)) and practitioners (Nilsson (2014) or more commonly known as the Willy report) to investigate potential market price manipulations that led to the 2014 Bitcoin bubble.[5] That said, our analysis focuses on a different period within the data and investigates potential volume manipulations rather than price manipulations. Section 1 provides more institutional details and summary statistics for our data.

Section 2 presents our direct evidence of wash trading on the major bitcoin exchange Mt.Gox: Among 16 million buy/sell records, representing 8 million unique trades, we find more than 115 thousand trades with both sides (buy and sell) having the same trader ID. These trades first emerged in June 26$^{\text{th}}$, 2011, immediately after the exchange restored itself following a one-week halt of service due to a hacker attack, and have been prevalent ever since until May 20$^{\text{th}}$, 2013, when the exchange was investigated by regulators for a separate litigation case. Within the June 26$^{\text{th}}$, 2011 - May 20$^{\text{th}}$, 2013 period, these trades constitute more than 2% of the total number of trades. Because a single person or entity could potentially register multiple trader IDs on Mt.Gox (using different email addresses), this number likely underestimates the true extent of wash trading in reality. Nevertheless, the finding does provide both indisputable direct evidence and a lower bound of wash trading activities. The remainder of the section then further characterizes these

[5] See also a Wall Street Journal article about the Willy report.

detected wash trades. We find that wash trades tend to be smaller in size, and are more likely to come from frequent traders. However, there do not seem to be trader IDs dedicated to wash trading — that is, those trader IDs who commit wash trading also do normal trades, and they are more likely to conduct wash trades following less active market conditions.

A major advantage of our direct evidence is that it provides a unique opportunity to empirically evaluate various indirect estimation techniques proposed for testing the presence of wash trades. Section 3 dives into this direction. We first find that those indirect estimation techniques based on the conformity with the Benford's law, the pattern of trade size clustering at rounded numbers, deviations from log-normal distributions, and the presence of structural breaks are all indicative of the presence of wash trades. We also point out that for a large sample with a wide range of trade sizes, the size clustering technique requires a careful choice of round number definitions — Based on this observation, we then further provide guidance for defining rounded trades among ones with vastly different sizes. We also find that the structural break-based approach is more effective when applied to logarithmically-transformed trading volumes, as opposed to non-transformed series or number of trades. Finally, we point out that indirect estimation techniques based on tail distributions' fit to the power law tend to give opposite predictions compared to our direct evidence, suggesting that such techniques may not be that reliable. Since these performance evaluations lay foundations for many further conclusions in the literature drawn from various indirect wash trading estimation techniques, we view the exercises in this section to be the main contribution of our paper.

Aside from assessing the respective merits and limitations of various indirect estimation techniques, our direct evidence also allows us to uniquely evaluate the power of simple heuristics for detecting the potential presence of wash trading. By the end of Section 3, we further use the facts that wash trades tend to inflate the apparent number of trades and tend to be smaller in size to construct a simple "wash trading index" based on publicly available information only. Specifically, the heuristic index follows the spirit of Comerton-Forde and Putniņš (2011) and estimates the probability of any particular short trading period to contain wash trading.[6] Using Receiver

[6] Comerton-Forde and Putniņš (2011) were developed in the stock market context and for a different purpose of detecting close-price manipulation. However, both approaches share the spirit of using direct evidence to train a

operating characteristic (ROC) curves from cross-validations, we find the simple heuristics to enjoy significant out-of-sample predictive powers. To put the finding in perspective, we also adapt the indirect estimation techniques presented earlier in Section 3 to construct similar indexes. The comparison shows that the performance of our simple heuristics is at par with other indirect techniques proposed in the literature.

One limitation of our work is that because different trader IDs may be controlled by the same person or entity on Mt.Gox, our direct evidence may only serve as a lower bound of the actual extent of wash trading. Section 4 further uses the trader ID-level information unique to our data to explore various methods to detect "hidden" wash trades. Our analysis adopts two different approaches: First, in what we call a volume-match-based approach, we look for patterns in which multiple traders collude to trade among themselves only to end up in the same position before trading. Although the volume-match-based approach has been proven useful on decentralized exchanges for wash trading detection, we find that it does not identify a significant number of "hidden" wash trades on our centralized exchange. Second, in what we call an insider-based approach, the premise is that to the extent the trades among exchange insiders themselves do not expose them to market risk, such trades provide another source of potential wash trading. Based on available information on exchange insiders' involvement in wash trading, the second approach does find more "hidden" wash trades. That said, overall our augmented wash trading sample still shows a significantly lower wash trading percentage than those reported in more recent samples. We discuss differences in market environments that are likely to account for such discrepancies.

Finally, Section 5 documents the impact of wash trades on the Bitcoin market, with the caveat of no causal claims. Using a vector autoregression (VAR) framework to characterize the joint dynamics between the intensity of wash trading and other market-level variables, we find that wash trades tend to have mild impacts on bitcoin returns and market liquidity, and are followed by significant yet short-lived positive responses in fee revenues to the exchange. We then discuss the relative merits of several possible explanations for these observations.

In sum, our finding provides direct evidence for the widely-suspected "fake volume" allegation

public measure and evaluate performances with cross-validation.

against cryptocurrency exchanges as well as external validations for various indirect techniques for detecting the presence of wash trades. That said, we remind readers that Mt. Gox operated at an early stage in the development of cryptocurrency markets, faced relatively less competition, and had a different strategy than other exchanges. This could generate different patterns of wash trading on Mt. Gox than on other exchanges.

Related literature Besides those papers mentioned in the introduction on indirect estimations of anomalous transactions in the cryptocurrency market, this paper is also related to the emerging literature documenting various manipulations in the cryptocurrency market. For example, using the same data as ours but focusing on a different time period, Gandal, Hamrick, Moore and Oberman (2018) and the Willy report relate the late-2013-early-2014 bitcoin bubble/crash to price manipulations on the Mt.Gox exchange by a trading bot commonly known as Willy, who started operation in September 2013. Such price manipulation is different from wash trading identified in the current paper, as the former is done by only one-sided buys with funds apparently created out of thin air, while the latter involves buying and selling simultaneously between oneself. Furthermore, wash trading in our sample also occurs prior to the time when Willy starts operation.

Griffin and Shams (2020) document another price manipulation, which relates the 2017 bitcoin bubble to Tether issuance from a single large bitcoin address. Specifically, this single address, allegedly belonging to Bitfinex, appears to "mint" Tether out of thin air and purchases Bitcoin mostly around major bitcoin price support levels so as to create an illusion of strong buying pressure and to push bitcoin prices higher. This price manipulation also involves only one-sided buying, rather than simultaneous buying and selling as in wash trading studied in this paper.

The third type of price manipulation concerns so-called pump-and-dump schemes. For example, Dhawan and Putniņš (2021), Li, Shin and Wang (2019), and Xu and Livshits (2019) provide direct evidence of pump-and-dump schemes in the cryptocurrency market using communication records on Telegram. Hamrick, Rouhi, Mukherjee, Feder, Gandal, Moore and Vasek (2018) further cover communication records on Discord. Such manipulations involve buying first and then selling later, instead of simultaneous buying and selling that features wash trading.

Finally, Dyhrberg, Foley and Svec (2019) examine pervasive undercutting behaviors in the cryptocurrency market, which is arguably manipulative. The authors also use their findings to show that tick sizes (in general) can be too small. Such undercutting manipulations do not involve simultaneous buying and selling at the price, and therefore also differ from wash trading studied in our paper.

In a broader sense, our findings also contribute to the emerging literature on the bitcoin and cryptocurrency market. For example, Makarov and Schoar (2020), Choi, Lehar and Stauffer (2020), and Yu and Zhang (2020) document large and recurrent arbitrage opportunities across exchanges and especially across borders. Li and Yi (2019), Liu and Tsyvinski (2021), Liu, Tsyvinski and Wu (2022), and Cong, Karolyi, Tang and Zhao (2022) study the factor structures in cryptocurrency returns. Shams (2020) and Benetton and Compiani (2020) relate crypto-asset returns to investor demands, while Schwenkler and Zheng (2021) relate them to co-mentions in news. Biais, Bisiere, Bouvard, Casamatta and Menkveld (2018) relate Bitcoin prices with changes in transactional benefits and costs of bitcoin. Augustin, Rubtsov and Shin (2020) studies the impact of introducing the Bitcoin futures contract on the spot market. Foley, Karlsen and Putniņš (2019) and Li, Baldimtsi, Brandao, Kugler, Hulays, Showers, Ali and Chang (2021) study the illegal usage of cryptocurrencies. Guo, Li, Luo and Wang (2022) studies the liquidity of Bitcoin options contracts. Our paper also touches upon the regulation of the crypto market in general, as in Li and Mann (2018, 2021).[7]

Our approach of observing actual manipulative conduct, characterizing the activity, constructing public measures to infer such activity, and using private data to test/validate "public" measures of the activity also shares the spirit of Comerton-Forde and Putniņš (2011), even though that paper is in a different domain of measuring closing price manipulations.

The remainder of the paper is organized as follows: Section 1 describes the data. Section 2 presents direct evidence of wash trading. Section 3 evaluates various indirect wash trading measures using our direct evidence. Section 4 explores different approaches to uncover "hidden" wash trades. Section 5 investigates the market impact of wash trading. Section 6 then concludes.

[7] For more examples within the now vast literature of blockchains in general, see Cong, He and Li (2021), Halaburda, He and Li (2021), Li (2021), Chatzigiannis, Baldimtsi, Griva and Li (2022), He, Li and Wu (2023), Li (2023), Li and Wu (2023), and Petryk and Li (2023), etc.

1 Data description

Our data comes from the internal trading records of Mt.Gox, which is widely regarded as the first, and for a long time during its life (before its collapse in February 2014), the only major bitcoin exchange in the world. Figure 1 Panel A presents the market shares of Mt. Gox in terms of trading volume over time. Individuals could deposit various fiat currencies to the exchange and purchase bitcoin, or conversely, deposit bitcoin to the exchange and sell for various fiat currencies of their own choice.

[Figure 1 about here.]

Mt.Gox profits from collecting fees from all transactions. Within each fiat-bitcoin trade, the trader who buys bitcoin with fiat currency incurs a bitcoin fee (in units of bitcoin) and the one who sells bitcoin for fiat currency incurs a fiat fee (in units of the receiving fiat currency). Both fees are calculated as percentages (typically 0.60%) of their corresponding trade sizes, and each trade has to pay a minimum fee (typically 1 satoshi, i.e. 10^{-8} BTC for bitcoin fees and ¥0.001 for fiat fees) regardless of its size. In the original file, fiat fees are labeled as "money fees" so we will use fiat fees and money fees interchangeably in the rest of the paper. The fee revenues to the exchange sum up both legs for each trade after currency conversions. Figure 1 Panel B presents a fee calculator used by traders on Mt.Gox to illustrate the bitcoin fees and fiat fees.

Figure 2 plots the evolution of bitcoin prices during our sample period and chronicles several major events throughout Mt.Gox's life. Of particular interest is the date of June 19$^{\text{th}}$, 2011, when Mt.Gox recovered from a hack that brought down its website for about a week. Wash trading first emerges immediately after this incident.

[Figure 2 about here.]

The data were first leaked by hackers in February 2014 upon the collapse of Mt.Gox. The entire leaked zip file in its original form contains 60 "trade" files (at least one for each month from April 2011 to November 2013). There are more trade files than the number of months because for some months by the end of the sample, as bitcoin trading volume spiked there was one file for each week.

Also for April 2013, there is an extra anonymized log in addition to the full log, which differs in two and only two places that we will elaborate further in Appendix B for identifying an exchange insider nicknamed "MagicalTux".

Each trade file includes second-by-second transaction records, with information that includes transaction ID, trader ID, amount (in both units of BTC and fiat currencies), time, order type (buy/sell), currency, JPY exchange rate, user location (country, state), a dummy for whether the trader's location is based in Japan or not (JP/NJP), and the two type of fees (fiat fees and Bitcoin fees).[8]

Merging all the trade files renders approximately 9 million unique transactions on Mt.Gox from April 1st, 2011 to November 30th, 2013. The original log however contains many duplicate transactions. We use a de-duplication process similar to that of Feder, Gandal, Hamrick and Moore (2017). Specifically, we remove duplicate transaction records with the same transaction ID, trader ID, transaction time, transaction type (buy/sell), and transaction amount.[9] Removing duplicates narrows the data to about 8 million unique transactions, which is closer to the daily volumes reported on bitcoincharts.com than the original leaked data.[10]

Following Nilsson (2014), we further exclude all "quartet transactions" before March 2013 by trader IDs TIBANNE_LIMITED_HK and THK, which are the only non-numerical IDs in our data, and are widely believed in the bitcoin community to belong to a "super user" created by Mt.Gox's parent company Tibanne Co. Ltd.[11] It is argued that before March 2013, this super user's main role was to facilitate cross-currency trades.[12] Table 1 reports summary statistics for the cleaned

[8]Figure A1 in the appendix presents a snapshot of the original files.

[9]The de-duplication procedure in Feder, Gandal, Hamrick and Moore (2017) also relies on fields including trader ID, transaction time, transaction type (buy/sell), and transaction amount, but does not include transaction IDs. We believe the inclusion of transaction IDs renders a more accurate result since, for example, a trading bot may post two identical orders, which are cleared at the same time, but go by different transaction IDs. Both methods nevertheless generate similar results.

[10]To validate our de-duplication process, Figure A2 in the appendix compares the daily volume calculated from our transaction records.

[11]Tibanne registered a company in Hong Kong on May 19th 2011, creating a "super user" with ID TIBANNE_LIMITED_HK, and later changed to THK in March 2013. This ID enjoys an exceptionally high privilege as it can trade without incurring any fees. For example, its first trade on August 27th 2011 7:48 bought 1BTC with USD and then at the same time sold 1 BTC to JPY, neither incurring any fees. The ID THK traded in a total of 2.8M BTC and 350M USD till the shutdown of Mt.Gox. As far as we know, THK is first documented in a post by pseudo-name HTCFOX, and then popularized in Reddit.

[12]Since March 2013, which was regarded as the start of the 2013 bitcoin bull run, however, all of THK's trades were BUYs and no SELLs.

sample. These transactions represent 108 million Bitcoins in trading volume (54 million Bitcoins buy/sell) from April 1st, 2011 to November 30th, 2013.

[Table 1 about here.]

Mt.Gox serves a global customer base from more than 160 countries and supports 17 fiat currencies. Mt.Gox users can opt to verify their accounts on Mt.Gox to shorten the waiting periods upon withdrawal. For verified users, Federal Information Processing Standard (FIPS) state codes will be entered into the User_Country and sometimes User_State variables in the corresponding trade file entries. The two variables remain empty for unverified users, take values of "`!!`" for users who failed verification, and for a small number of anomalous trader IDs (about 50) take a value of "`??`".[13] Although Mt.Gox is headquartered in Japan and uses JPY as the clearing currency internally, the United States is the largest market with about 34% of total transactions, and the U.S. dollar is used among more than 85% of the transactions.

2 Direct evidence of wash trading

The U.S. Commodity Futures Trading Commission defines wash trading as "*entering into, or purporting to enter into, transactions to give the appearance that purchases and sales have been made, without incurring market risk or changing the trader's market position.*" Designating wash trades in practice is however nontrivial because it is often disputable what a trader's true intention behind a trade is. We hence start with a narrow wash trading definition, which labels a transaction as a wash trade if and only if both its buy side and sell side have the same trader ID.[14] This narrow definition likely underestimates the actual magnitude of wash trading, it nevertheless is one of the most intuitive ways to identify direct evidence for cryptocurrency wash trading and enjoys the advantage of being non-controversial. We thus first follow this narrow definition to provide indisputable evidence for the presence of wash trading, and present characteristics of both wash trades and wash traders. Section 4 later will discuss different approaches to further uncover "hidden" wash trades and augment the wash trades sample.

[13] The "`??`" trader IDs will be used later for exchange insider detection in Appendix B.

[14] Therefore, we do not consider tax-motivated wash trades as in Cong, Landsman, Maydew and Rabetti (2023).

[Figure 3 about here.]

Figure 3 illustrates the presence of "self-self" wash trades in our sample. Panel A plots the daily wash trading volume on Mt.Gox during our entire sample (April 1$^{\text{st}}$, 2011 to November 30$^{\text{th}}$, 2013). As the figure demonstrates, wash trading emerges immediately after the 2011 Mt.Gox hack which brought down the exchange for about a week.[15] Wash trading has been prevalent ever since this incident until a sudden stop in 2013 following the U.S. Department of Homeland Security's seizure of $5 million in assets from Mt.Gox's U.S. branch. Therefore, wash trading has been in existence on Bitcoin exchanges far earlier than it becomes aware to the community. In the rest of the paper, unless otherwise noted, we will focus our attention on the period in which wash trading was active — that is, from June 26$^{\text{th}}$, 2011 to May 20, 2013, a period with 5,548,867 transactions and a total volume of 43,515,392 BTC. These numbers will be used in subsequent analyses as the denominators for calculating various percentages of wash trading. Panel B further plots the daily share of wash trading volumes within the total daily trading volumes on Mt.Gox. As the figure demonstrates, the volume share of wash trades varies over time and could account for as high as 60% of the daily trading volume.

[Table 2 about here.]

Overall, during the period between June 26$^{\text{th}}$, 2011 and May 20$^{\text{th}}$, 2013, we find 115,275 transactions with both the buy and sell side having the same trader ID. These wash trades account for 2.1% of the total number of transactions and 1.4% of the total amount of bitcoin traded. These wash trades involve 2,887 unique trader IDs, whom we denote as *wash trader IDs*. There are 1,821,366 transactions among these wash trader IDs during the June 26$^{\text{th}}$, 2011 and May 20$^{\text{th}}$, 2013 period, which account for about 32.8% of the total number of transactions and 32.5% of the total amount of bitcoin traded during the same period. This percentage reveals active interactions among the 2,887 wash trader IDs, as they are only a tiny fraction (about 2.9%) of the 98,391 total number of unique trader IDs during the same period. Table 2 tabulates these statistics.

[15] See Roy (2018) for a vivid narration, where several famous figures in the Bitcoin community, including Jesse Powell, who later created a major cryptocurrency exchange Kraken, and Roger Ver, who later served as the CEO of Bitcoin.com and co-created Bitcoin Cash, helped with bringing the site back online. Figure A3 in the appendix illustrates this incident.

2.1 Remarks on the magnitude

The magnitude of wash trading identified in this paper is significantly lower than those estimated in other studies. For example, based on a statistical model using trade-size clustering patterns, Cong, Li, Tang and Yang (2020) estimated about 70% of wash trades on unregulated exchanges; Using proprietary methods, Forbs and Bitwise (2019) claim 51% and 95% of fake volume on cryptocurrency exchanges, respectively. It thus begs answers for why such discrepancies arise. We list two potential explanations below:

1. Different market environments: Our sample period is within 2011-2013, which is now widely regarded as the early days in Bitcoin's history. For a significant proportion of this sample period, Mt.Gox was the only Bitcoin exchange. Even though other exchanges entered the industry near the end of our sample, Mt.Gox remained the dominant one. On the other hand, the estimates in other studies all come from periods after 2018, when the industry for cryptocurrency exchanges has been featuring fierce competition among hundreds of exchanges, and the fake volume problem on cryptocurrency exchanges began to receive wide attention. In a market with fierce competition among exchanges, the incentives for wash trading tend to be much higher because an apparent high volume could help an exchange attract more traders (and thus collect more trading commissions) as liquidity begets liquidity, and it could also help an exchange attract more entrepreneurs to list their coins on the exchange (and thus collect more listing fees). Neither incentive, however, was present during our sample period. Therefore, we expect a much lower magnitude of wash trading during our sample period compared to more recent samples.[16]

2. Different methodologies: As mentioned earlier, to establish indisputable direct evidence of wash trading, as a first pass we adopt a narrow definition for wash trades. Besides, another feature of our data also tends to underestimate the magnitude of wash trades: Since Mt.Gox imposed no mandatory know-your-customer (KYC) requirement (user verification is

[16] Amiram, Lyandres and Rabetti (2020) empirically establish a positive relationship between wash trading and market competition. The fact we find a low intensity of wash trading on an exchange not facing fierce competition may also be viewed as an external validation of their results.

optional), the same person may open as many accounts as possible. In other words, it is possible that multiple different trader IDs in our data are controlled by the same person or entity. Therefore, our narrow definition of wash trades as being transactions whose buyer and seller share the same trader ID may only detect a subset of wash trades, and our evidence so far should be interpreted as a lower bound of wash trading magnitude, rather than a precise quantification. On the other hand, estimates in other studies rely on statistical patterns to detect "anomalous" trades, which may or may not be exclusively wash trades. Therefore, we expect a much lower magnitude of wash trading under our method compared to those in other studies.

Although our direct evidence may under-represent the actual extent of wash trading, they are nevertheless useful for evaluating the indirect estimation techniques proposed in the literature. Section 3 will take on this task. Section 4 will further explore potential approaches to uncover "hidden" wash trades beyond our narrow definition. Before moving on, however, we first provide more in-depth characterizations of our detected wash trades.

2.2 Characterizing wash trades

Given the identified wash trades, we further characterize such manipulative conducts at both the trade and trader levels.

[Figure 4 about here.]

Figure 4 presents a trade-level comparison between wash and non-wash trades. We are interested in the patterns that wash trades exhibit regarding both trade sizes and clustering at round sizes. As the figure illustrates, both wash and non-wash trades exhibit trade size clustering at round numbers. Although not visually obvious, as we calculate the average trade size, we further find that wash trades tend to have smaller trade sizes compared to regular trades. Indeed, the average (median) trade size of non-wash trades is 7.90 (0.62) BTC, while the average (median) trade size of wash trades is 5.26 (0.01) BTC.

[Table 3 about here.]

With the trade-level characterization, we further explore trader-level information. Specifically, we decompose each wash trader ID's transactions by its counterparties. Table 3 then presents the number of trades, volume (BTC), and volume (¥) attributed to a wash trader ID's transactions with (1) self (own ID), (2) other Wash Trading IDs (i.e. traders that have self-self trades), and (3) Regular IDs (i.e. traders that have no self-self trades). Table 3 reveals several salient characteristics of wash traders. First, they tend to trade a lot more than non-wash traders: Even though wash traders only account for about 2.9% of the entire trader population, for an average wash trader, transactions with wash traders account for about one-half of all her trades. This finding, together with the trade-level evidence of wash trades having smaller trade sizes on average documented in the previous paragraph, is consistent with the community knowledge that fake orders tend to feature small trade sizes but large total amounts (Vigna and Osipovic (2018)). Second, there do not seem to be trader IDs that are primarily registered for wash trading purposes; instead, wash traders tend to engage in self-self trades, trades with other wash traders, and trades with regular non-wash traders as well. For an average wash trader, self-self wash trades only account for 3.35% (1.61%) of her total number of trades (total BTC trading volume), with the maximal percentage across the wash trader population being 49.72% (47.81%). Therefore, wash traders conduct both wash and non-wash trades, rather than doing wash trades only.[17]

[Table 4 about here.]

Since wash traders conduct both wash and non-wash trades, it is natural to ask when they choose one over the other. We therefore look for variables associated with wash trading. Table 4 presents market conditions under which wash traders choose to conduct wash trades versus normal (non-wash) trades. Specifically, like in Table 3, we decompose all wash trader's transactions into three groups: self-self wash trades, trades with other wash traders, and trades with non-wash trades. Within each group, we calculate the mean and standard deviation of various variables of market conditions, measured by the previous-minute genuine (non-wash) trading volume, market illiquidity as measured by the Amihud measure, Bitcoin return, and realized volatility, respectively.

[17]This conclusion is admittedly subject to the caveat over our narrow definition of wash trades. Section 4 will explore various approaches to further uncover "hidden" wash trades.

Overall, we find more wash trades following lower non-wash volumes and lower liquidity (i.e. higher Amihud measure), as well as slightly lower returns and higher volatility.[18] The evidence suggests that traders choose wash trading over normal (non-wash) trading when the market is less active. One plausible reason for this finding is that some of the wash traders may be controlled by the exchange itself, and they conduct wash trading at inactive times to "reinvigorate" the market. Section 4.2 and 5 will further explore evidence that evaluates this economic interpretation (as well as alternative economic channels). That said, we caution readers that while it is tempting to label market inactivity and the associated variables as the "determinants" of wash trading, the discussions here may not necessarily have causal interpretations due to a lack of exogenous shocks.

3 Evaluating indirect estimation techniques

Because of the wide interest in the cryptocurrency community over volume manipulations on exchanges, in addition to our direct evidence on wash trading, many studies have also looked into this topic. However, without having access to internal trading records as we do, all other studies have to resort to indirect estimations. It is then natural to ask how accurate these indirect estimations are, especially given that indirect estimations based on statistical patterns may detect patterns not solely driven by wash trading but by various other trading strategies. Our internal trading records from the exchange thus provide a unique opportunity to evaluate the merit of those proposed indirect estimation techniques.

Based on a review of the literature, we evaluate the following indirect estimation techniques: compliance with Benford's law, patterns of trade size clustering at round numbers, the shape of the power-law tail distribution, trading volume (number of trades) deviations from the log-normal distribution, and the presence of structural breaks identified by non-parametric machine-learning algorithms. The subsequent subsections further explain these techniques and present our evaluation results.

[18] This finding concerns the extensive margin, that is, regarding the occurrence of wash trades. At the intensive margin, that is, regarding the size of wash trades, we also find these effects to be more salient among small wash trades: As we further decompose self-self wash trades into large and small ones (whose size is above or below the median wash trades sizes), we find that it is small wash trades that are preceded by even lower genuine volumes, lower liquidity, lower returns, and higher volatility.

3.1 Benford's Law

The first indirect estimation technique we evaluate takes advantage of Benford's law (Benford (1938)), which suggests that the first significant digits of trade sizes on an exchange should follow a logarithmic distribution benchmark. Specifically, the percentage of trade sizes with a leading digit of $i \in \{1, 2, \cdots, 9\}$ should ideally be $\log_{10}(i + 1) - \log_{10}(i)$, and the extent to which the empirical distribution deviates from the benchmark (measured by a χ^2 statistic) is arguably useful for detecting anomalous trading behaviors. Building on this assumption, Cong, Li, Tang and Yang (2020) infer that a majority of unregulated cryptocurrency exchanges exhibit anomalous activities during a period circa July 2019 - Nov 2019, while regulated exchanges do not show anomalies. Based on this observation, the authors infer that many (less prominent) unregulated exchanges are likely involved in wash trading. An alternative specification around Benford's law has also been used in Amiram, Lyandres and Rabetti (2020) to investigate the cross-sectional relationship between anomalous trading and market competition.

While Benford's law has been used widely in other settings, without an evaluation against direct evidence, it remains an empirical question about the validity of such indirect estimations in identifying wash trading. We fill this void by taking advantage of the direct evidence in our data. Specifically, we compare the sample of wash trades detected from our direct evidence and the sample of remaining non-wash regular trades and investigate the extent to which they respectively conform to Benford's law. Figure 5 Panel A illustrates the comparison and formally presents the χ^2-test results.

[Figure 5 about here.]

As Figure 5 Panel A demonstrates, in the sample of wash trades, the empirical distribution of the first digits of trade sizes indeed significantly deviates from the theoretical benchmark stipulated by Benford's law: The χ^2 statistics is 69.71 with a p-value that significantly rejects the null hypothesis that the empirical distribution and the theoretical Benford benchmark do not differ. On the other hand, in the sample of non-wash trades, the empirical distribution of the first digits of trade sizes does not significantly deviate from the Benford benchmark: The χ^2 statistics is only 7.91 with a p-

value of 0.443, which fails to reject the null hypothesis of an indistinguishable empirical distribution from the Benford benchmark. Overall, the evaluation based on our direct evidence provides support for the effectiveness of inferences based on Benford's law.

3.2 Trade size clustering

The second indirect estimation technique we evaluate relies on the fact that human-generated trades tend to cluster at round numbers (Rosch, 1975). To formally evaluate the effectiveness of trade-size clustering patterns for detecting the presence of wash trading, we follow Cong, Li, Tang and Yang (2020) and examine whether trade-size clustering appears at rounded sizes.[19] To quantify the magnitude of trade-size clustering, we compare trade frequencies at rounded trade sizes with the highest frequency of nearby unrounded trades, where trade frequency is calculated as the ratio between the number of trades at a specific trade size and the total number of trades within the nearby window. We then conduct Student's t-tests of the frequency differences between rounded and unrounded trades, for the wash trades subsample and non-wash trades subsample, respectively.

[Table 5 about here.]

Table 5 presents results from the above analysis. First, we find that a naïve application of the specification in Cong, Li, Tang and Yang (2020) is not helpful for detecting wash trades: In the fourth row of Panel I and II, respectively, we follow Cong, Li, Tang and Yang (2020) and first define a "base unit" as 10^{-4} BTC. We then define one "rounded-size unit" as 100 or 500 base units (100X or 500X), and report the t-statistics from comparing frequencies of trades with sizes that are multiples of rounded-size units and the highest frequencies within their nearby windows, that is, within (100X-50, 100X+50] or (500Y-100, 500Y+100]. As the table shows, we find all differences to be significantly negative, suggesting that trade frequency at round sizes is actually lower than unrounded ones by a large margin regardless of trade type (non-wash or wash trades), which is inconsistent with the findings in Cong, Li, Tang and Yang (2020). We also note that this

[19] As an illustration of trade-size clustering patterns, Figure A4 in the appendix also plots the histograms of trade sizes to illustrate potential clustering, for both the wash trading sample and non-wash trading sample. Comparing the histograms of wash trades and non-wash trades, we observe trade-size clustering around round numbers in both groups, although visually the clustering tends to be more salient among non-wash trades. The findings suggest that the patterns in trade-size clustering have the potential to serve as a noisy indicator of the presence of wash trading.

inconsistency is not due to specifications of rounded-size units — as we alter them from 100×10^{-6}, 100×10^{-5}, and up to 100×10^{-1} BTC, we keep finding similar patterns, as rows 1-3 and 5-6 in Panel I and II of Table 5 further report.

A hasty conclusion from the above finding would seem to refute the trade-size clustering approach in Cong, Li, Tang and Yang (2020). However, as we dig deeper, we uncover that the trade-size clustering approach is still valid once applied properly. We first note that the reason behind the apparent inconsistency between our finding above and those in Cong, Li, Tang and Yang (2020) lies in the fact that the choices of rounded-size units are crucial for the effectiveness of the size-clustering approach — specifically, an appropriate definition for rounded-size units should take trade size into account. For example, for small trades with sizes around say 1 BTC, it may be reasonable to regard trades with sizes of multiples of 0.1 BTC as rounded, but for large trades with sizes around say 300 BTC, it may be more reasonable to regard trades with sizes of multiples of 10 BTC as rounded.[20] Cong, Li, Tang and Yang (2020) do not suffer from this problem because that paper focuses on a short period (from July 9th 2019 to November 3rd, 2019, during which Bitcoin prices stayed within the same order of magnitude (within about \$7,500 to \$12,000). However, when applying the method to a sample with a longer time horizon, during which Bitcoin prices moved across a range of orders of magnitudes (within about \$1 to \$400) and accordingly a very wide range of trade sizes (e.g. from one satoshi up to more than 34K BTC in our sample), a careful selection of round numbers and nearby window sizes becomes important.

Given the reasoning above, in Panel III and IV we further decompose trades into different size buckets. Within each size bucket, we then apply different rounded trade definitions to reevaluate whether differences exist between wash and non-wash trades in terms of trade-size clustering patterns. The results in general confirm the findings in Cong, Li, Tang and Yang (2020): On the one hand, for non-wash trades across all size buckets, trade frequencies at round sizes are higher than those at unrounded sizes by a large margin (all significant at the 1% level except for one, which is significant at the 5% level); On the other hand, the differences between trade frequencies at round

[20] On the one hand, too large a rounded trade definition would render all small trades unrounded; On the other hand, as larger orders at the tail of the trade size distribution tend to be sparsely distributed, too small a rounded trade definition would render non-rounded large orders to take 100 percentage shares within their respective nearby windows and thus overwhelm the overall test statistics.

sizes and those at unrounded sizes are significantly smaller among wash trades. Out of the twelve size bucket/rounded size pairs we evaluate, eleven of them see no significantly higher frequencies at rounded sizes than at unrounded sizes. The remaining one also sees a lower difference than compared to non-wash trades.

Overall, we demonstrate that once trade size roundness is appropriately defined, our direct evidence supports using trade clustering patterns at rounded sizes to evaluate the presence of wash trading. Specifically, for various magnitudes of trade sizes, we confirm that the frequencies of rounded-size trades are significantly higher than those of unrounded-size trades among non-wash trades, but the same pattern either does not exist or is weaker among wash trades. Panel III and IV of Table 5 also compares rules-of-thumb for pairing size cutoffs and corresponding roundness definitions/window sizes when dealing with our long sample. For example, the size-clustering approach appears to work better when rounded trades are defined as those whose sizes are multiples of 5 (at different orders of magnitudes), which is also consistent with similar observations in Cong, Li, Tang and Yang (2020).

3.3 Power tail distribution

The third indirect estimation technique we evaluate takes advantage of the fact that typical financial markets expect heavy tails on the distribution of large transaction sizes (Gabaix, Gopikrishnan, Plerou and Stanley, 2003). Specific to the crypto market, Cong, Li, Tang and Yang (2020) test whether the cumulative distribution of trade sizes satisfies

$$\mathbb{P}(X > x) \sim x^{-\alpha}, \tag{1}$$

where the power-law exponent or the tail exponent α is expected to fall in the Pareto–Lévy regime, that is, α is expected to fall in the range of $(1, 2)$.

We test the hypothesis for both the wash trading subsample and the non-wash trading subsample. Following Cong, Li, Tang and Yang (2020), we first set the start of the tail as the top 10% of the largest trades for each subsample, and then apply both the OLS estimator and the Hill estimator to estimate α. The former takes the logarithm of the empirical probability density function and fits the log-log data to power-law distribution by Ordinary Least Square (OLS), while the latter calculates the Hill estimator as defined by (Clauset, Shalizi and Newman (2009); Hill (1975)):

$$\hat{\alpha}_{Hill} \equiv 1 + n \left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right)^{-1}, \tag{2}$$

where n is the number of observations and x_{min} is start of the tail.

Figure 5 Panel B illustrates tails of trade-size distributions and the fitted power-law lines on log-log plots for both the wash trading sample and non-wash trading sample.[21] Fitted power-law lines are plotted with parameters estimated by Ordinary Least Square (OLS) and Maximum Likelihood Estimation (MLE), shown in black and red lines, respectively. Blue dots represent empirical data points for trade-size frequencies. Overall, we find the tails of trade sizes to feature power-law distributions for both the wash trading and non-wash trading sample.

Figure 5 Panel B also presents the coefficients from OLS and MLE fittings for the wash trading subsample and the non-wash trading subsample. We find that the α coefficients estimated from wash trades feature $\hat{\alpha}_{OLS} = 1.883$ and $\hat{\alpha}_{Hill} = 1.580$, which fall within the Pareto-Lévy range ($\hat{\alpha} \in (1, 2)$); On the other hand, the coefficients estimated from non-wash trades feature $\hat{\alpha}_{OLS} = 2.674$ and $\hat{\alpha}_{Hill} = 2.054$, which fall out of the Pareto-Lévy range ($\hat{\alpha} \notin (1, 2)$). These findings stand in contrast with the assumption in Cong, Li, Tang and Yang (2020), who theorize that normal trades tend to have tails in trade sizes that follow power-law distributions with α coefficients within the Pareto-Lévy range. Therefore, based on our direct evidence we suggest that the power law test may not be an accurate method for testing the presence of wash trades in a sample.

3.4 Distance from log-normal distribution

The fourth indirect estimation technique we evaluate concerns the distributional properties of all trades rather than only those at the tail. Richardson, Sefcik and Thompson (1986) and Ajinkya and Jain (1989) argue that in the absence of evident manipulation, trading volume tends to follow a log-normal distribution. In the context of cryptocurrency trading, Amiram, Lyandres and Rabetti

[21] As in Cong, Li, Tang and Yang (2020), we apply the Python package Powerlaw as discussed in Alstott, Bullmore and Plenz (2014) to fit the data and calculate exponents.

(2020) first compute, for each month, the mean and standard deviation of log trading volume (as well as the number of trades) over ten-minute intervals, and then quantify the distance between the cumulative distribution function (c.d.f.) of the resulting empirical distribution and that of a normal distribution with the same mean and variance (the reference distribution) using the Kolmogorov-Smirnov (KS) statistic, that is, the supremum of the distance between the cumulative distribution functions of the empirical distribution and the reference distribution. We apply their method to our direct observed sample of wash trades and non-wash trades respectively to evaluate the merit of the KS measure.

[Table 6 about here.]

Table 6 presents formal statistical tests comparing the cumulative distribution function of both trading volumes and the numbers of trades (both transformed by natural logarithm) and the cumulative distribution functions of normal distributions with the same means and variances — for the respective samples of wash trades and non-wash trades. We observe that compared to the wash trades sample, the non-wash trades sample tends to have cumulative distribution functions that more closely align with the theoretical benchmark, regardless of whether the cumulative distribution function is calculated for trading volumes or the numbers of trades. We also observe that within the sample of non-wash trades, the alignment is closer for the number of trades than for the trading volume.[22] Specifically, the cumulative distribution function of the log trading volume for non-wash trades is significantly closer to normal than wash trades, with average Kolmogorov-Smirnov distances of 0.14 versus 0.05 in the wash and non-wash sample, respectively. Even stronger results are obtained for number-of-trades-based Kolmogorov-Smirnov distances, with the average Kolmogorov-Smirnov distances being 0.22 versus 0.04 in the wash and non-wash sample, respectively.

Overall, the finding provides support for Amiram, Lyandres and Rabetti (2020) regarding their use of the deviations from log-normal distributions as indirect measures to gauge the extent of wash trading. Besides, we also record a few additional findings: First, as mentioned earlier it seems that the number of trades conforms better with the log-normal test than the volume in our

[22]Figure A5 in the appendix further illustrates both observations in a representative month (July 2011).

sample. Second, we caution that although the log-normal test is *qualitative* powerful, its *quantitative* implications may be sample dependent. For example, while the average volume-based and number-of-trades-based KS distances for our wash trading sample are 0.14 and 0.22 respectively, the sample of Amiram, Lyandres and Rabetti (2020) in comparison has the lowest volume-based KS across exchanges to be 0.15 (Binance) and the highest to be 0.60 (ZB) while the lowest #-trade-based KS across exchanges is 0.14 (Binance) and the highest is 0.52 (ZB). This finding suggests that in practical applications we may get more reliable results solely from using qualitative comparisons.[23]

3.5 Structural breaks

The final indirect estimation technique we evaluate is based on the assumption that a sample plagued by wash trading tends to demonstrate more structural breaks, that is, the mean shifts or trend shifts in the time series of trading volume or number of trades. Specifically, Amiram, Lyandres and Rabetti (2020) adopt the so-called E-Divisive with medians (EDM, see e.g., James, Matteson et al. (2015) and James, Kejariwal and Matteson (2016)) algorithm to detect structural breaks. The EDM is a non-parametric machine-learning algorithm developed by computer scientists to flag unusual patterns in a time series, as it attempts to determine whether a new chunk of time series data is considerably different from the previous chunk through comparisons of various permutations of the data.

EDM has been argued to be suitable for detecting wash trading in that it is robust to the presence of short-lived abnormalities in the trading volume series, as trading volume data tend to have occurrences of peaks, which, despite often being a result of a natural data-generating process, may appear as anomalies when data series are aggregated into short intervals. In addition, EDM is nonparametric, so it accommodates the fact that volumes resulting from wash trading may not conform to a particular distribution, and the trading volume and number of trades series often contain more than one structural break.

[Table 7 about here.]

Following Amiram, Lyandres and Rabetti (2020), we use the monthly EDM-based number of

[23] Our later analysis in Section 3.6 will reflect this observation.

structural breaks as our measure of fake trading. Specifically, for each month we calculate the number of structural breaks within the time series of ten-min trading volumes and the number of trades. We then compare monthly summary statistics across the wash trading and non-wash trading sample. For fair comparisons, we only include full months in our sample, that is, from July 2011 to May 2013 (and exclude the last couple of days in June 2011).

Table 7 presents the summary statistics for the number of structural breaks within the time series of log ten-min trading volume (in the upper-left panel), log ten-min number of trades (in the upper-right panel), ten-min trading volume (in the lower-left panel), and ten-min number of trades (in the lower-right panel) within each month.[24] For both logarithmic series, we find that wash trades indeed tend to contain more structural breaks than non-wash trades. For example, in the time series of log ten-min trading volume, we find a monthly average of 7.32 structural breaks within the wash trading sample, versus no structural break within the non-wash trading sample; in the time series of log ten-min number of trades, we find a monthly average of 21.96 structural breaks within the wash trading sample, versus 1.18 structural breaks within the non-wash trading sample. For both specifications, the differences between wash and non-wash samples are statistically significant at the 1% level. Overall, these findings confirm the effectiveness of using structural breaks to detect the presence of wash trades.

That said, we also note that the EDM algorithm tends to be sensitive to nonlinear transformations. For example, if we do not take the logarithm of the volume or number of trade series, the number of structural breaks significantly drops. Even though we still see more structural breaks within the wash trading sample than the non-wash sample, the difference is not statistically significant. By visual inspection, we find that the EDM algorithm tends to label more structural breaks within the logarithmic data rather than raw ones, and overall gives the most accurate results in the logarithmic volume series in our sample.

[24] We follow James, Kejariwal and Matteson (2016) and Amiram, Lyandres and Rabetti (2020) to set the minimum bucket size, that is, the lowest number of observations (trading volume or the number of trades within a ten-minute interval) between structural breaks, at 6, as well as the degree of penalization at 1 and β at 0.008. See package and tutorial for more details about implementing the EDM algorithm. As an illustration of the formal results, Figure A6 in the appendix further uses the March 2012 series of log volumes as an example and visualizes the performance of the EDM algorithm.

Summary In sum, our evaluations of different families of indirect estimation techniques used in contemporaneous studies render a handful of new findings: First, they provide external validation for four families of methods: those based on Benford's Law, trade-size clustering patterns at round numbers, short-horizon trading activities' conformity with the log-normal distribution, and the presence of structural breaks in trading activities; On the other hand, we find the method based on power law distributions of tails tend to give opposite predictions than those in the literature, suggesting that the tail-distribution based approach may not be that reliable.

Second, within the four supported methods, our investigations above also allow us to compare the various techniques. Specifically, Benford's law is the easiest to implement as it requires no parameter choices, the trade-size clustering approach requires selecting appropriate round-size definitions according to trade sizes, the method based on conformity to log-normal distributions tends to work better on the number of wash trades than trading volumes, and the one based on the presence of structural breaks tends to work better on log-transformed trading volumes than non-transformed trading volumes or the number of trades.

Finally, we clarify that all the indirect estimation techniques we evaluate look for different statistical patterns within a *sample* of trades. In other words, all these methods apply to a population and reveal "group" characteristics, rather than serving as classification tools to label each individual trade as "wash" or "non-wash." In the next subsection, we continue with exploring a related but less demanding task of evaluating whether these indirect methods are helpful for labeling short trading horizons as ones with "wash" or "non-wash" trades. In this process, we also develop a "wash-trading index" based on simple heuristics.

3.6 Detecting periods with wash trading using public information

A unique advantage of our direct evidence is that it allows us to develop and evaluate simple heuristics for detecting whether a specific trading period is plagued by wash trading. Inspired by Comerton-Forde and Putniņš (2011), we develop the following simple heuristics for labeling whether a short trading horizon contains wash trading or not:

First, we look for variables that are associated with the presence of wash trading. From the trade-level characterization presented in Section 2.2, we already know that wash trades tend to be smaller in size, come from frequent traders, and artificially inflate the number of trades. Therefore, we expect periods with wash trading to feature both a higher number of observed trades as well as a lower median trade size. We then use both the number of observed trades and the median trade size within each ten-min window to construct an index for classifying trading windows into ones with wash trading and ones without. Specifically, the index Z_t for window t is defined as

$$Z_t = \sum_{\tau=6}^{6+6\times 24} (\#_t - \#_{t-\tau})^+ \times (\text{Size}_t - \text{Size}_{t-\tau})^-, \tag{3}$$

where $\#_t$ and Size_t respectively denote the number of trades and the median trade sizes within window t. In other words, Equation (3) looks into the whole day ending one hour before t, and counts the number of 10-min windows within the day that have both fewer numbers of trades than that in window t and larger median trade sizes than that in window t.[25] By comparing the number of observed trades and the median trade size in window t with the corresponding variables within a nearby day, we account for lower frequency inter-temporal trends in the underlying variables; By skipping one hour, we avoid the contamination from potential trade intensity clustering; By retaining only the sign of the differences rather than the magnitude of the differences, we avoid sample-specific quantities and only rely on qualitative comparisons to allow for potentially better out-of-sample classification performances.[26] We thus develop a wash trading index Z_t only using information publicly available to market participants.

To evaluate the accuracy of our wash trading index, we test it against our direct evidence. Specifically, we summarize the Type I/II errors from our index by plotting a Receiver Operating Characteristic (ROC) Curve. An ROC curve illustrates the performance of a classification algorithm by plotting the proportion of true positives (labeled as "sensitivity") against that of false positives (labeled as one minus "specificity") as one varies the classification thresholds. Therefore it

[25]We choose the benchmark window as one day with a gap of one hour for computational efficiency. The index remains valid if we alter the specification, e.g. taking the benchmark window as one week with a gap of one day.

[26]Therefore, our index construction follows the spirit of Comerton-Forde and Putniņš (2011). There is, however, one notable difference: Comerton-Forde and Putniņš (2011) construct close-price manipulation indexes for stocks so they can further use cross-sectional information to construct "difference-in-difference" measures. Since Bitcoin is the only asset traded on Mt.Gox, we do not have the luxury of cross-sectional information. This limitation tends to hamper the accuracy of our index, although we will show that it still enjoys significant accuracy.

characterizes the trade-off between Type I and II errors without a priori specifying a classification threshold. Indeed, it examines classification accuracy under all possible classification thresholds as opposed to a specific one. In our context, a ROC curve describes the necessary proportion of time windows out of ones that contain no wash trades but are erroneously classified as containing wash trades, in order to correctly classify a given proportion of time windows that indeed contain wash trades. For a classification algorithm to be effective, it should stay above the 45° line, and the larger the area under the ROC curve is, the more accurate the classification method is.

[Figure 6 about here.]

As the upper-left panel of Figure 6 illustrates, although our classification is only based on simple heuristics, it obtains favorable trade-offs between true and false positives. To put the classification accuracy in context, in the other panels of Figure 6, we further illustrate the performance of classification methods based on the various indirect estimation techniques discussed earlier. For example, the upper-right panel presents the ROC curve from the Benford χ^2-statistics — that is, we obtain the Benford χ^2-statistic for each ten-min window, count the number of 10-min windows within the whole day ending one hour before t that have smaller χ^2-statistics than that in window t to obtain anomalous χ^2-statistics, and then present the ROC curve. A similar procedure applies to the other panels, with the Benford χ^2-statistics in each window replaced by the percentage of unrounded trades (the lower-left panel) and the Kolmogorov-Smirnov (KS) distance between log trade sizes and normal distributions with the same mean and variance.[27]

As the panels illustrate, all methods demonstrate significant out-of-sample classification effectiveness. Notably, our simple heuristic index performs at par with, if not better than other indirect estimation techniques.[28] Our index could thus provide guidance to regulators as well as other market participants who would like to investigate potential wash trades but face resource constraints: Based on publicly available information only, they can first calculate the index for each time window, which can then advise them to start with windows that have the highest index values and

[27]Structural breaks do not fit the classification exercise here as they are developed for lower frequencies.

[28]In the appendix, we further combine all four measures using a logistic regression model with LASSO model selection and test out-of-sample predictive powers with both cross-validation and external validation. Since we find the performance enhancement from the "cocktail" model overall our simple Z-score is negligible, we prefer the simpler model presented here.

gather additional information to continue with further investigations.

Remark The construction of our index requires transaction-level data, which are typically not widely available to researchers. For example, while Kaiko provides trade-level data for about 80 centralized exchanges, other popular data sources such as CoinMarketCap and CryptoCompare cover over 500 exchanges but only provide lower-frequency (hourly) aggregated prices and trading volumes. Since many researchers rely on CryptoCompare and CoinMarketCap data for their empirical research, it would be helpful to explore whether one can construct predictors based on less granular data such as low-frequency trading volume and price changes. We thus further explore this possibility using our unique direct evidence. Specifically, we adapt our index in (3) by replacing the number of trades and the median trade sizes within each 10-minute window with the trading volumes and price returns with each hour and construct a new predictor over the prevailing week (skipping one day). Figure A12 reports our findings: Overall, a less granular index still has some predictive power (as ROCs still lie above the 45° line), although its accuracy is much lower than our more granular measure. Therefore, solely based on our data, it appears that resource-constrained researchers may partially mitigate wash trading concerns in their analyses with hourly volume and return data, even though we do not strongly recommend so for accuracy concerns. A useful path for future research is to further improve on this aspect and explore the possibility of constructing wash trading predictors based on low-frequency and publicly available data.

4 Identifying "hidden" wash trades

Section 2 has detected wash trading based on a narrow definition of same-user (self-self) trades. While this definition indicates indisputable wash trading behaviors, it may very well represent the tip of the iceberg and only provide a lower bound of the true magnitude of wash trading. On the one hand, because a trader could easily generate multiple IDs on Mt.Gox, it is possible that multiple IDs were controlled by the same party so that many trades that appear to take place between different IDs are actually wash trades as well; On the other hand, it is also possible that a group of associated traders collude and clear orders among themselves (instead of a single trader clearing her own orders) to "cook" fake transaction records. Trades in the latter case are typically referred to by market participates as painting the tape, and are closely related to the narrowed-defined wash trades earlier. To the extent that the associated traders enjoy shared financial interests, "painting the tape" trades also satisfy the CFTC's definition of wash trades as "*entering into, or purporting to enter into, transactions to give the appearance that purchases and sales have been made, without incurring market risk or changing the trader's market position.*"

In this section, we explore alternative methods to further uncover "hidden" wash trades using our unique data. We adopt two approaches to identify such transactions. The first approach uses a volume-matching algorithm, which aims to detect a "circle" of investors whose trades within a short span of time leave each investor's positions unchanged. The second approach uses multiple sources to identify trader IDs that are likely controlled by exchange insiders; To the extent all exchange insiders share the same financial interests, the trades among themselves can then be further classified as (broadly-defined) wash trades.

4.1 Volume-matching-based wash trading detection

While trader-ID level information on centralized cryptocurrency exchanges is typically unavailable except for rare cases like in our study, trades on decentralized exchanges do have public address-level information. As a result, studies in the computer science and engineering literature have proposed sophisticated detection algorithms to identify wash trading on decentralized exchanges. For example, Victor and Weintraud (2021) apply a volume-matching algorithm to decentralized exchanges IDEX and EtherDelta. Their method adopts a broader definition of wash trades, so that when multiple addresses execute a series of transactions among themselves, after which they all end up with the same market positions that they initially had, then all involved transactions are deemed wash trades. For example, a three-user scheme could involve trader A selling to trader B, trader B selling to trader C, and finally, trader C selling to trader A, so that after the whole sequence of transactions their market positions remain the same. Such complicated trading "circles" also satisfy the CFTC's definition of wash trading, as the intent of the involved parties is to execute fictitious trades that are not subject to market risk, even though they cannot be picked up by simply looking at "self-self" trades.

We adapt the volume-matching algorithm in Victor and Weintraud (2021) to our sample to potentially uncover more "hidden" wash trades. The algorithm looks for groups of trader IDs who conduct frequent transactions among themselves so that their trades within a given time window result in all involved trader IDs' ending positions equaling their beginning ones.[29] The exercise indeed reveals more trades that are part of some complicated trading chains. However, we find that the number of such "circle" trades is negligible compared to self-self trades: Requiring "circle" trades to finalize within one day, we only find 1,956 non-self-self circle trades.[30] Compared to the 115,275 self-self trades we identify in Section 2, these "circle" trades account for only 1.70% of the total number of (self-self) wash trades; Given 5,548,867 trades during our sample period, adding "circle" trades to augment the wash trades sample would only increase the share of wash trades within the total number of trades by 3.53 basis points.

[Figure 7 about here.]

As another illustration of the lack of frequent "circle" trading among colluding trader IDs, the left panel of Figure 7 plots the relationship between each trader ID's number of trades and unique trading partners across the entire trader ID population. Each cross represents one trader ID, positioned by its number of trades (horizontal axis) and the number of unique trading counterparties (vertical axis). For comparison, we also include Figure 2 in Victor and Weintraud (2021) on the right, which conducts the same analysis on decentralized exchange IDEX and EtherDelta. While they identify many traders that perform a large number of trades with only a few other counterparties (as we highlight in red circles), we do not find similar patterns on the centralized exchange Mt.Gox.

Our finding suggests that techniques useful for wash trading detection on decentralized exchanges may not be applicable to centralized exchanges. There are two potential explanations for this difference. First, decentralized exchanges are suits of smart contracts that operate on public

[29] In practice, to account for frictions we follow Victor and Weintraud (2021) and allow one's beginning and ending positions to differ by no more than 0.1 BTC; Lowering this buffer would only bias us toward finding fewer "hidden" wash trades.

[30] Even if we adopt an extremely relaxed assumption to allow "circle" trades to finalize within one week, we still only find 2,292 non-self-self circle trades.

blockchains, so that the addresses involved in all transactions on decentralized exchanges are publicly observable. Because of this transparency, any self-self trade on a decentralized exchange can be publicly detected and verified, and users with wash trading intent would have strong incentives to hide self-self trades behind more complicated trading "circles" to obfuscate their behavior. On the other hand, trader IDs for transactions on a centralized market are not observable to the market, so users with wash trading intent have much lower incentives to hide self-self trades. Second, the volume-matching algorithm requires each involved party in a trading "circle" to end up with the same market position before and after the sequence of trades. This requirement may be restrictive, however, if the involved trader IDs all belong to the same entity so that they have other means to balance each individual account off the exchange. For example, if the exchange itself controls multiple trader IDs, then these IDs will have little incentive to ensure the trades among themselves always maintain each ID's initial position. The next section thus continues to explore this possibility.

4.2 Insider-based wash trading detection

Even though our data do not allow us to exactly match all trader IDs to their corresponding real persons or entities behind, we are nevertheless able to do so for a subset of them. Specifically, we can associate a few trader IDs with Mt.Gox itself (with various levels of confidence) based on a few additional sources. To the extent that all trades among exchange-controlled IDs do not incur market risk to the exchange itself, these trades offer an additional source of identifying (some) wash trades.

We follow the spirit of Willy report and identify a sequence of (increasingly inclusive) sets of suspicious insider IDs (with decreasing confidence). Appendix B provides details on the identification process. In summary, we consider an increasingly inclusive sequence of possible insiders: (1) Known Insiders, (2) Hijackers, (3) Double Users, (4) Zero-Fee Traders, and (5) All Wash Traders.[31] Table 8 Panel A characterizes these various categories, and for each category summarizes the number of

[31]Known insiders are our most restrictive, but also most confident exchange insider designations. They include a trader ID nicknamed MagicalTux (a pseudo-name of Mark Karpelés, owner and CEO of Mt.Gox), as well as trader ID #1, the first registered trader on Mt.Gox.

unique trader IDs in it, the number of wash trader IDs in it, the number of transactions among traders in this category as well as its percentage out of all transactions during the June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013 period, and the amount of bitcoins traded among traders in this category as well as its percentage out of all transactions in the June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013 period.

[Table 8 about here.]

Using the various candidate insider definitions in Table 8 Panel A, we augment the wash trades identified in Section 2 with transactions among insiders. To the extent that the insider IDs are controlled by Mt.Gox itself, the augmented set will further include what are commonly known as painting-the-tape trades and may more comprehensively reflect the extent to which wash trading prevails. Specifically, with increasing inclusiveness and decreasing confidence, we define various groups of (augmented) wash trades: Group 1 (G1) is the set union of all wash trades and all transactions among Known Insiders. G2 to G5 are similarly defined: Group 2 (G2) is the union of all wash trades and all transactions among Known Insiders and Hijackers; Group 3 (G3) is the union of all wash trades and all transactions among Known Insiders, Hijackers, and Double-Users; Group 4 (G4) is the union of all wash trades and all transactions among Known Insiders, Hijackers, Double-Users, and Zero-Fee Traders; and Group 5 (G5) is the union of all wash trades and all transactions among Known Insiders, Hijackers, Double-Users, Zero-Fee Traders, and All Wash Traders. For ease of exposition, we also refer to the set of all (self-self) wash trades as G0.

Table 8 Panel B then compares the various groups of augmented wash trading transactions over several dimensions, including the number of transactions counts in each set and the percentage share out of all transactions, as well as the amount of bitcoins traded within each set and the percentage share out of all transactions. Overall, we find that the insider-based approach does detect more "hidden" wash trades. Depending on how aggressively we assign insiders, we get different levels of estimated hidden wash trades. That said, because of the fuzzy identification of exchange insiders, we caution readers that this part of the analysis is less conclusive than our main results in previous sections (which is especially true for G5, our most liberally defined set of wash trades).

5 Impact of wash trading

While this paper's main contribution is in presenting direct evidence of wash trading and evaluating various indirect estimation techniques, in this section we nevertheless provide additional results from leveraging our direct evidence to explore a follow-up question about what impact wash trading has on the cryptocurrency market. Specifically, we associate the intensity of wash trading with subsequent market conditions and document several stylized facts. Since these variables may all be endogenous and feature complicated joint dynamics, we adopt a vector autoregression (VAR) analysis framework.[32] Specifically, we estimate the following model:

$$Y_t = \sum_{i=1}^{4} A_i Y_{t-i} + \epsilon_t,$$

where A_i is a 5×5 matrix to be estimated, Y_t is a 5×1 vector corresponding to 1) the log wash trading volume (in units of BTC), 2) the log fee revenues collected from non-wash transactions (in units of BTC), 3) the log return of Bitcoin prices, 4) the liquidity of the Bitcoin market as measured by the Amihud measure, and 5) the realized volatility of Bitcoin return. The first two time series (wash trading volume and fee revenues from non-wash trades) are added by one before entering log transformations to avoid taking the log of zero; The latter two time-series (Amihud measure and realized volatility) are calculated within 30-minutes rolling windows. All ratios (return and Amihud) are winsorized at the 1% level.[33] We confirm that all time series are stationary: they contain no time trends, and significantly reject the unit root hypothesis by the Augmented Dickey-Fuller test.

[Figure 8 about here.]

Figure 8 plots the impulse response functions. Overall, we find a significantly positive response in (non-wash) fee revenues following a positive impulse of wash trading of about 1.5%, and the impulse response persists for the next 60 minutes. On the other hand, the responses from the

[32] The VAR approach has been widely used in financial market research related to the impact of trades on market conditions (e.g. Hasbrouck (1991)).

[33] Two-way winsorization for returns, which could take both positive and negative values, and one-way winsorization for the Amihud measure, which is always non-negative.

Amihud measure (of illiquidity) and Bitcoin return are much milder, and are order-of-magnitude smaller than that of fee revenues.

Remarks We continue to offer an economic analysis of the links between wash trading and other variables of interest uncovered in Figure 8. Before discussing these results, however, we caution that the usual caveat for VAR interpretations applies here: Since we do not have exogenous sources of variation, the relations presented in Figure 8 imply no causal claims.

On the one hand, because wash trades only inflate trading volume by clearing a simultaneous buy and sell order, it is perhaps not surprising that wash trading has mild responses in subsequent returns. This finding also reflects the difference between wash trading as studied in this paper from price manipulations as studied by Gandal, Hamrick, Moore and Oberman (2018). Although Gandal, Hamrick, Moore and Oberman (2018) use the same Mt. Gox data as we do, the two studies indeed cover different time periods: As we have explained earlier in the paper, wash trading stopped in May 2013, and it was only several months after May 2013 did the dramatic price run-up studied in Gandal, Hamrick, Moore and Oberman (2018) — which have been attributed to rampant price manipulations — begin.

On the other hand, the significantly positive responses in fee revenues following impulses in wash trading could be consistent with several plausible theories.[34] Our discussions below thus take an even-handed approach and explore their relative merits and limitations:

First, the positive fee response is consistent with suspicions in the Bitcoin community that exchanges may pump up volumes to look more attractive to traders and thus boost fee collections. That said, a hasty conclusion may face multiple challenges. For example, if the exchange indeed fakes volume to boost fees, it may be puzzling to observe that the response to wash trading shocks is so immediate (hits the peak in about 3 minutes) and quickly phases out (within 60 minutes) — as this finding stands an apparent contrast with that in Amiram, Lyandres and Rabetti (2020), who suggest evidence for long term effect of wash trading. We view this first challenge addressable given that a likely reason behind this difference is that our sample and that in Amiram, Lyandres

[34] It is worth noting that the fees in Figure 8 only count those collected from non-wash trades (that is, fees collected from wash trades are excluded), therefore the connection between wash trading and subsequent increase in fee revenues is not mechanically due to clustering of wash trades over time.

and Rabetti (2020) live in two dramatically different market environments — while Mt.Gox was the dominant if not the only Bitcoin exchange during our sample period, there are hundreds of competing exchanges during the sample period of Amiram, Lyandres and Rabetti (2020). In a competitive environment, exchanges can hike up rankings among peer exchanges and thus look more attractive to *prospective* customers when they inflate apparent trading volumes. This channel, however, is nonexistent in the almost monopolistic market environment that Mt.Gox lived in. Indeed, the higher fees following wash trading impulses in our sample are almost exclusively coming from existing investors — if we drop fee revenues from new investors, the impulse response will see no discernible changes. That existing investors drive fee changes also explains why the response in fees to wash impulse is immediate. Regarding why the responses tend to be short-lived, we point to a likely reason in that the existing investors lured by wash trades may quickly realize that the market is actually not that liquid: As Figure 8 has also shown, although liquidity sees an immediate but mild improvement after a wash trading shock (as the Amihud measure slightly decreases), it quickly reverts back to normal at a similar pace with how the response of fee revenues fades.

That said, other challenges remain if one investigates the trade-offs involved in wash trading: On the one hand, if the positive responses in fees following wash trading impulses indeed have causal interpretations, then we can quantify the overall economic benefit of committing wash trading to be in the magnitude of five thousand bitcoins, which is a significant amount.[35] On the other hand, there does not seem to be any obvious cost to wash trading for an exchange insider — while reputation concerns are natural candidates, they do not seem likely given that Mt.Gox's internal book was not available to be the public,[36] so wash trades may not be easily detected.[37] This tension thus throws doubt on a causal interpretation of wash trading enhancing fee revenues: If there is indeed a benefit to wash trading from fee revenue increases but no disciplinary mechanism against it, then why the exchange does not increase wash trades as much as it can?

[35] Given an average log wash trading volume per minute of 0.017 BTC, about 1 million minutes in our sample, and a cumulative impulse response of 0.14 of log non-wash bitcoin fee to log wash trading volume, the total amount of extra (both bitcoin and money) fee collected is in the order of magnitude of $2 \times 0.14 \times 0.017 \times 10^6 = 4,760$ BTCs (which is about $200K given the average bitcoin price of $40 during our sample).

[36] At least it was not intended to be public until hackers doxxed it upon Mt.Gox's collapse, which enabled our study as well as several others in the literature as we have reviewed earlier.

[37] As far as we know, the first public awareness of potential wash trading did not appear until Ribes (2018) when the intensity of wash trading was arguably much higher than during our sample period.

The challenge above thus suggests the potential of alternative explanations behind the positive responses in fees following impulses in wash trading. One possibility is that there is actually no causal relationship between wash trading and subsequent increase in fee revenues, rather both are just reactions to some other market variables, say increased volatility. Indeed, both wash trading and fee revenues feature positive responses to a positive impulse in volatility, and both impulses are persistent as they last beyond 60 minutes.[38] Another possibility points to potential mistakes in trading bots' algorithms: Given that wash trades tend to come from more frequent traders (who are likely bots), it is possible that their algorithms occasionally submit erroneous orders that eventually cross with oneself.[39]

Ultimately, due to the lack of exogenous shocks, one cannot conclusively argue whether the observed positive impact on fee revenues by wash trades is necessarily causal. Furthermore, it is possible that multiple economic channels are simultaneously at play behind the observations, so it is difficult to indisputably pinpoint a single motivation behind wash trades. Nevertheless, we hope the evidence in this section may still offer useful information and help readers make updates based on their individual priors.

6 Conclusion

The major contributions of this paper are two-fold: First, it provides first-hand direct evidence of wash trading in the cryptocurrency market. While the community has only in recent years become suspicious of such market manipulations, we find that this practice actually has a much longer history. In fact, wash trading has been taking place on the world's first Bitcoin exchange, Mt.Gox, back in the early days of cryptocurrency trading. From June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013, a relatively small set of (2,887 out of 98,391, or 2.93%) trader IDs in Mt.Gox were found to be involved in 115,275 wash trading transactions. These trader IDs include ones who have close ties with the exchange itself, and the number of transactions among them accounts for a disproportionately 32.8% of all transactions on Mt.Gox during the same period. Wash trading first

[38]Figure A7 in the appendix illustrates both impulse response functions.

[39]If such mistakes are the pure cause behind the observed wash trades, they may also explain the low level of wash trading identified in our sample.

emerges after a one-week halt of Mt.Gox's service in 2011, a time when it needs to rebuild liquidity, and stops in 2013 when regulators step in for an investigation.

Second, given that our data is so far the only individual-level information available to the community, our direct evidence gives us a unique opportunity to evaluate various alternative indirect estimation techniques proposed in the literature. In particular, our analysis lends support to indirect techniques based on Benford's law, trade clustering patterns at round sizes, conformity to log-normal distributions, and structural breaks, which have been used by Cong, Li, Tang and Yang (2020) and Amiram, Lyandres and Rabetti (2020) to evaluate the extent of wash trading in today's market and connections with exchange competitions, respectively. We also provide suggestions on how to most effectively apply some of these methods based on their performance on our data, point out that indirect estimation techniques based on tail distributions may not be reliable, and develop an index of wash trading based on simple heuristics and public information only to guide regulators locate periods with high wash trading probabilities.

We hope our first-hand direct evidence can shed new light on the community's suspicions over fake volume in cryptocurrency exchanges as well as our existing toolkit to detect wash trading from public information. Given the heated debate on how to appropriately regulate cryptocurrency exchanges, we hope our findings could help develop ways to safeguard the healthy development of the cryptocurrency market.

References

Ajinkya, Bipin B, and Prem C Jain. 1989. "The behavior of daily stock market trading volume." *Journal of Accounting and Economics*, 11(4): 331–359. 21

Alexander, Gordon J, and Mark A Peterson. 2007. "An analysis of trade-size clustering and its relation to stealth trading." *Journal of Financial Economics*, 84(2): 435–471. 3

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz. 2014. "powerlaw: a Python package for analysis of heavy-tailed distributions." *PloS one*, 9(1): e85777. 21

Amiram, Dan, Evgeny Lyandres, and Daniel Rabetti. 2020. "Competition and Product Quality: Fake Trading on Crypto Exchanges." *Available at SSRN 3745617.* 3, 13, 17, 21, 22, 23, 24, 34, 35, 37, 66, 81

Augustin, Patrick, Alexey Rubtsov, and Donghwa Shin. 2020. "The impact of derivatives on cash markets: Evidence from the introduction of bitcoin futures contracts." *Available at SSRN*. 8

Benetton, Matteo, and Giovanni Compiani. 2020. "Investors' Beliefs and Asset Prices: A Structural Model of Cryptocurrency Demand." *University of Chicago, Becker Friedman Institute for Economics Working Paper*, , (2020-107). 8

Benford, Frank. 1938. "The law of anomalous numbers." *Proceedings of the American philosophical society*, 551–572. 17

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, Catherine Casamatta, and Albert J Menkveld. 2018. "Equilibrium bitcoin pricing." *Available at SSRN 3261063*. 8

Bitwise. 2019. "Bitwise Asset Management Presentation to the U.S. Securities and Exchange Commission." 3, 13

Chatzigiannis, Panagiotis, Foteini Baldimtsi, Igor Griva, and Jiasun Li. 2022. "Diversification across mining pools: Optimal mining strategies under pow." *Journal of Cybersecurity*, 8(1): tyab027. 8

Choi, Kyoung Jin, Alfred Lehar, and Ryan Stauffer. 2020. "Bitcoin microstructure and the kimchi premium." *Available at SSRN 3189051*. 8

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. "Power-law distributions in empirical data." *SIAM review*, 51(4): 661–703. 21

Comerton-Forde, Carole, and Tālis J Putniņš. 2011. "Measuring closing price manipulation." *Journal of Financial Intermediation*, 20(2): 135–158. 5, 8, 25, 26

Cong, Lin William, G. Andrew Karolyi, Ke Tang, and Weiyi Zhao. 2022. "Value Premium, Network Adoption, and Factor Pricing of Crypto Assets." *working paper*. 8

Cong, Lin William, Wayne Landsman, Edward Maydew, and Daniel Rabetti. 2023. "Tax-Loss Harvesting with Cryptocurrencies." *Journal of Accounting and Economics, forthcoming*, 101607. 11

Cong, Lin William, Xi Li, Ke Tang, and Yang Yang. 2020. "Crypto Wash Trading." *working paper*. 3, 13, 17, 18, 19, 20, 21, 37, 54, 66, 72, 76, 81, 84

Cong, Lin William, Zhiguo He, and Jiasun Li. 2021. "Decentralized mining in centralized pools." *The Review of Financial Studies*, 34(3): 1191–1235. 8

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels. 2020. "Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability." *2020 IEEE Symposium on Security and Privacy (SP)*, 566–583. 2

Dhawan, Anirudh, and Tālis J Putniņš. 2021. "A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets." *Review of Finance, forthcoming*. 7

Dyhrberg, Anne Haubo, Sean Foley, and Jiri Svec. 2019. "When bigger is better: The impact of a tiny tick size on undercutting behavior." *Available at SSRN 3194932.* 7

Feder, Amir, Neil Gandal, J.T. Hamrick, and Tyler Moore. 2017. "The impact of DDoS and other security shocks on Bitcoin currency exchanges: evidence from Mt. Gox." *Journal of Cybersecurity*, 3, 2: 137–144. 10, 59

Foley, Sean, Jonathan R Karlsen, and Tālis J Putniņš. 2019. "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?" *The Review of Financial Studies*, 32(5): 1798–1853. 8

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and H Eugene Stanley. 2003. "A theory of power-law distributions in financial market fluctuations." *Nature*, 423(6937): 267–270. 20

Gandal, Neil, JT Hamrick, Tyler Moore, and Tali Oberman. 2018. "Price manipulation in the Bitcoin ecosystem." *Journal of Monetary Economics*, 95: 86–96. 4, 7, 34, 82

Griffin, John M, and Amin Shams. 2020. "Is Bitcoin really untethered?" *The Journal of Finance*, 75(4): 1913–1964. 7

Guo, Yang, Jiasun Li, Mei Luo, and Yintian Wang. 2022. "Illiquid bitcoin options." *Available at SSRN 4149934.* 8

Halaburda, Hanna, Zhiguo He, and Jiasun Li. 2021. "An economic model of consensus on distributed ledgers." National Bureau of Economic Research. 8

Hamrick, JT, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. 2018. "The economics of cryptocurrency pump and dump schemes." *CEPR Discussion Paper No. DP13404.* 7

Hasbrouck, Joel. 1991. "Measuring the information content of stock trades." *The Journal of Finance*, 46(1): 179–207. 33

He, Zhiguo, Jiasun Li, and Zhengxun Wu. 2023. "Don't Trust, Verify: The Case of Slashing from a Popular Ethereum Explorer." 1078–1084. 8

Hill, Bruce M. 1975. "A simple general approach to inference about the tail of a distribution." *The annals of statistics*, 1163–1174. 21

James, Nicholas A, Arun Kejariwal, and David S Matteson. 2016. "Leveraging cloud data to mitigate user experience from 'breaking bad'." 3499–3508, IEEE. 23, 24

James, Nicholas A, David S Matteson, et al. 2015. "ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data." *Journal of Statistical Software*, 62(i07). 23

Kuo, Wei-Yu, Tse-Chun Lin, and Jing Zhao. 2015. "Cognitive limitation and investment performance: Evidence from limit order clustering." *The Review of Financial Studies*, 28(3): 838–875. 60

Li, Jiasun. 2021. "DeFi as an Information Aggregator." 171–176, Springer. 8

Li, Jiasun. 2023. "On the security of optimistic blockchain mechanisms." *Available at SSRN 4499357.* 8

Li, Jiasun, and Guanxi Yi. 2019. "Toward a factor structure in crypto asset returns." *The Journal of Alternative Investments*, 21(4): 56–66. 8

Li, Jiasun, and William Mann. 2018. "Digital tokens and platform building." 8

Li, Jiasun, and William Mann. 2021. "Initial coin offerings: Current research and future directions." *The Palgrave Handbook of Technological Finance*, 369–393. 8

Li, Jiasun, and Zhengxun Wu. 2023. "Arbitrary Message Passing across Blockchains." *Available at SSRN.* 8

Li, Jiasun, Foteini Baldimtsi, Joao P Brandao, Maurice Kugler, Rafeh Hulays, Eric Showers, Zain Ali, and Joseph Chang. 2021. "Measuring illicit activity in DeFi: The case of Ethereum." 197–203, Springer. 8

Li, Tao, Donghwa Shin, and Baolian Wang. 2019. "Cryptocurrency pump-and-dump schemes." *Available at SSRN 3267041.* 7

Liu, Yukun, Aleh Tsyvinski, and Xi Wu. 2022. "Common risk factors in cryptocurrency." *The Journal of Finance*, 77(2): 1133–1177. 8

Liu, Yukun, and Aleh Tsyvinski. 2021. "Risks and returns of cryptocurrency." *The Review of Financial Studies*, 34(6): 2689–2727. 8

Makarov, Igor, and Antoinette Schoar. 2020. "Trading and arbitrage in cryptocurrency markets." *Journal of Financial Economics*, 135(2): 293–319. 8

Nilsson, Kim. 2014. "The Willy Report." *https://willyreport.wordpress.com/.* 4, 10, 82

Petryk, Mariia, and Jiasun Li. 2023. "(Data) Mining the Ethereum Open-Source Development Community." *Available at SSRN 4652260.* 8

Ribes, Sylvain. 2018. "Chasing Fake Volume: A Crypto-Plague." *https://medium.com/@sylvainartplayribes/chasing-fake-volume-a-crypto-plague-ea1a3c1e0b5e.* 3, 35

Richardson, Gordon, Stephan E Sefcik, and Rex Thompson. 1986. "A test of dividend irrelevance using volume reactions to a change in dividend policy." *Journal of Financial Economics*, 17(2): 313–333. 21

Rosch, Eleanor. 1975. "Cognitive reference points." *Cognitive psychology*, 7(4): 532–547. 18

Roy, Levin. 2018. "Mt Gox Hack Explained: Full History & Information Guide." *https://totalcrypto.io/mt-gox/.* 12

Schwenkler, Gustavo, and Hannan Zheng. 2021. "Peer Co-Movement in Crypto Markets." 8

Shams, Amin. 2020. "The structure of cryptocurrency returns." *Charles A. Dice Center Working Paper*, , (2020-11). 8

Tibeaudou, M. 2019. "Problems with the Bitwise Fake Volume Report." *https://hedgetrade.com/bitwise-exchange-report-raises-questions/*. 3

Underwood, BD. 2018. "Virtual Market Integrity Initiative Report." *Report, Office of the New York State Attorney General, Albany.* 2

Victor, Friedhelm, and Andrea Marie Weintraud. 2021. "Detecting and quantifying wash trading on decentralized cryptocurrency exchanges." 23–32. 29, 30, 48

Vigna, Paul, and Alexander Osipovic. 2018. "Bots Are Manipulating Price of Bitcoin in Wild West of Crypto." *Wall Street Journal*, 2. 15

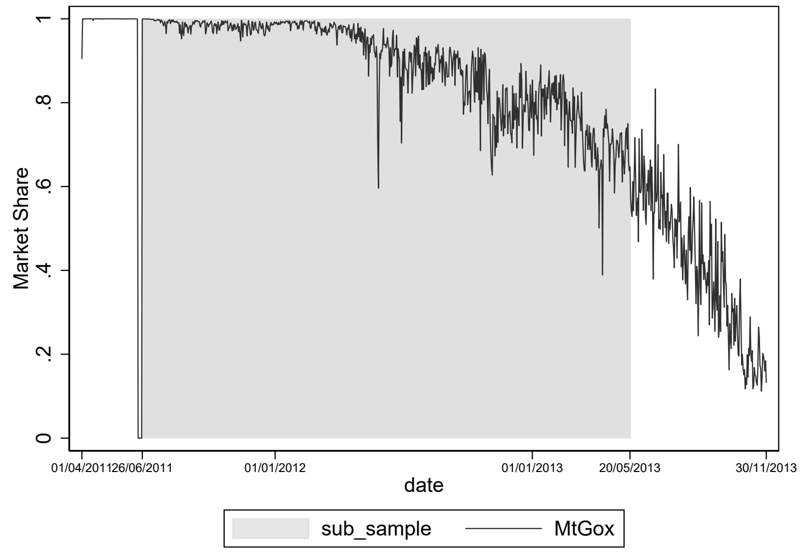
Xu, Jiahua, and Benjamin Livshits. 2019. "The anatomy of a cryptocurrency pump-and-dump scheme." *Proceedings of 28th USENIX Security Symposium*, 1609–1625. 7

Yu, Yang Gloria, and Jinyuan Zhang. 2020. "Flight to Bitcoin." *Jinyuan, Flight to Bitcoin (Oct 26, 2020).* 8

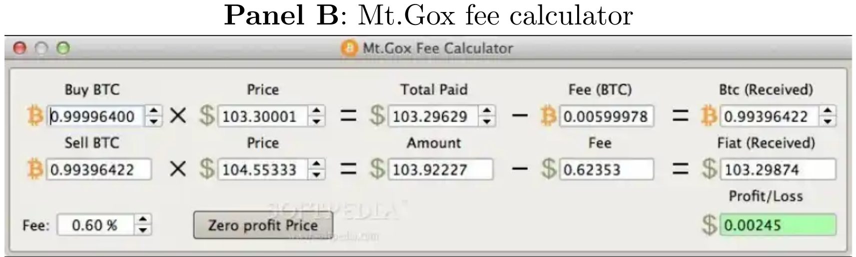
Figures

Figure 1: Background Information about Mt.Gox

Panel A: Mt.Gox's market share over time

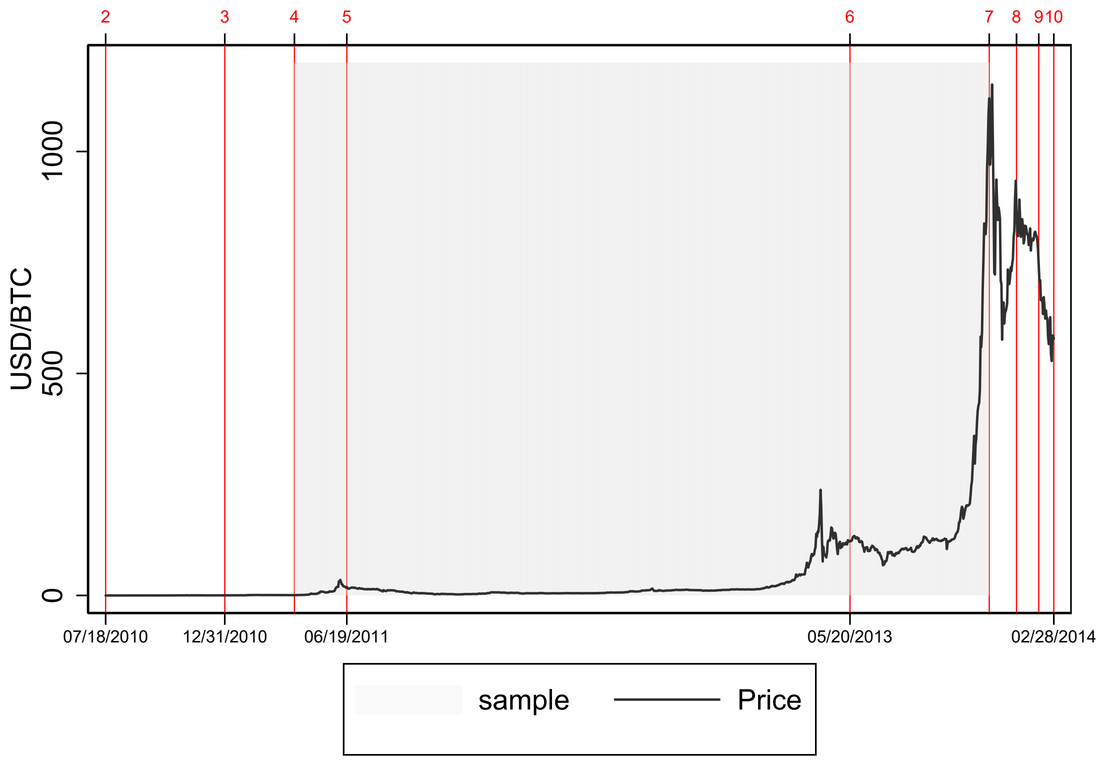


Panel B: Mt.Gox fee calculator



This figure provides background information about Mt.Gox. Panel A presents its market shares by trading volume over our entire sample period (April 1$^{\text{st}}$, 2011 to November 30$^{\text{th}}$, 2013). The shaded area corresponds to the sub-sample period with active wash trading (June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013). We notice that Mt.Gox is the largest, and for about half of our sample period, the monopolistic bitcoin exchange. Data for daily trading volume are obtained from Bitcoinity. However, following a June 2011 hack of Mt. Gox (as indicated by the sudden drop of market share to 0 right before the shaded sub-period), Bitcoinity reports missing volumes for many days, and we thus replace the Mt. Gox trading volume data from June 26$^{\text{th}}$, 2011 to November 29$^{\text{th}}$, 2011 using data from BitcoinCharts. Our calculations are consistent with third-party estimates, such as those reported in Medium. Panel B presents Mt.Gox's fee calculator, which illustrates bitcoin fees and fiat fees. Bitcoin fees (in units of BTC) and fiat fees (in units of fiat currencies) are commissions taken by the exchange from transactions. Only the bitcoin buyer (who pays fiat currencies in exchange for bitcoins) in a transaction pays the bitcoin fee; the bitcoin seller does not incur bitcoin fees. Similarly, only the bitcoin seller (who pays bitcoins in exchange for fiat currencies) in a transaction pays the fiat fee; the bitcoin buyer does not incur fiat fees. The bitcoin fee takes the form of a fixed percentage (typically 0.60%) of the number of bitcoins a buyer purchases, with a typical minimum of 1 satoshi (10^{-8} BTC); The fiat fee takes the form of a fixed percentage (typically 0.60%) of the amount of fiat currency one gets from selling bitcoins, with a typical minimum of ¥0.001.

Figure 2: Timeline and Historical Events

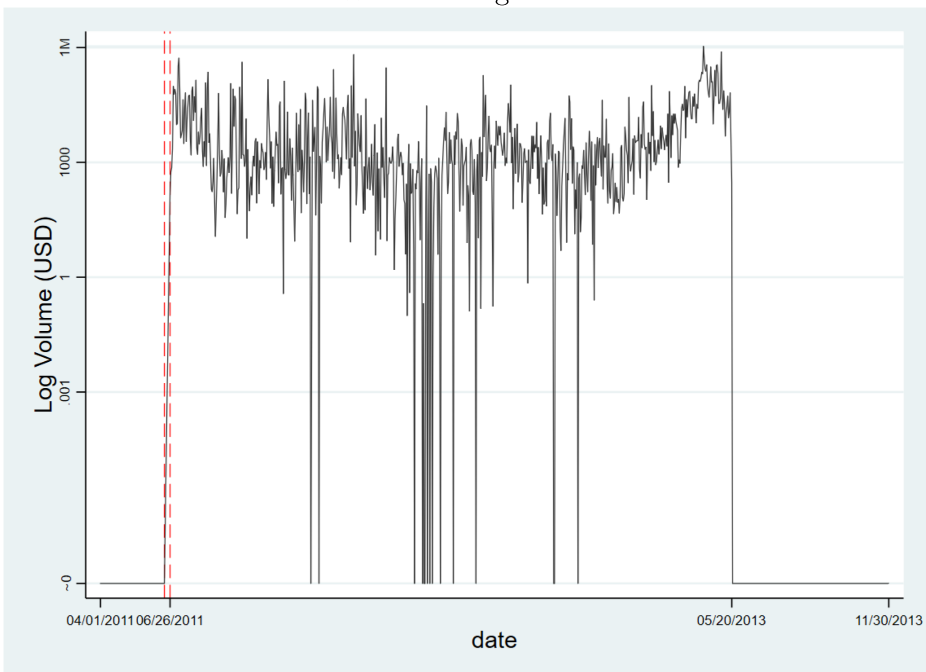


This figure plots the evolution of bitcoin prices during our sample period. The shaded area corresponds to a period during which wash trading was prevalent. The red vertical dashed lines mark several major historical events, including:

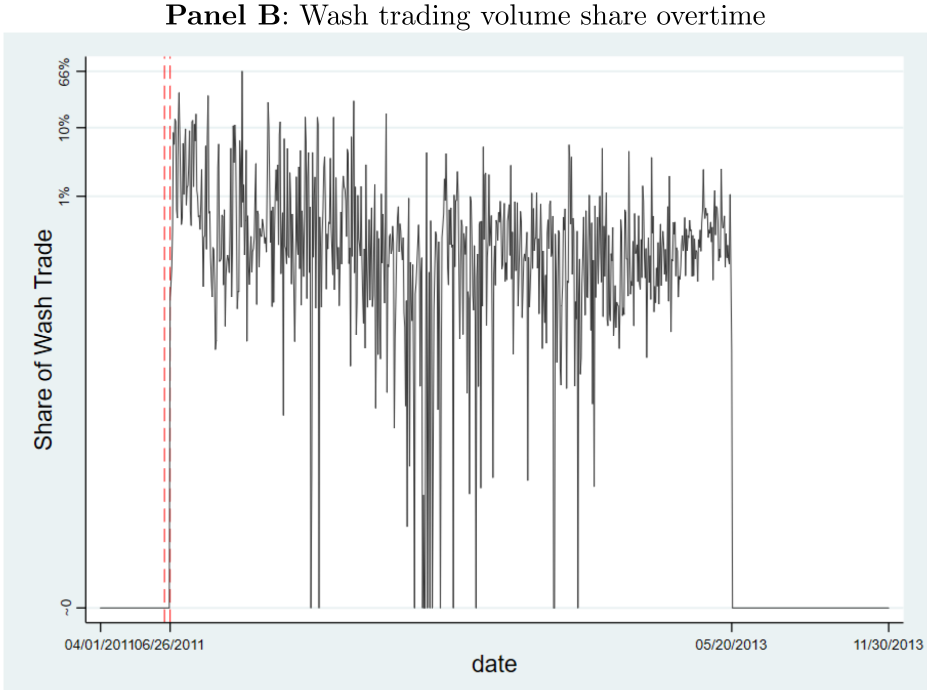
1. late 2007: Jed McCaleb founded "Magic: The Gathering Online Exchange" (Mt.Gox) as a trading venue for the card game "Magic: The Gathering".
2. July 18$^{\text{th}}$, 2010: Mt.Gox started quoting prices of Bitcoin. Popularity boomed.
3. *circa* the end of 2010: Mark Karpelés bought 88% share of Mt.Gox from McCaleb and revamped the website. McCaleb later went on to create the payment startup Ripple.
4. April 1$^{\text{st}}$, 2011: Sample starts.
5. June 19$^{\text{th}}$, 2011: Mt.Gox hacked; site offline for about a week (further illustrated in Figure A3).
6. May 20$^{\text{th}}$, 2013: Coinlab sued Mt.Gox claiming \$75M over a non-materialized deal on the exchange's US-based customers, followed by the seizure of around \$5M from the company's bank accounts by U.S. Department of Homeland Security in the investigation.
7. November 30$^{\text{th}}$, 2013: end of the sample.
8. January 7$^{\text{th}}$, 2014: the presence of trading bot Willy was confirmed on Mt.Gox during a site glitch.
9. February 7$^{\text{th}}$, 2014: Mt.Gox froze all Bitcoin withdrawals.
10. February 28$^{\text{th}}$, 2014: Mt.Gox filed for bankruptcy, with 744,408 bitcoins belonging to customers and 100,000 to the company "lost".

Figure 3: Wash Trading Overtime

Panel A: Wash trading volume overtime

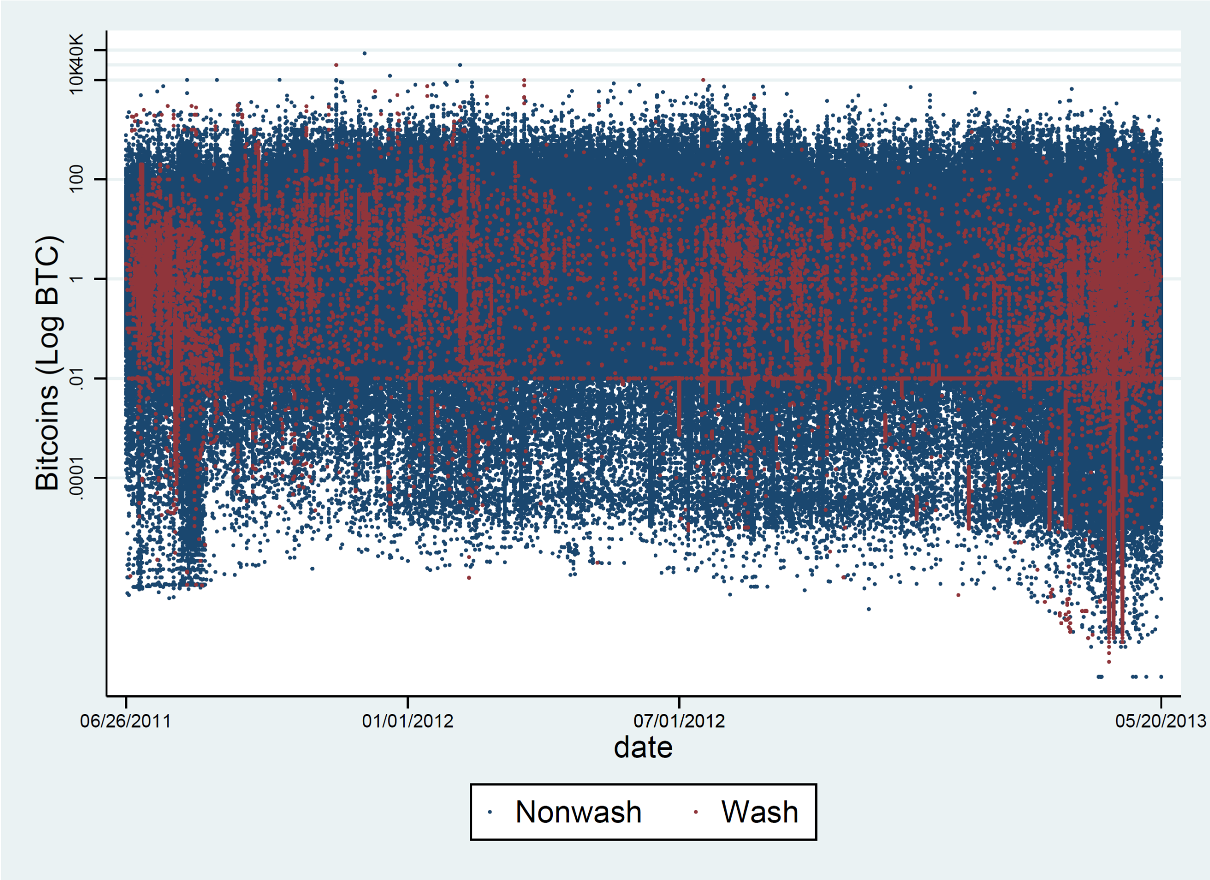


Panel B: Wash trading volume share overtime



This figure illustrates the presence of wash trades. Panel A plots the daily wash trading volume (in USD) on Mt.Gox, and Panel B plots the daily share of wash trading volumes within the total daily trading volumes on Mt.Gox. We see that wash trading first arises circa June 26$^{\text{th}}$, 2011 (after a technical glitch brought down Mt.Gox for about a week, which is marked by vertical dotted red lines). Wash trading had then been prevalent until May 20, 2013, when Mt.Gox faced an investigation by the regulator.

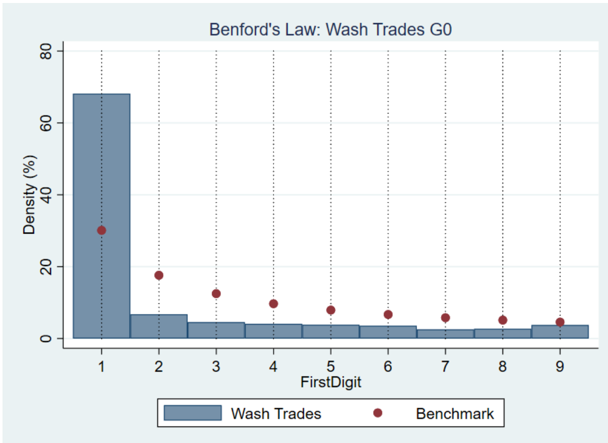
Figure 4: Wash versus Non-Wash Trading Patterns

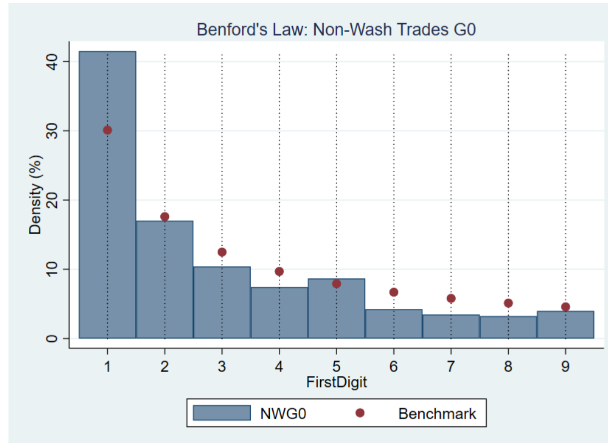


This figure visualizes the trade size distribution over time for both wash trades and non-wash trades, respectively. The horizontal axis is time and the vertical axis is trade size in the logarithm scale. We use blue dots to indicate non-wash trades and red dots to indicate wash trades. Wash trading started on June 26^{th}, 2011 (one week after the attack on Mt Gox on June 19^{th}, 2011) and ended on May 20^{th}, 2013 (five days after being issued a warrant by the US Department of Homeland Security (DHS) to seize money from Mt.Gox's U.S. subsidiary's account with payment processor Dwolla.

Figure 5: Fits from Indirect Estimation Techniques

Panel A: Testing indirect estimation by Benford's law

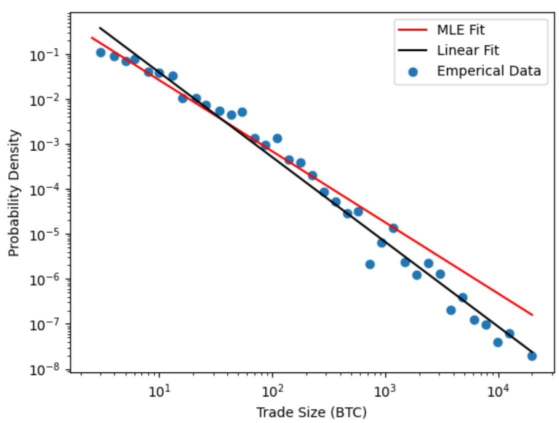


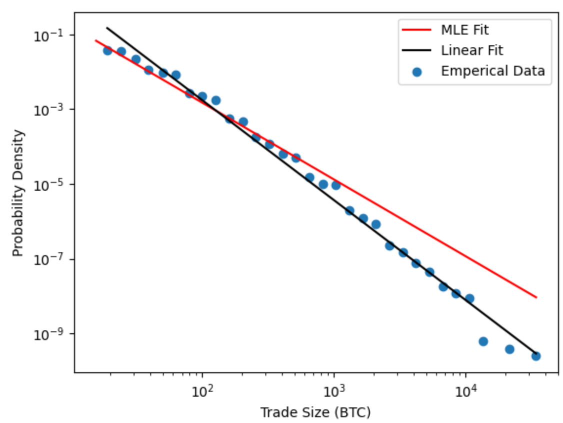


Wash trades:
χ^2 statistics: 69.71; p-value: 0.000

Non-wash trades:
χ^2 statistics: 7.91; p-value: 0.443

Panel B: Tail distribution and power-law fitting



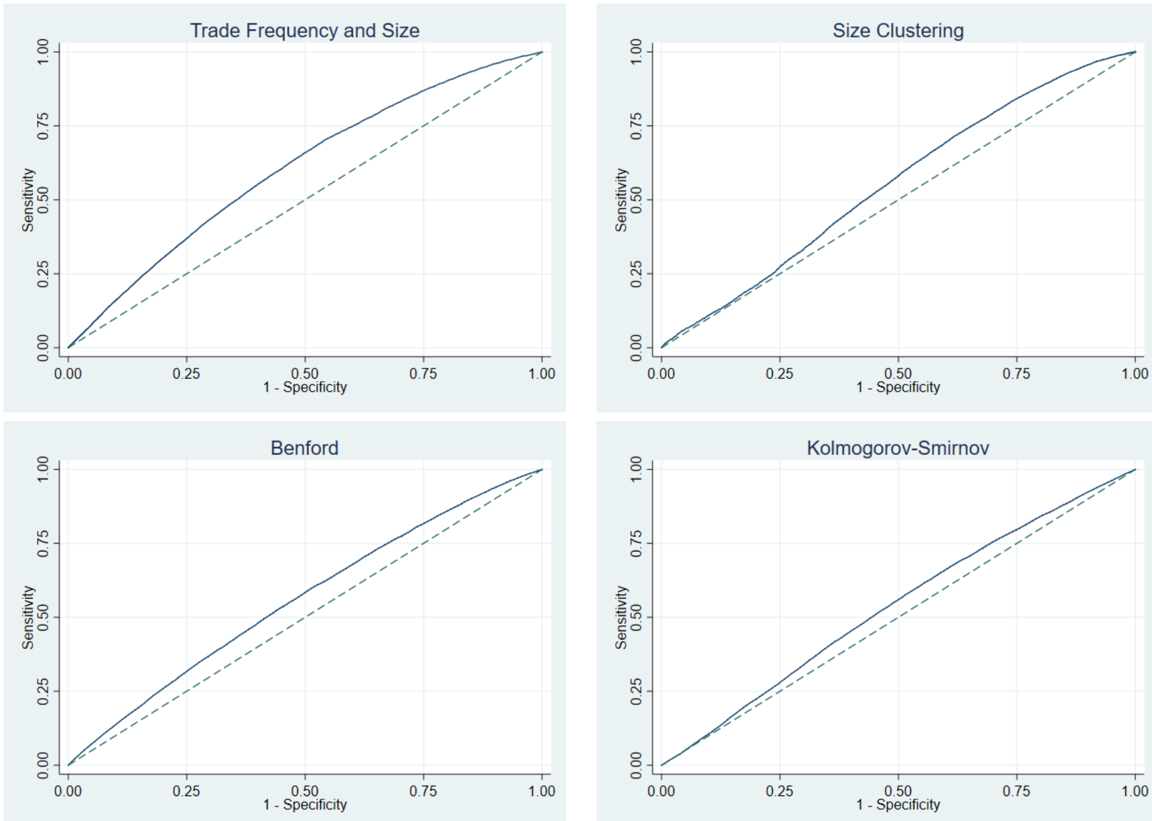


Wash trades:
$\hat{\alpha}_{OLS} = 1.883$; $\hat{\alpha}_{Hill} = 1.580$

Non-wash trades:
$\hat{\alpha}_{OLS} = 2.674$; $\hat{\alpha}_{Hill} = 2.054$

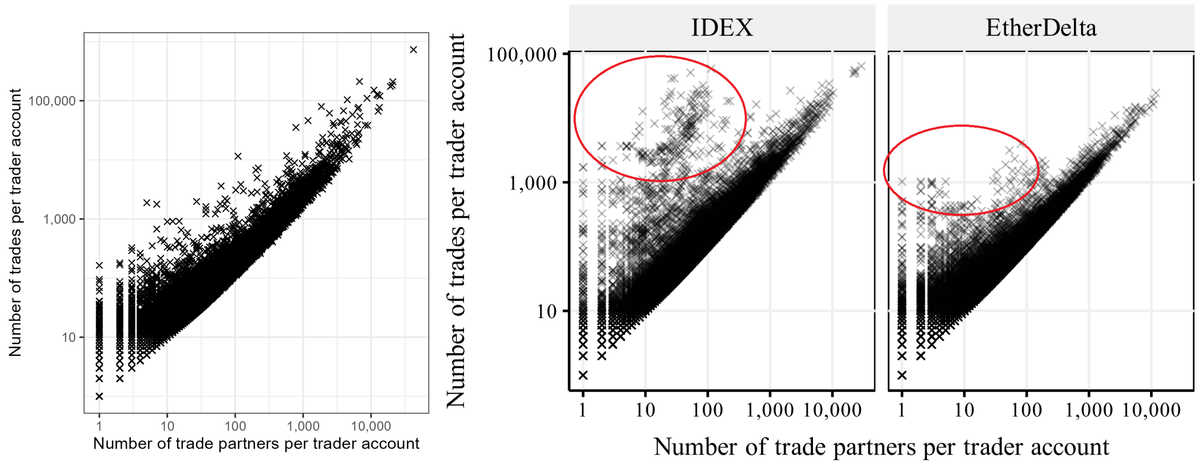
This figure illustrates the fit of two indirect estimation techniques. Panel A plots the distribution of the first significant digits of trade sizes (blue bars), for wash and non-wash trades, respectively. We compare the empirical distributions to theoretical benchmarks suggested by Benford's law (red dots) and both the plots and ξ^2 statistics suggest that the distribution of wash trades significantly deviates from the Benford law, while it does not for non-wash trades. Our direct evidence thus supports the effectiveness of Benford's law for wash trading inference. Panel B plots the tails of trade-size distributions and the fitted power-law lines on log-log plots, for wash and non-wash trades, respectively. Fitted power-law lines are plotted with parameters estimated by Ordinary Least Square (OLS) and Maximum Likelihood Estimation (MLE), shown in black and red lines, respectively. Blue dots represent empirical data points for trade-size frequencies. The sample is between June 26$^{\text{th}}$, 2011 and May 20, 2013 in both panels.

Figure 6: Comparing Indirect Estimation Techniques Using ROC Curves



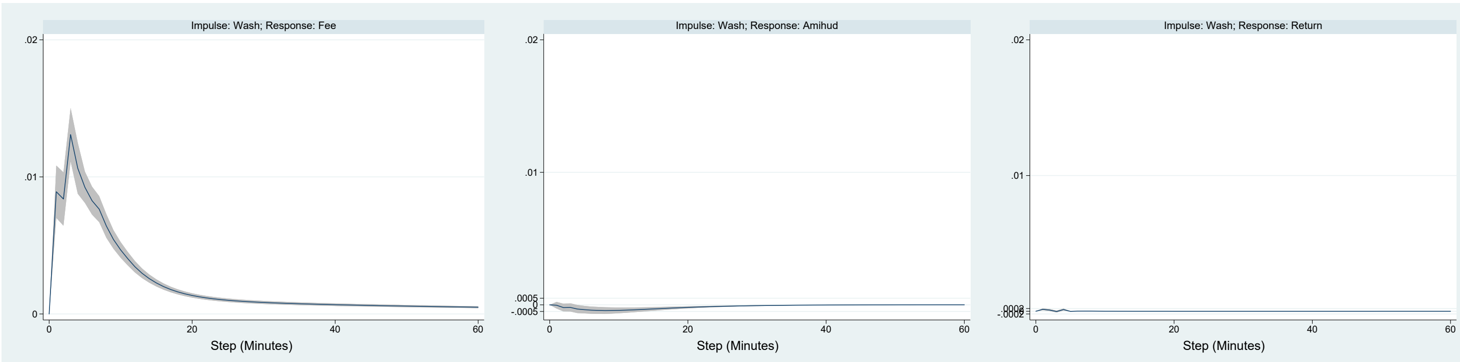
These figures plot the Receiver Operating Characteristic (ROC) Curves from applying various indirect methods for labeling whether a ten-min trading window contains wash trades or not. The upper-left panel presents the ROC curve from our wash trading index, which is built upon the observation that periods with wash trading tend to see anomalously high numbers of trades and low median trade sizes. Specifically, the index at window t looks into the whole day ending one hour before t and counts the number of 10-min windows within the whole day that have both fewer numbers of trades than that in window t and larger median trade sizes than that in window t. The higher the index in window t is, the more likely that it contains wash trades. The other panels then put the classification accuracy of our index in context by further illustrating the performance of classification methods based on the various indirect estimation techniques discussed earlier. For example, the upper-right panel presents the ROC curve from the percentage of unrounded trades — that is, we obtain the percentage of unrounded trades for each ten-min window, count the number of 10-min windows within the whole day ending one hour before t that have smaller percentages of unrounded trades than that in window t to obtain anomalous percentages of unrounded trades, and then present the ROC curve. A similar procedure applies to the other panels, with the percentage of unrounded trades in each window replaced by the Benford χ^2-statistics (the bottom-left panel), the Kolmogorov-Smirnov (KS) distance between log trade sizes and normal distributions with the same mean and variance (the bottom-right panel). All curves are significantly above the 45° line, suggesting that they are all effective for classifying time windows into ones with wash trading and ones without.

Figure 7: Lack of Trading Partner Concentration



This figure visualizes the potential concentration of a trader's trading partners. In the left panel, each cross represents a trader ID in our Mt.Gox sample, positioned by its number of trades (horizontal axis) and the number of unique trading counterparties (vertical axis). For comparison, we also include Figure 2 in Victor and Weintraud (2021) on the right, which conducts the same analysis but on decentralized exchange IDEX and EtherDelta. We highlight in red circles that the two decentralized exchanges have many traders that perform a large number of trades with only a few other counterparties, while we note that the centralized exchange Mt.Gox does not show similar patterns.

Figure 8: Impulse Response Functions



These figures plot impulse responses for the following VAR analysis: $Y_t = \sum_{i=1}^{4} A_i Y_{t-i} + \epsilon_t$, where A_i is a 5×5 matrix to be estimated, Y_t is a 5×1 vector corresponding to 1) the log wash trading volume (in units of BTC), 2) the log fee revenues collected from non-wash transactions (in units of BTC), 3) the log return of Bitcoin prices, 4) the liquidity of the Bitcoin market as measured by the Amihud measure, and 5) the realized volatility of Bitcoin return. The first two time series (wash trading volume and fee revenues from non-wash trades) are added by one before entering log transformations to avoid taking the log of zero; The latter two time series (Amihud measure and realized volatility) are calculated within 30-minute rolling windows. All ratios (return and Amihud) are winsorized at the 1% level, with two-way winsorization for returns, which could take both positive and negative values, and one-way winsorization for the Amihud measure, which is always non-negative. Shaded areas represent 95% confidence intervals. There is a significantly positive response in (non-wash) fee revenues of about 1.5% following a positive impulse of wash trading, which persists for the next 60 minutes. In comparison, the responses from the Bitcoin return and Amihud measure (of illiquidity) are negligible.

Tables

Table 1: Summary Statistics for All Transactions

Panel A: Full Sample: April 1, 2011 to November 30, 2013

	Mean (1)	Min (2)	1^{st} Qu. (3)	Median (4)	3^{rd} Qu. (5)	Max (6)	Std.dev (7)
Order size (BTC)	6.65	0.00	0.04	0.50	3.00	34274.76	44.57
Price (JPY)	10833.08	0.00	570.46	4071.52	12089.74	150000.00	19852.68
Price (USD trades only; $)	103.16	0.50	6.51	25.93	119.63	1333.33	190.36
BTC fee (buy orders; %)	0.41%	0.00%	0.29%	0.43%	0.55%	100.00%	1.32%
Money fee (sell orders; %)	0.38%	0.00%	0.27%	0.43%	0.55%	100.00%	0.42%

Panel B: Sub-sample: June 26th 2011 ad May 20th 2013

	Mean	Min	1^{st} Qu.	Median	3^{rd} Qu.	Max	Std.dev
Order size (BTC)	7.84	0.00	0.05	0.57	3.99	34274.76	50.34
Price (JPY)	4039.30	0.00	436.46	952.14	8080.29	26379.39	5183.94
Price (USD trades only; $)	39.77	1.00	5.26	11.40	74.02	397.21	51.06
BTC fee (buy orders; %)	0.38%	0.00%	0.28%	0.40%	0.55%	100.00%	0.57%
Money fee (sell orders; %)	0.36%	0.00%	0.27%	0.40%	0.53%	100.00%	0.38%

This table presents summary statistics for the entire sample, which includes all transaction records on Mt.Gox from April 1^{st} 2011 to November 30^{th}, 2013, as well as the period with active wash trading, which includes all transaction records on Mt.Gox from June 26^{th} 2011 to May 20^{th}, 2013. The BTC and money fees contain outliers because of the presence of minimum bitcoin/money fees, in that those who trade a very small amount of bitcoin (e.g. below the minimum fee level) will end up incurring 100% fees. These incidents are rare though, as the 99% percentile of the BTC fee is 0.60%, which is in line with the rest of the spectrum.

Table 2: Direct Evidence of Wash Trading

	TX counts	BTC volume
	(1)	(2)
Wash trading (amount)	115,275	606,698
Wash trading (% out of total)	2.1%	1.4%
Trades among wash traders (amount)	1,821,366	14,135,295
Trades among wash traders (% out of total)	32.8%	32.5%
All trades	5,548,867	43,515,392

This table summarizes the direct evidence of wash trading in our sample. In the first two rows, we respectively present the absolute amount of wash trades (either in terms of the total number of transactions in column (1) and total bitcoin volume in column (2)) as well as their percentage shares out of all transactions. Wash trades are defined as trades with the same trader ID on both sides (buy/sell), and the sample period is from June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013 during which there is active wash trading on Mt.Gox. Our directly detected wash trades account for 2.1% of all transactions during the period, as well as 1.4% of the total trading volume. In the next three rows, we present auxiliary information to put the magnitude of wash trading in the first two rows in perspective. In rows three and four, we respectively present the total number of transactions (column (1)) and total bitcoin volume (column (2)) as well as their percentage shares out of all transactions, for trades among wash traders, that is, those trader IDs that have ever engaged in wash trading (i.e., IDs that show up in either side (buy or sell) of at least one wash trade). Such trades are not necessarily wash trades but are indicative of the behavior of those IDs ever involved in wash trading. These trades account for 32.8% of all transactions during the period, as well as 32.5% of the total trading volume. The fifth row presents the total number of transactions as well as the total BTC volume of all transactions within the sample period.

Table 3: Characterizing Wash trader ID's transactions

Panel A: Characterizing Wash Trader IDs' Trades with Self (Own ID)

	Mean (1)	Min (2)	1st Qu. (3)	Median (4)	3rd Qu. (5)	Max (6)	Std.dev (7)
# trades	39.93	1.00	1.00	2.00	6.00	61,213.00	1,147.40
Share (%)	3.35	0.00	0.33	1.05	2.88	49.72	6.57
Volume (BTC)	210.15	0.00	0.34	3.15	19.25	264,499.62	5,199.50
Share (%)	1.61	0.00	0.06	0.32	1.26	47.81	3.89
Volume (¥, 10^3)	238.10	0.00	0.64	6.12	41.25	98,047.58	2,563.15
Share (%)	1.55	0.00	0.05	0.28	1.16	47.88	3.80

Panel B: Characterizing Wash Trader IDs' Trades with Other Wash Trading IDs

	Mean	Min	1st Qu.	Median	3rd Qu.	Max	Std.dev
# trades	1,221.84	1.00	44.00	142.00	527.50	393,961.00	9,392.44
Share (%)	53.58	1.53	45.45	55.21	63.14	95.61	13.72
Volume (BTC)	9,582.01	0.00	86.13	529.72	2,862.77	2,029,205.33	63,500.46
Share (%)	50.78	0.03	42.73	52.68	60.97	99.81	15.53
Volume (¥, 10^3)	17,235.04	0.00	186.00	1,081.21	6,591.40	2,774,225.84	88,412.56
Share (%)	48.22	0.03	40.29	49.33	57.74	100.00	15.56

Panel C: Characterizing Wash Trader IDs' Trades with Regular (Non-Wash Trading) IDs

	Mean	Min	1st Qu.	Median	3rd Qu.	Max	Std.dev
# trades	912.06	0.00	35.00	117.00	412.50	277,184.00	6,576.07
Share (%)	43.08	0.00	33.56	41.84	52.06	97.27	14.72
Volume (BTC)	7,185.17	0.00	88.97	484.00	2,480.40	1,091,445.77	43,268.45
Share (%)	47.61	0.00	37.74	46.01	56.00	99.94	15.76
Volume (¥, 10^3)	19,227.07	0.00	187.67	1,176.18	7,344.01	3,152,531.75	100,015.25
Share (%)	50.23	0.00	41.01	49.49	58.54	99.94	15.92

This table presents summary statistics across the wash traders population regarding their number of trades (percentage within all of one's trades), volumes in BTC (percentage within one's total BTC trading volume), and volumes in ¥ (percentage within one's total ¥ trading volume) attributed to transactions with different counterparties: (1) oneself (in Panel A), (2) other Wash Trading IDs (i.e. traders with many self-self trades, in Panel B), and (3) Regular IDs (i.e. traders that have never engaged in self-self trades, in Panel C). The sample period is from June 26th 2011 to May 20th, 2013.

Table 4: Market Variables Associated with Wash Traders' Transactions

Panel A: Comparison of Prevailing Market Conditions for Different Transactions

Wash Traders' Transactions with		NonWash Volume (1)	Illiquidity (Amihud) (2)	Return (3)	Volatility (4)
		Extensive Margin			
Self	Mean	0.091	0.021	-0.001	0.010
	S.D.	0.387	0.039	0.015	0.005
Other Wash Traders	Mean	0.229	0.014	0.000	0.009
	S.D.	0.892	0.030	0.012	0.006
Non-Wash Traders	Mean	0.216	0.015	0.000	0.009
	S.D.	0.824	0.031	0.012	0.007
Wash vs. Non-Wash Trades	Diff.	-0.130	0.006	-0.001	0.001
	t.stat	-51.50	66.95	-29.83	45.51
		Intensive Margin			
Small Wash Trades	Mean	0.043	0.026	-0.009	0.011
	S.D.	0.206	0.042	0.025	0.006
Large Wash Trades	Mean	0.099	0.020	0.000	0.010
	S.D.	0.407	0.038	0.013	0.005
Small vs. Large Wash Trades	Diff.	-0.056	0.006	-0.009	0.002
	t.stat	-16.55	18.14	-65.41	36.74

Panel B: Logit Regression of Wash Trade on Prevailing Market Conditions

	Intercept	NonWash Volume	Illiquidity (Amihud)	Return	Volatility
Coefficient	-3.92***	-0.871***	3.500***	-5.949***	12.221***
S. E.	0.005	0.014	0.070	0.208	0.341

Panel A explores market conditions under which wash traders choose to conduct wash trades versus normal (non-wash) trades. Specifically, we decompose all wash trader's transactions into three groups: self-self wash trades, trades with other wash traders, and trades with non-wash trades. Note that the first group constitutes wash trades according to our definition, while the second and third groups together constitute non-wash trades. Within each group, we calculate the mean and standard deviation of various variables of market conditions, measured by the previous minute's Bitcoin return, realized volatility, market illiquidity as measured by the Amihud measure, and genuine (non-wash) trading volume, respectively. To improve readability, the Amihud measure takes the ratio between the absolute value of the return and volume multiplied by 1,000, and the non-wash trading volume is divided by 1,000. Note that if more than one type of trade happens at the same time, then the same previous-minute variables of market conditions would contribute to the mean and standard deviation of each group. To give a more comprehensive picture, we further decompose self-self wash trades into large and small ones (whose size is above or below the median wash trades sizes). Panel B formalizes the comparison in Panel A in the following logit regression: For each transaction by a wash trader, we regress a dummy of whether it is a wash trade (1) or not (0) on the previous minute market conditions, ranging from Bitcoin return, realized volatility, market illiquidity as measured by the Amihud measure, and genuine (non-wash) trading volume, respectively. The regression coefficients remain stable after the sixth iteration. Statistical significance at levels of 1%, 5%, and 10% are denoted with ***, **, and *, respectively.

Table 5: Student's t-Tests for Trade-Size Clustering

trade size (BTC)	Round size unit (BTC)	Wash Trade		Non-Wash Trade	
		Difference (1)	t-stat (2)	Difference (3)	t-stat (4)
(0, Max]	100×10^{-1}	-0.349	-7.129	-0.296	-11.711
(0, Max]	100×10^{-2}	-0.379	-15.564	-0.245	-20.769
(0, Max]	100×10^{-3}	-0.502	-41.477	-0.351	-64.892
(0, Max]	100×10^{-4}	-0.523	-69.634	-0.376	-139.943
(0, Max]	100×10^{-5}	-0.650	-142.638	-0.401	-283.843
(0, Max]	100×10^{-6}	-0.698	-203.682	-0.331	-380.130
Panel II					
(0, Max]	500×10^{-1}	-0.358	-9.054	-0.316	-16.105
(0, Max]	500×10^{-2}	-0.426	-22.260	-0.271	-29.247
(0, Max]	500×10^{-3}	-0.517	-50.912	-0.387	-90.035
(0, Max]	500×10^{-4}	-0.574	-91.550	-0.388	-175.494
(0, Max]	500×10^{-5}	-0.677	-166.753	-0.418	-358.877
(0, Max]	500×10^{-6}	-0.694	-211.838	-0.340	-442.001
Panel III					
(50, 500]	100×10^{-1}	-0.081	-1.642	0.113***	3.243
(5, 50]	100×10^{-2}	0.041	1.422	0.230***	7.995
(0.5, 5]	100×10^{-3}	0.001	0.031	0.128***	3.804
(0.05, 0.5]	100×10^{-4}	0.064***	2.906	0.206***	5.923
(0.005, 0.05]	100×10^{-5}	0.002	0.052	0.072**	2.186
(0.0005, 0.005]	100×10^{-6}	-0.055	-4.192	0.128***	6.203
Panel IV					
(25, 250]	500×10^{-2}	0.014	0.377	0.129***	4.071
(2.5, 25]	500×10^{-3}	0.036	1.438	0.197***	5.465
(0.25, 2.5]	500×10^{-4}	-0.002	-0.099	0.129***	4.377
(0.025, 0.25]	500×10^{-5}	0.042	1.617	0.145***	3.738
(0.0025, 0.025]	500×10^{-6}	-0.015	-0.453	0.039	1.279
(0.00025, 0.0025]	500×10^{-7}	-0.093	-4.610	0.092***	4.136

This table examines whether frequencies of trades at round sizes are higher than those at unrounded sizes within the wash trading subsample and the non-wash trading subsample, respectively. In Panels I and II, we follow Cong, Li, Tang and Yang (2020) and carry out two sets of t-tests with different rounded size and surrounding window definitions: Multiples of 100 units with a window radius 50 (100X-50, 100X+50), and multiples of 500 units with a window radius 100 (500X-100, 500X+100), respectively, where one unit ranges from $\times10^{-6}$ to $\times10^{-1}$ BTC. Specifically, for each rounded size (e.g. 200×10^{-4} BTC) and its nearby window (e.g. between $(200-50)\times10^{-4}$ and $(200+50)\times10^{-4}$ BTC, excluding 200×10^{-4} BTC), we compare the frequency of those trades at the rounded size versus the highest frequency among all other non-rounded trades within the window. The differences between the rounded and the highest unrounded frequencies are then carried out over all rounded sizes, and the t-test results are reported with ***, **, and * denoting *positive* differences with statistical significance at levels of 1%, 5%, and 10%, respectively (A positive difference indicates that frequencies at round sizes are higher than the rest within the observation window, suggesting trade-size clustering). Panels III and IV further decompose all trades into subsamples of different sizes and report results from similar tests using different round-size units and surrounding window sizes (e.g. in Panel III a rounded trade with size within $(50, 500]$ BTC has a size in multiples of its corresponding round size unit, i.e. $100\text{X}\times10^{-1}$ and a window of $((100\text{X-}50)\times10^{-1}, (100\text{X}+50)\times10^{-1}]$; In Panel IV a rounded trade with size within $(25, 250]$ BTC has a size of $500\text{X}\times10^{-2}$ and a window of $((500\text{X-}100)\times10^{-2}, (500\text{X}+100)\times10^{-2}]$.

Table 6: Deviations from log-normal distributions

	Volume			#Trades		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0.139	0.051	0.088***	0.223	0.035	0.187***
SD/t	(0.071)	(0.020)	(5.894)	(0.071)	(0.010)	(12.873)

This table illustrates the respective extent to which trading volumes of wash trades and non-wash trades conform to log-normal distributions. For every full month in our sample (July 2011 to May 2013), within either the wash trading sample or the non-wash trading sample, we first aggregate trading volume in ten-minute bins, and then calculate the Kolmogorov-Smirnov (KS) distances between the empirical cumulative density function of their natural logarithms and the cumulative distribution function of a normal distribution with the same mean and variance. In the left three columns, we report the mean of the monthly KS distances for the wash trading sample and non-wash trading sample, as well as their difference. Standard deviations (for the mean statistics) and t-statistics (for the difference) are reported below in parenthesis. The right three columns repeat the exercise, with trading volume replaced by the number of trades.

Table 7: Structural breaks

	Volume (log)			#Trades (log)		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	7.318	0.000	7.318***	21.955	1.182	20.773***
SD	(6.869)	(0.000)	(4.997)	(28.429)	(2.130)	(3.418)
	Volume			#Trades		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0.182	0.000	0.182	0.136	0.000	0.136
SD	(0.664)	(0.000)	(1.283)	(0.640)	(0.000)	(1.000)

This table illustrates the respective extent to which trading volumes of wash trades and non-wash trades contain structural breaks. For every full month in our sample (July 2011 to May 2013), within either the wash trading sample or the non-wash trading sample, we use the time series of the log trading volume, the log number of trades, the (non-transformed) trading volume, and the (non-transformed) number of trades aggregated into ten-minute bins to calculate the number of structural breaks. We then report the mean of the monthly number of structural breaks for the wash trading sample and non-wash trading sample, as well as their difference. Standard deviations (for the mean statistics) and t-statistics (for the difference) are reported below in parenthesis.

Table 8: Insider-Based Detection of "Hidden" Wash Trades

Panel A: IDs Suspected to Be Associated with Exchange Insiders

Insider Type	#IDs	#Wash-trader IDs	#In-group TXs	(%) All	BTC sum	(%) All
	(1)	(2)	(3)	(4)	(5)	(6)
Known Insiders	3	2	211	0.00%	189,084	0.43%
Double-Users	71	7	66,572	1.20%	55,826	0.13%
Hijackers	325	10	66,870	1.21%	139,522	0.32%
Zero-Fee Traders	1,032	5	3,060	0.06%	53,621	0.12%
All Wash Traders	2,887	2,887	1,821,366	32.82%	14,135,295	32.48%

Panel B: Suspected Wash Trades Associated with Exchange Insiders

	TX counts	(%)	BTC sum	(%)
	(1)	(2)	(3)	(4)
G0	115,275	2.08%	606,698	1.39%
G1	115,383	2.08%	706,308	1.62%
G2	147,207	2.65%	887,332	2.04%
G3	147,481	2.66%	894,256	2.06%
G4	153,967	2.77%	963,824	2.21%
G5	1,904,178	34.32%	14,963,660	34.39%

This table augments our wash trading sample by identifying exchange insiders. Panel A compares the variously-defined categories of likely exchange insiders. Each column (from left to right) reports the number of IDs within each category, the number of IDs within each category that have ever engaged in wash trading, the number of (both direct and indirect wash) transactions within each category (as well as the percentage within all transactions in the June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013 subperiod), and the number of bitcoins traded within each category (as well as the percentage within all transactions in the June 26$^{\text{th}}$, 2011 to May 20$^{\text{th}}$, 2013 subperiod). Panel B then compares the variously-defined sets (G1-G5) of wash trades and painting-the-tape trades by exchange insiders. G1 is the set union of all wash trades and all transactions among Known Insiders. G2 is the set union of all wash trades and all transactions among Known Insiders and Double-Users; G3 is the set union of all wash trades and all transactions among Known Insiders, Double-Users, and Hijackers; G4 is the set union of all wash trades and all transactions among Known Insiders, Double-Users, Hijackers, and Zero-Fee Traders; and G5 is the set union of all wash trades and all transactions among Known Insiders, Double-Users, Hijackers, Zero-Fee Traders, and All Wash Traders.

Appendix

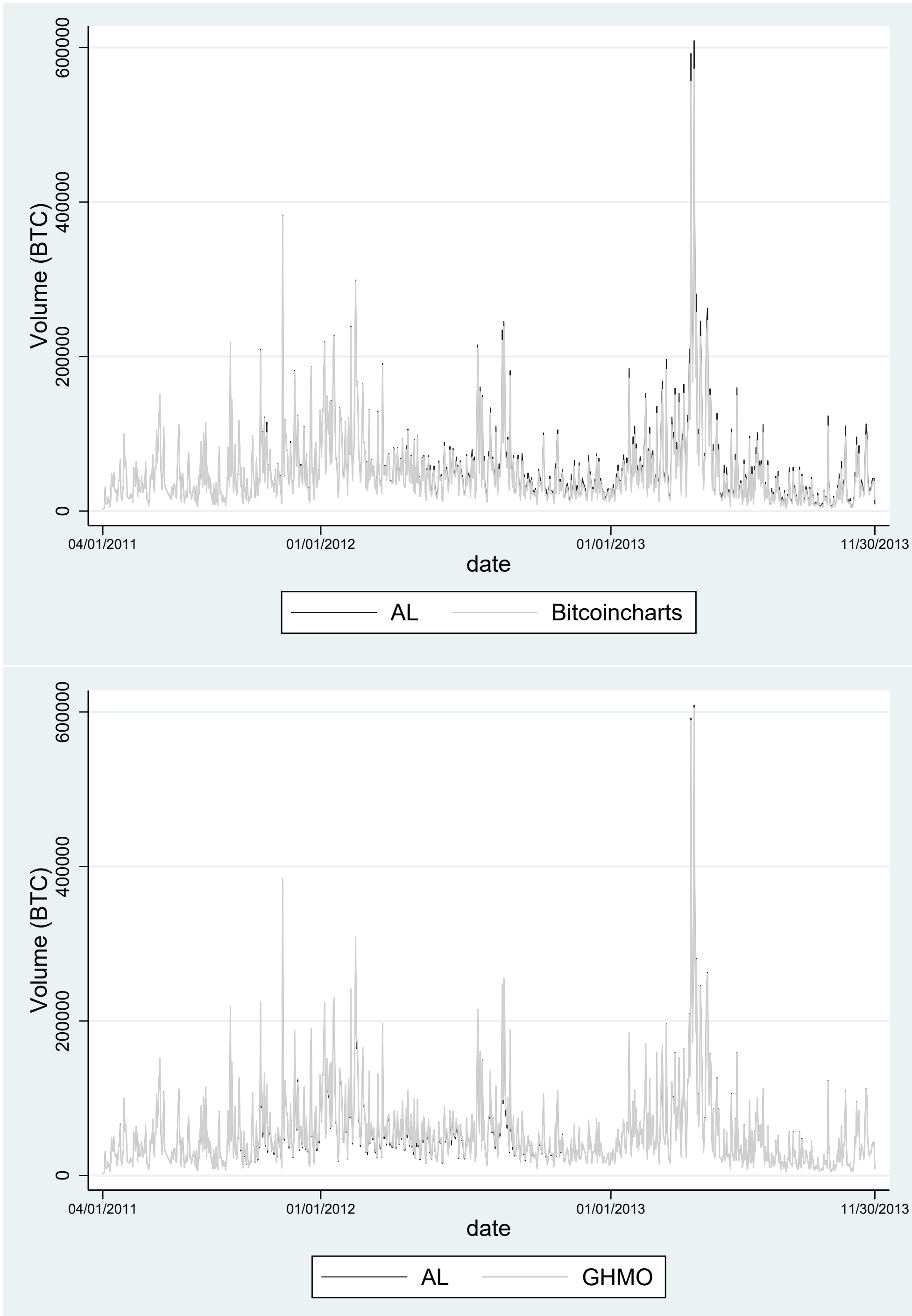
A Additional figures and tables

Figure A1: Snapshots of Original Files

Trade_Id	Date	User_Id	Japan	Type	Currency	Bitcoins	Money	Money_Rate	Money_JPY	Money_Fee	Money_Fee_Rate	Money_Fee_JPY	Bitcoin_Fee	Bitcoin_Fee_JPY
218861	19/06/2011 17:06	36272	NJP	sell	USD	0.122	2.155	80.385423966363	173.23058864751	0.0062	80.385423966363	0.49838962859145	0	0
218862	19/06/2011 17:06	30430	NJP	buy	USD	1	17.6	80.385423966363	1414.783461808	0	80.385423966363	0	0.006	7.692634402593
218862	19/06/2011 17:06	36272	NJP	sell	USD	1	17.6	80.385423966363	1414.783461808	0.114	80.385423966363	9.1639383321654	0	0
218863	19/06/2011 17:11	20311	NJP	buy	USD	1.06	18.56	80.385423966363	1491.9534688157	0	80.385423966363	0	0.006	7.692634402593
218863	19/06/2011 17:11	45295	NJP	sell	USD	1.06	18.56	80.385423966363	1491.9534688157	0.1206	80.385423966363	9.6944821303434	0	0
218864	19/06/2011 17:12	20311	NJP	buy	USD	0.522	9.14	80.385423966363	734.72277505256	0	80.385423966363	0	0.003	3.8463172012965
218864	19/06/2011 17:12	52686	NJP	sell	USD	0.522	9.14	80.385423966363	734.72277505256	0.05922	80.385423966363	4.760424807288	0	0
218865	19/06/2011 17:12	20311	NJP	buy	USD	1.03	18.035	80.385423966363	1449.7511212334	0	80.385423966363	0	0.006	7.692634402593
218865	19/06/2011 17:12	52686	NJP	sell	USD	1.03	18.035	80.385423966363	1449.7511212334	0.1173	80.385423966363	9.4292102312544	0	0
218866	19/06/2011 17:12	20311	NJP	buy	USD	0.008	0.14	80.385423966363	11.253959355291	0	80.385423966363	0	0	0
218866	19/06/2011 17:12	52686	NJP	sell	USD	0.008	0.14	80.385423966363	11.253959355291	8.0E-5	80.385423966363	0.006430833917309	0	0
218867	19/06/2011 17:13	20311	NJP	buy	USD	0.49	8.579	80.385423966363	689.62655220743	0	80.385423966363	0	0.003	3.8463172012965
218867	19/06/2011 17:13	35975	NJP	sell	USD	0.49	8.579	80.385423966363	689.62655220743	0.0559	80.385423966363	4.4935451997197	0	0
218868	19/06/2011 17:13	20311	NJP	buy	USD	2.02	35.37	80.385423966363	2843.2324456903	0	80.385423966363	0	0.013	16.667374538952
218868	19/06/2011 17:13	61019	NJP	sell	USD	2.02	35.37	80.385423966363	2843.2324456903	0.2292	80.385423966363	18.42433917309	0	0
1309108565842630	26/06/2011 17:16	634	JP	buy	USD	2	2	80.210970464135	160.42194092827	0	80.210970464135	0	0	0
1309108565842630	26/06/2011 17:16	634	JP	sell	USD	2	2	80.210970464135	160.42194092827	0	80.210970464135	0	0	0
1309109641289950	26/06/2011 17:34	634	JP	buy	USD	2	2	80.210970464135	2549.6814574718	0	80.210970464135	0	0	0
1309109641289950	26/06/2011 17:34	634	JP	sell	USD	2	2	80.210970464135	2549.6814574718	0	80.210970464135	0	0	0
1309110181198230	26/06/2011 17:43	634	JP	buy	USD	1	17.51001	80.210970464135	1404.4948949367	0	80.210970464135	0	0	0
1309110181198230	26/06/2011 17:43	634	JP	sell	USD	1	17.51001	80.210970464135	1404.4948949367	0	80.210970464135	0	0	0
1309110484425780	26/06/2011 17:48	7050	NJP	buy	USD	5.06992973	76.04895	80.210970464135	6099.9600822785	0	80.210970464135	0	0	0
1309110484425780	26/06/2011 17:48	53970	NJP	sell	USD	5.06992973	76.04895	80.210970464135	6099.9600822785	0	80.210970464135	0	0	0
1309110781131410	26/06/2011 17:53	60135	NJP	buy	USD	5	82.5	80.210970464135	6617.4050632911	0	80.210970464135	0	0	0
1309110781131410	26/06/2011 17:53	16131	JP	sell	USD	5	82.5	80.210970464135	6617.4050632911	0	80.210970464135	0	0	0
1309110960662670	26/06/2011 17:56	35825	NJP	buy	USD	45	742.5	80.210970464135	59556.64556962	0	80.210970464135	0	0	0
1309110960662670	26/06/2011 17:56	16131	JP	sell	USD	45	742.5	80.210970464135	59556.64556962	0	80.210970464135	0	0	0
1309110966830320	26/06/2011 17:56	35825	NJP	buy	USD	5	85	80.210970464135	6817.9324894515	0	80.210970464135	0	0	0

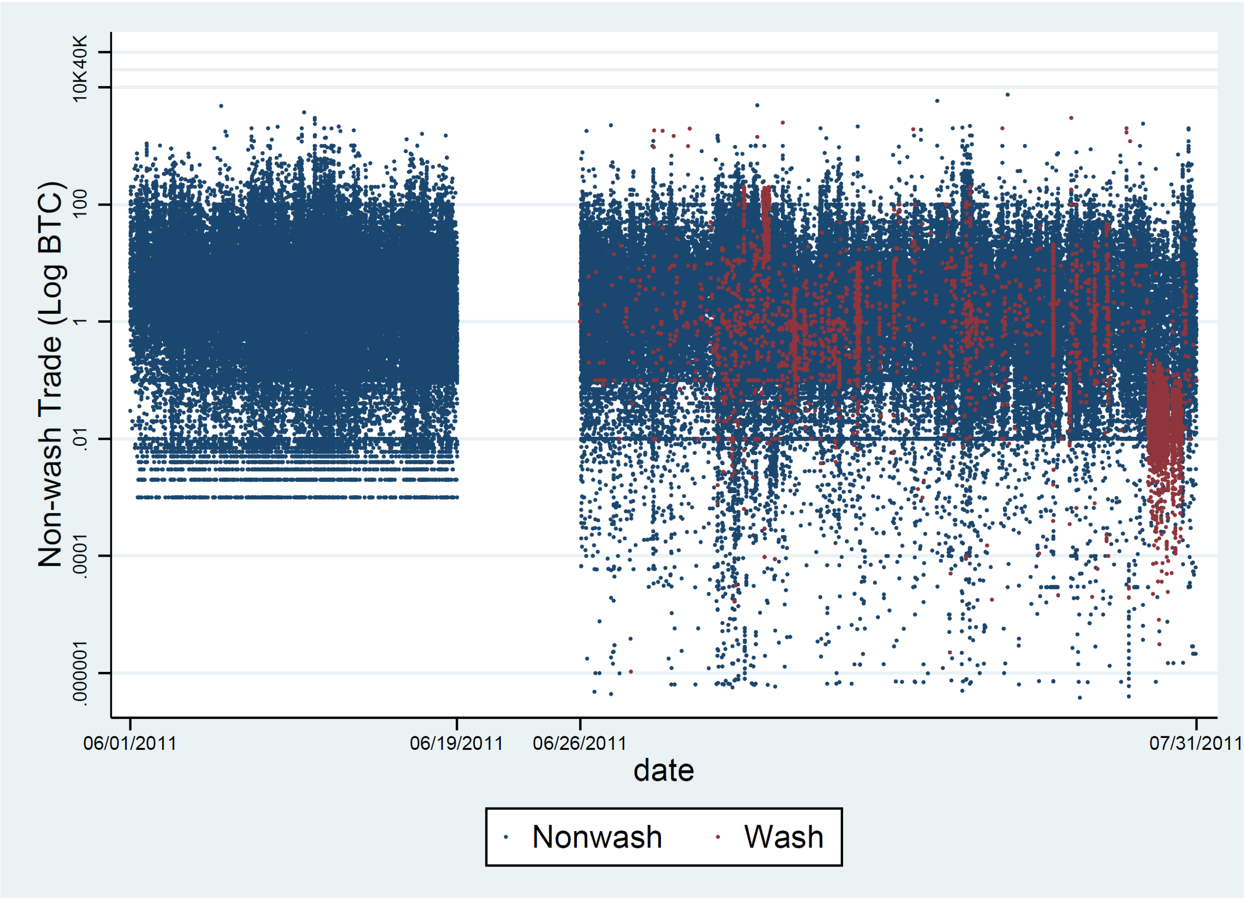
This figure presents a snapshot of the original data sorted by order sizes. The raw data contain information for each transaction a trade identifier, date and time, the user ID of the trader, whether the trade is in Japanese or not, whether it is a buy or sell leg, fiat currency involved, order size (in units of bitcoins and fiat currencies, as well as the converted value in Japanese Yen), the amount of fees paid (either in bitcoin or in fiat currency), country and location of the trader (for verified users only). We highlight several examples of wash trades among the largest transactions in red. These pairs of transactions share the same trade identifier and have the same user ID as both the buyer and seller at two legs of the same transaction.

Figure A2: Verification of the De-duplication Procedure



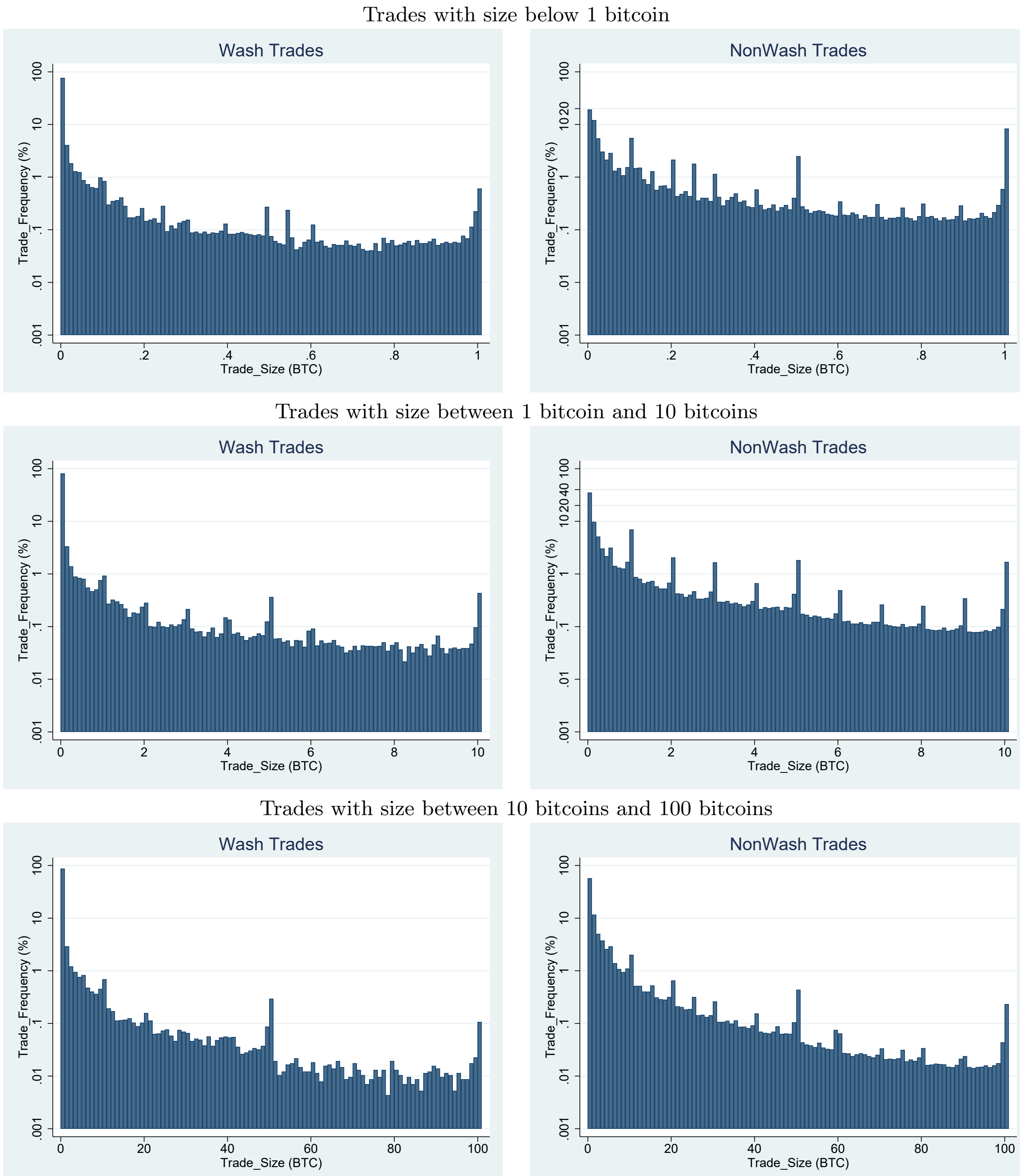
This figure compares the daily volume calculated from our de-duplication process (AL) with 1) that from Bitcoincharts in the top panel and 2) that from Feder, Gandal, Hamrick and Moore (2017) (FGHM) in the bottom panel, and confirms that our de-duplication process leads to volumes closely approximating those from external sources and FGHM.

Figure A3: The June 19th 2011 Hack on Mt.Gox that Halts Trading for a Week



This figure visualizes the order size distribution surrounding the June 19th 2011 hack. The horizontal axis is time and the vertical axis is order size in the logarithmic scale. We highlight wash trades in red while non-wash trades in blue to simultaneously present volume with/without wash trading and display the importance of wash behavior. Trading halts for about a week following the hack. It is after this hack that wash trading starts to emerge on Mt.Gox. Note that there is significant trade size clustering, which is also a well-recognized trader behavior in other markets (See e.g. Kuo, Lin and Zhao (2015)).

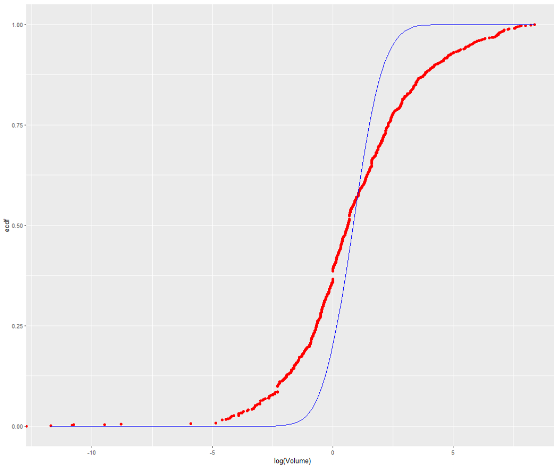
Figure A4: Trade-size Clustering



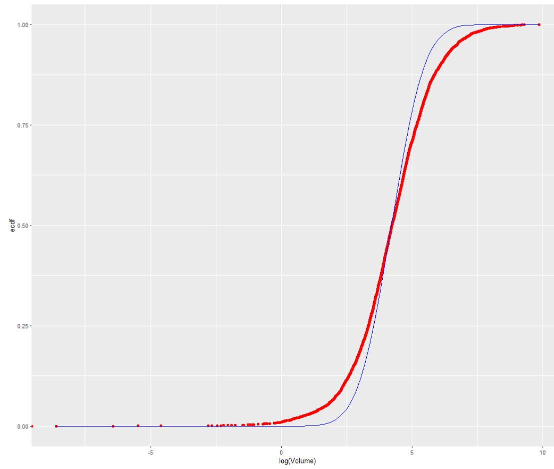
This figure plots the histograms of trade sizes for wash trades and non-wash trades, respectively, during our sample period between June 26$^{\text{th}}$, 2011 and May 20, 2013. We focus on three groups: trades with sizes below 1 bitcoin, ones with sizes between 1 bitcoin and 10 bitcoins, and ones with sizes between 10 bitcoins and 100 bitcoins, which together count for 99% of all transactions. Comparing the histograms of wash versus non-wash trades within each group, we find trade-size clustering around round numbers in both groups, although visually the clustering tends to be more salient among non-wash trades.

Figure A5: Deviations from log-normal distribution

Panel A: Trading volume

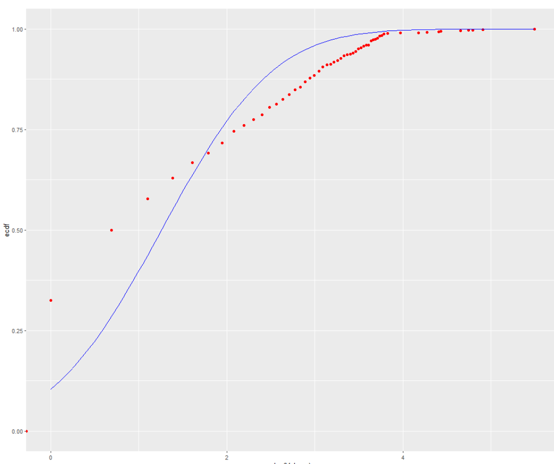


Wash trades

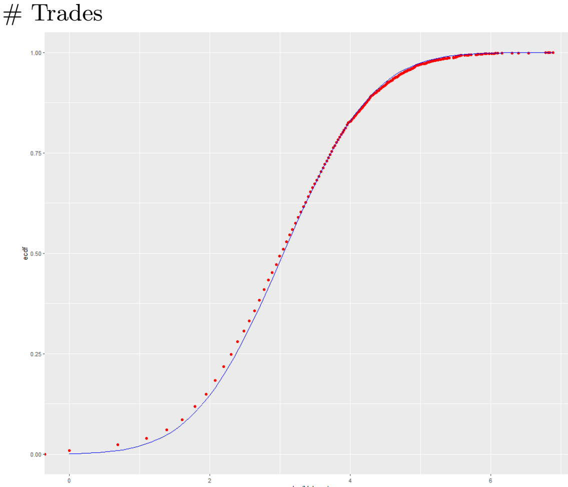


Non-Wash trades

Panel B: # Trades



Wash trades

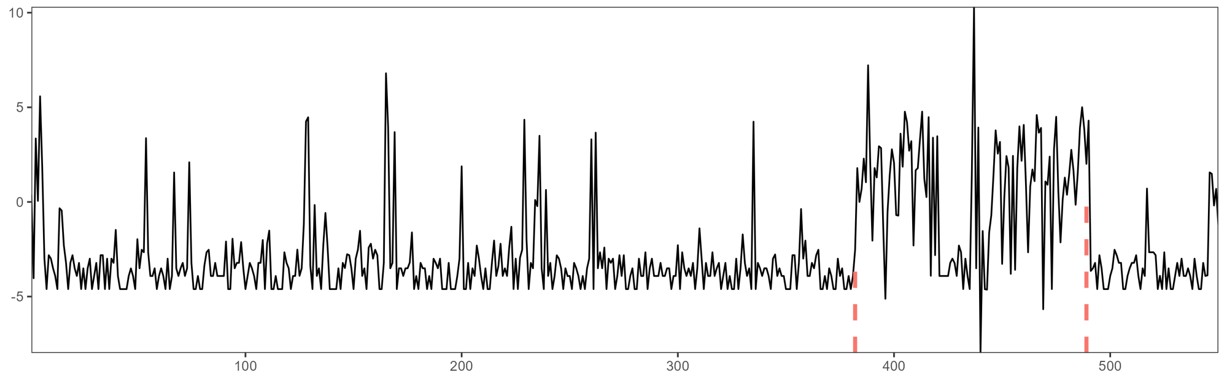


Non-Wash trades

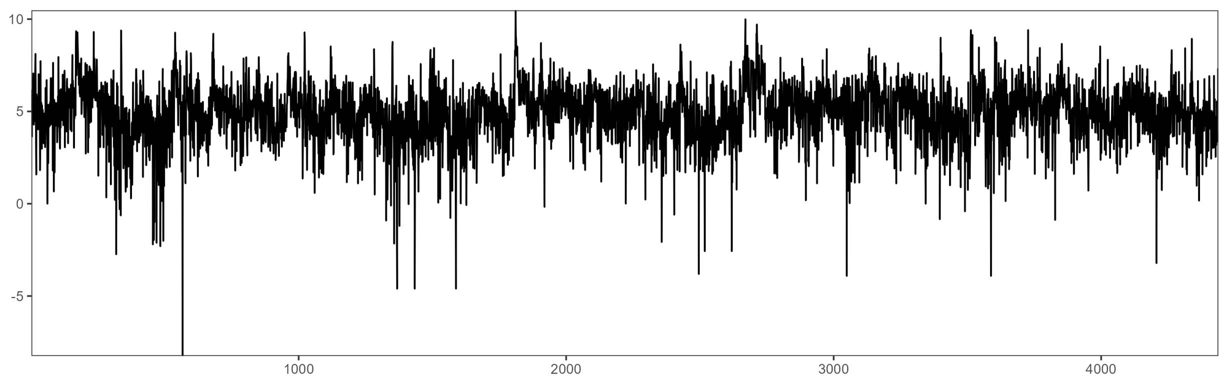
These figures illustrate the extent to which trading volumes and numbers of trades conform to lognormal distributions. We use the natural logarithm of trading volumes (in Panel A) and numbers of trades (in Panel B) aggregated into ten-minute bins in July 2011 as data, and plot the cumulative distribution function (c.d.f.) of the data (transformed by natural logarithm) in dotted red curves and the c.d.f. of a normal distribution with the same mean and variance in blue curves — for the respective samples of wash trades and non-wash trades.

Figure A6: Examples of the EDM measure

Panel A: Wash trades

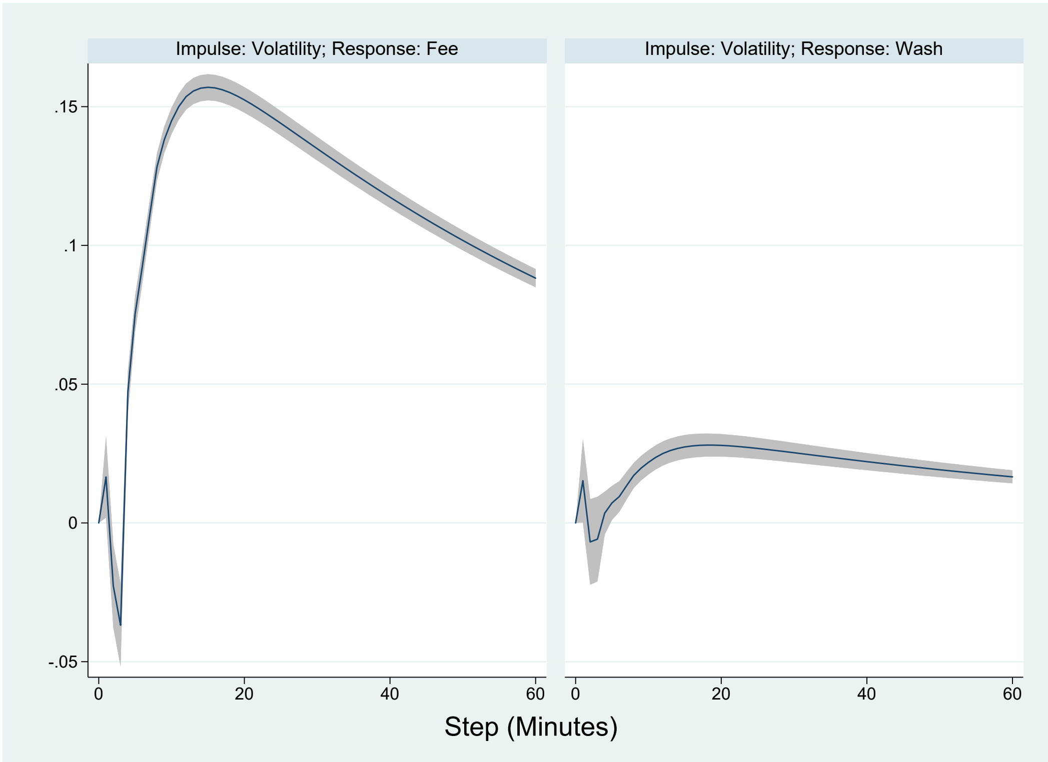


Panel B: Non-Wash trades



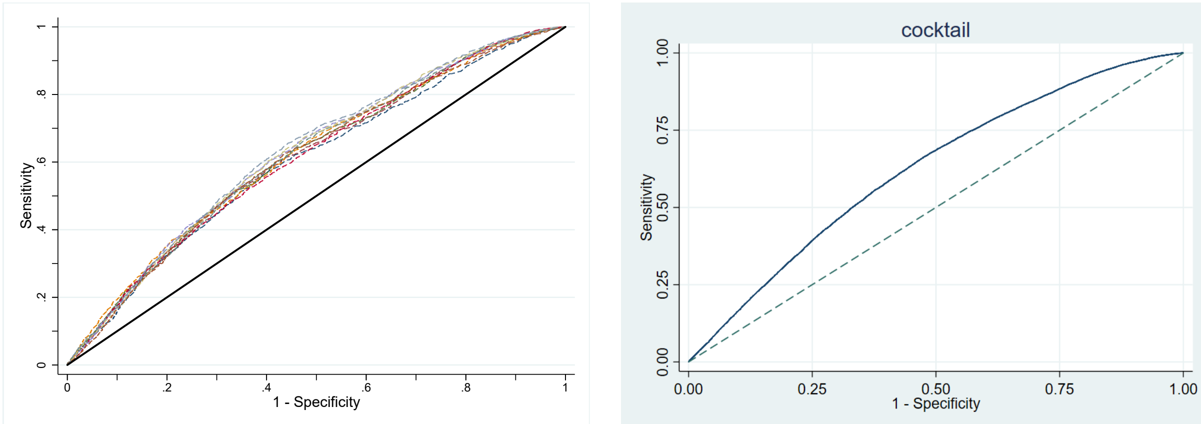
These figures illustrate the EDM measures on wash versus non-trading samples, respectively, by plotting the log ten-minute trading volumes in March 2012 and then applying the EDM algorithm to them. Panel A first looks at the wash trading sample, which is characterized by a step-like shape in its second half. The EDM algorithm detects two structural breaks, highlighted by dashed vertical lines. Panel B then looks at the non-wash trading sample, which has no visible structural breaks. The EDM algorithm identifies no breakouts in this series.

Figure A7: Responses to an Impulse in Volatility



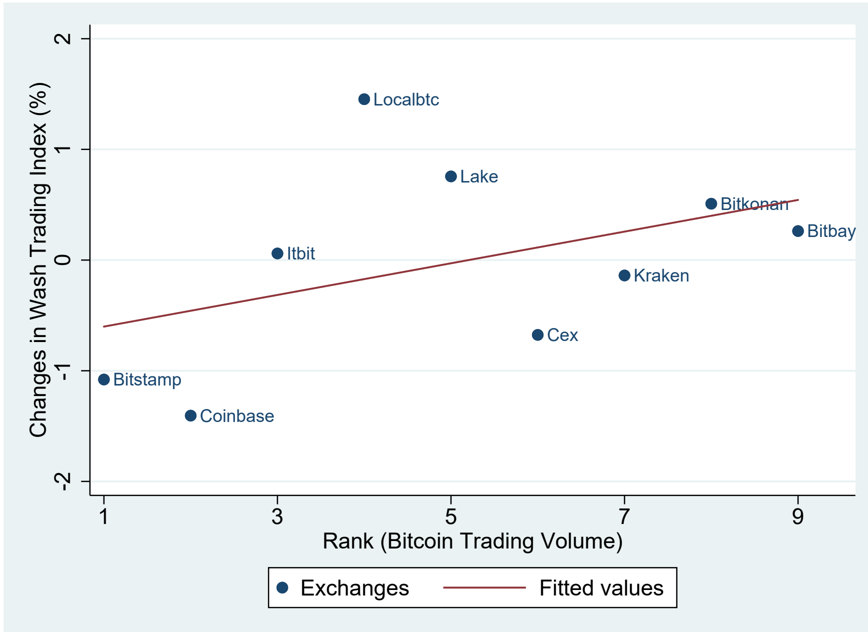
This figure plots the responses of wash trading volume and fee revenues following an impulse in volatility using the same VAR framework: $Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t,$ where A_i is a 5×5 matrix to be estimated, Y_t is a 5×1 vector corresponding to 1) the log wash trading volume (in units of BTC), 2) the log fee revenues collected from non-wash transactions (in units of BTC), 3) the log return of Bitcoin prices, 4) the liquidity of the Bitcoin market as measured by the Amihud measure, and 5) the realized volatility of Bitcoin return. The first two time series (wash trading volume and fee revenues from non-wash trades) are added by one before entering log transformations to avoid taking the log of zero; The latter two time series (Amihud measure and realized volatility) are calculated within 30-minute rolling windows. All ratios (return and Amihud) are winsorized at the 1% level, with two-way winsorization for returns, which could take both positive and negative values, and one-way winsorization for the Amihud measure, which is always non-negative. The VAR contains four lags based on the Bayesian information criterion. Shaded areas represent 95% confidence intervals. Responding to a one standard deviation increase in volatility, wash trading volume and fee revenues go up by more than 15% and 3% standard deviations, respectively. Both effects are persistent, as the impulses last beyond 60 minutes.

Figure A8: Cross-validation of Logistic Regression



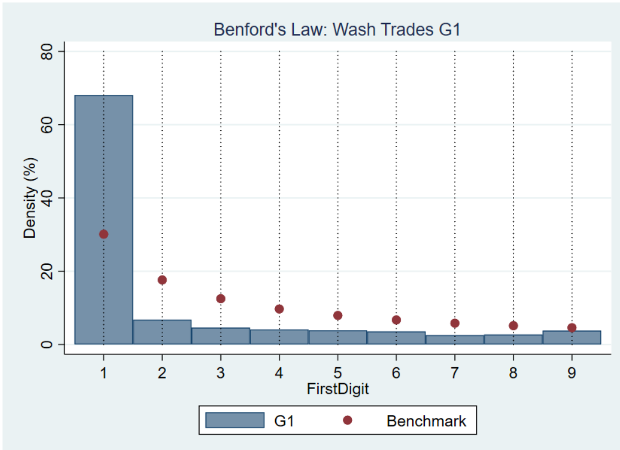
The left panel plots the Receiver Operating Characteristic (ROC) Curves from cross-validating a "cocktail" index in which we combine all four measured evaluated in Figure 6 into a logistic regression with LASSO model selection. Specifically, the cocktail index P_t is given by $P_t = \frac{1}{1+e^{-(-1.296+.0047\times Z_t+.0024\times R_t)}}$, where Z_t is our heuristic wash trading index developed in Section 3.6, and R_t is the anomalous ratio of un-rounded trades as presented in Figure 6. The cross-validation uses a default value of 5-folds, and the LASSO selection is based on the ROC area under curves from cross-validations. The index in window t quantifies its probability of containing wash trades. For comparison, the right panel also presents an in-sample ROC curve.

Figure A9: Changes in the Index on Other Exchanges

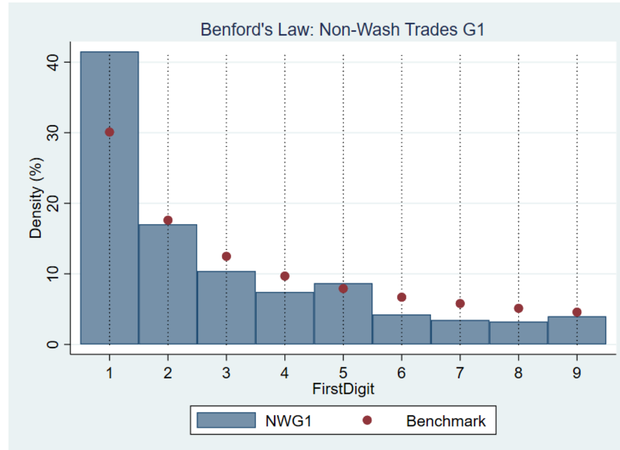


This figure applies the wash trading index to more recent samples from other Bitcoin exchanges, with trades data obtained from http://api.bitcoincharts.com/v1/csv/. We are able to get data from nine exchanges that were active from 2015 to 2018. Since our index is designed for capturing inter-temporal wash trading patterns within each exchange, we compare the average index between the 2015 - 2016 period and the 2017 - 2018 period and relate the differences (the vertical axis) to the exchange sizes measured by Bitcoin trading volumes (the horizontal axis, presented by exchange size rankings). As the figure shows, the intensity of wash trades on major exchanges (e.g. Coinbase and Bitstamp) have reduced in 2017-2018 compared to in 2015-2016 (in line with Cong, Li, Tang and Yang (2020)) while they have increased on smaller exchanges (consistent with Amiram, Lyandres and Rabetti (2020)).

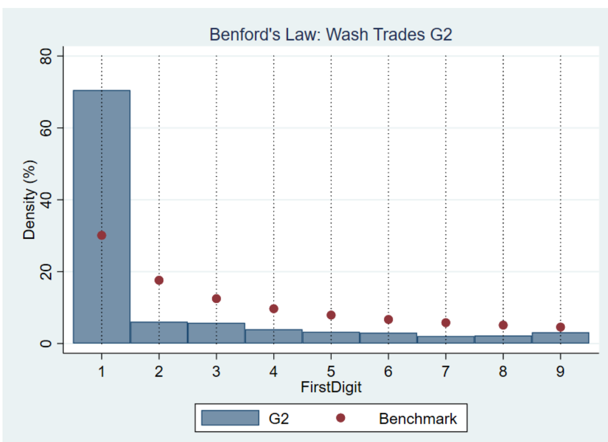
Figure A10: Evaluating Benford's Law (Alternative Wash Trading Definitions)



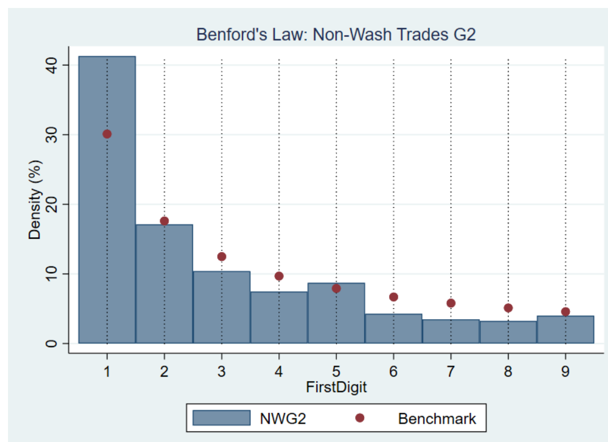
χ^2 statistics: 69.554; p-value: 0.000



χ^2 statistics: 7.907; p-value: 0.443



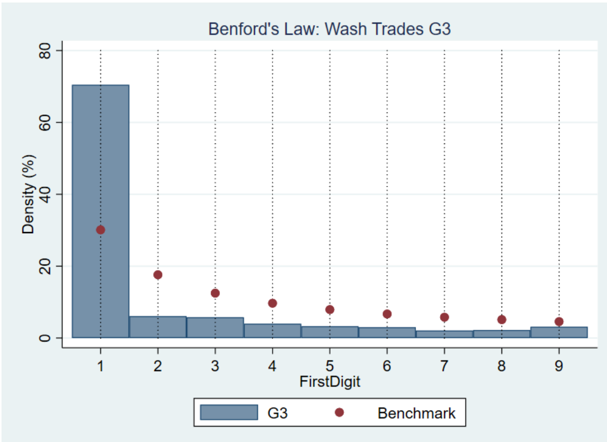
χ^2 statistics: 78.155; p-value: 0.000



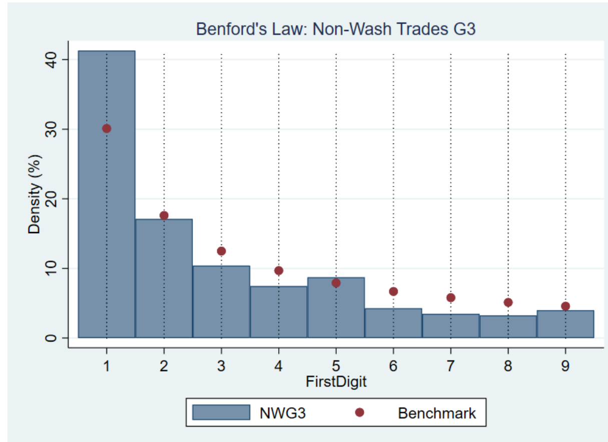
χ^2 statistics: 7.685; p-value: 0.465

This figure demonstrates the robustness of Benford's law as shown in Figure 5. We adopt various definitions of wash trades (G0 to G4) and compare the empirical distributions of the first significant digits of order sizes (bars) to theoretical benchmarks suggested by Benford's law for both wash and non-wash trades, respectively. We find that the distribution of wash trades significantly deviates from the Benford law, while it does not for non-wash trades. Our direct evidence thus provides support for the effectiveness of Benford's law for wash trading estimation. The sample is between June 26$^{\text{th}}$, 2011 and May 20, 2013.

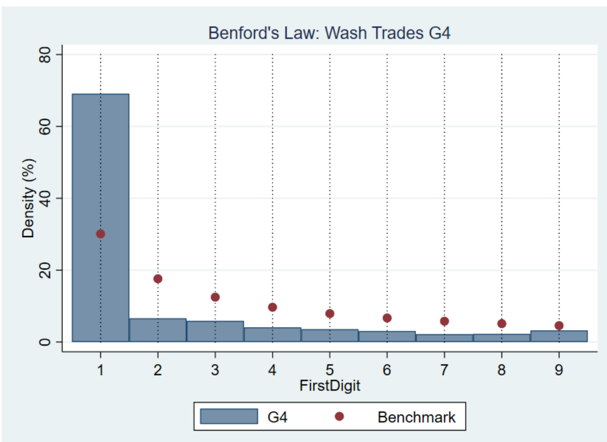
Figure A10: Evaluating Benford's Law (Alternative Wash Trading Definitions, Continued)



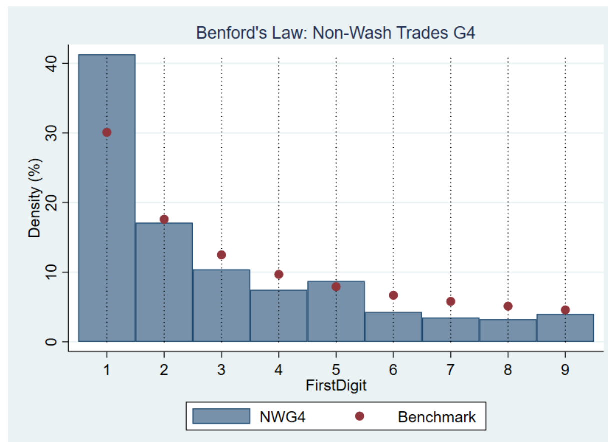
χ^2 statistics: 77.877; p-value: 0.000



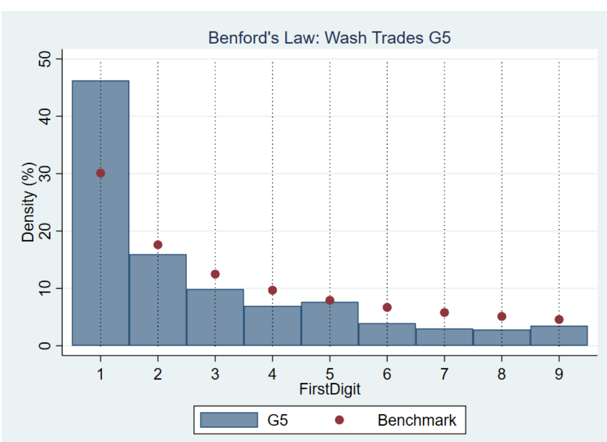
χ^2 statistics: 7.685; p-value: 0.465



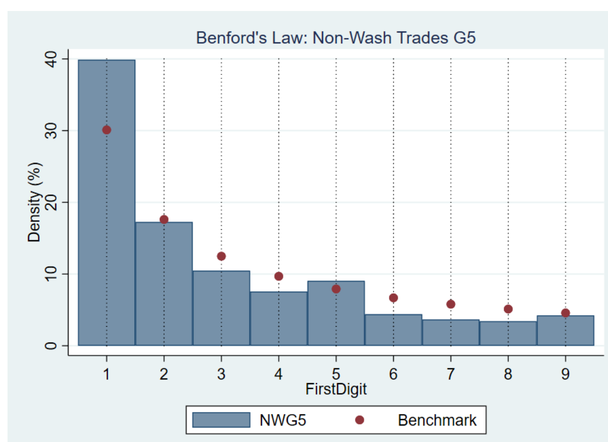
χ^2 statistics: 72.791; p-value: 0.000



χ^2 statistics: 7.685; p-value: 0.465



χ^2 statistics: 13.887; p-value: 0.085

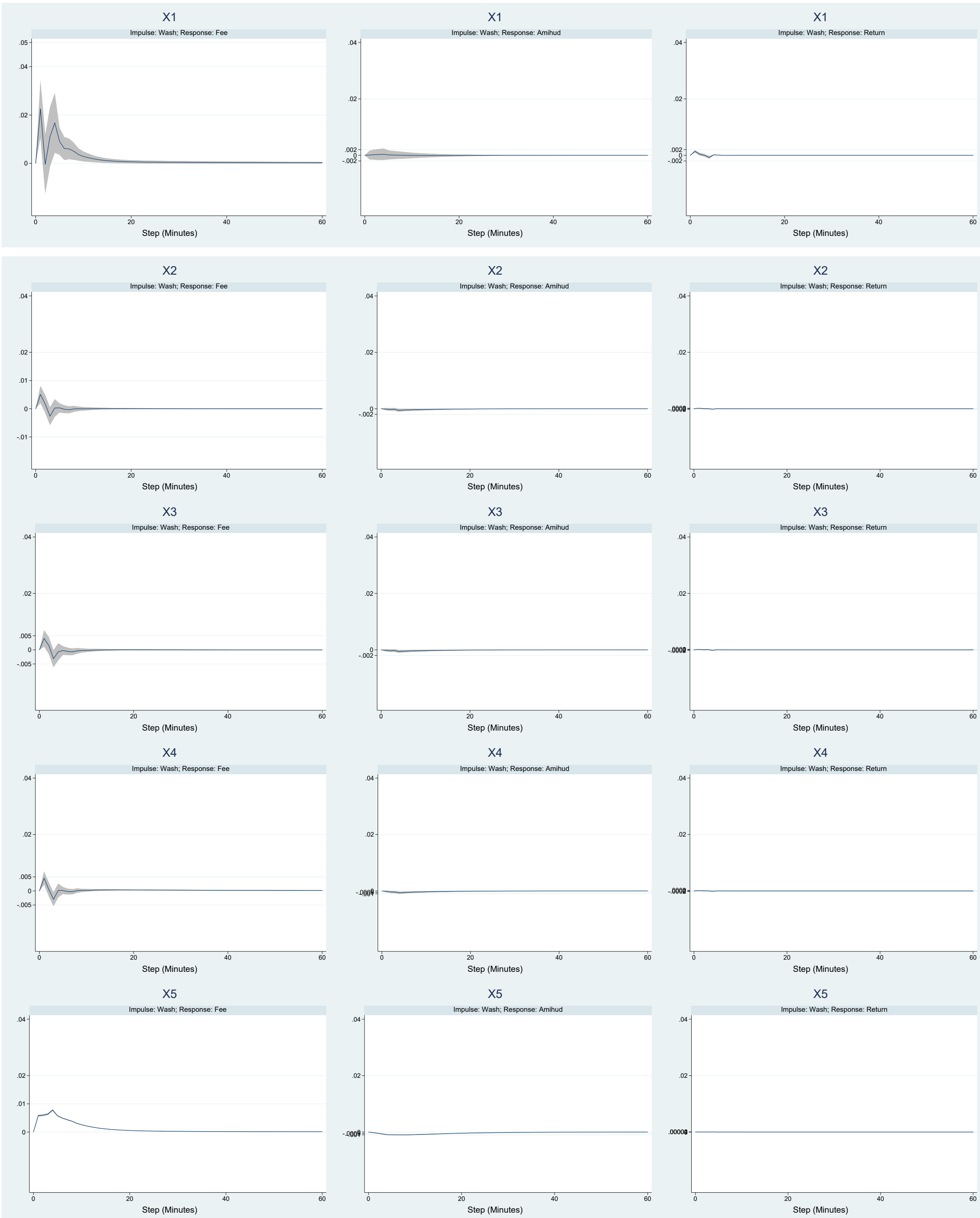


χ^2 statistics: 6.292; p-value: 0.615

Figure A11: Impulse response functions (Alternative Wash Trading Definitions)

These figures repeat the impulse response functions presented in Figure 8 but with wash trades alternatively defined by G1, G2, G3, G4, and G5, which augment the directly identified wash trades in Section 2 (denoted as G0) by further including trades among variously defined exchange insiders.

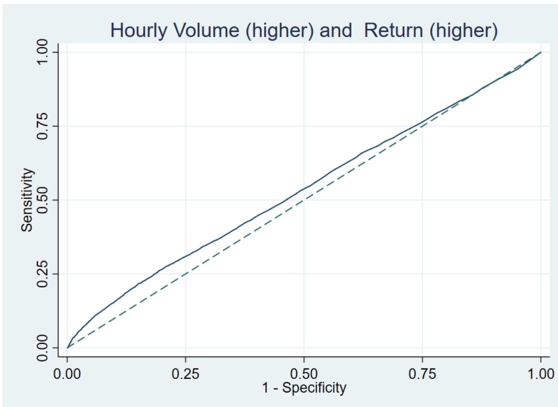
Figure A11: Impulse response functions (Alternative Wash Trading Definitions, Continued)



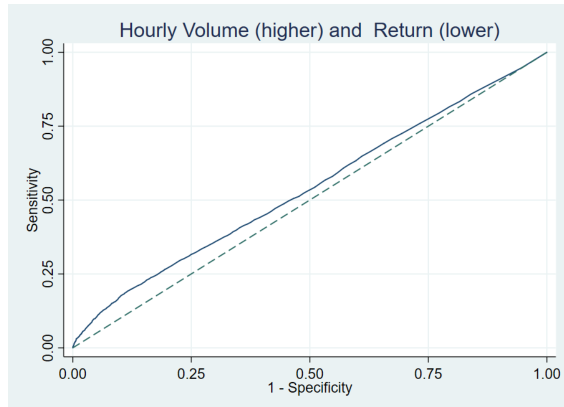
These figures repeat the impulse response functions presented in Figure 8 but with wash trades alternatively defined by G1, G2, G3, G4, and G5 with G0 removed (denoted as X1, X2, X3, X4, and X5, respectively) to avoid the directly identified wash trades overwhelming results.

Figure A12: Further Exploration into Constructing Wash Trading Index with Low-Frequency Data

Panel A



Panel B



These figures explore the feasibility of constructing wash trading indices with less granular public measures. Specifically, we plot the Receiver Operating Characteristic (ROC) Curves by applying adapted versions of our wash trading index, which in hour t look into the whole week ending one day before t and count the number of hours within the whole week that have both higher volumes than that in hour t and higher (Panel A) or lower (Panel B) returns than that in hour t, assuming that the higher the index in hour t is, the more likely that it contains wash trades.

Table A1: Size Clustering Patterns (Alternative Wash Trading Definitions)

trade size (BTC)	Round size unit (BTC)	Wash Trade Difference (1)	Wash Trade t-stat (2)	Non-Wash Trade Difference (3)	Non-Wash Trade t-stat (4)
		Panel I			
(0, Max]	$100{\times}10^{-1}$	-0.374	-7.821	-0.290	-11.428
(0, Max]	$100{\times}10^{-2}$	-0.384	-15.935	-0.242	-20.531
(0, Max]	$100{\times}10^{-3}$	-0.504	-41.783	-0.351	-64.793
(0, Max]	$100{\times}10^{-4}$	-0.523	-69.724	-0.376	-139.919
(0, Max]	$100{\times}10^{-5}$	-0.649	-142.313	-0.401	-283.877
(0, Max]	$100{\times}10^{-6}$	-0.697	-203.220	-0.331	-380.155
		Panel I - G2			
(0, Max]	$100{\times}10^{-1}$	-0.418	-9.485	-0.284	-11.119
(0, Max]	$100{\times}10^{-2}$	-0.387	-16.759	-0.239	-20.241
(0, Max]	$100{\times}10^{-3}$	-0.504	-42.545	-0.350	-64.572
(0, Max]	$100{\times}10^{-4}$	-0.521	-70.574	-0.375	-139.581
(0, Max]	$100{\times}10^{-5}$	-0.637	-141.414	-0.401	-283.706
(0, Max]	$100{\times}10^{-6}$	-0.690	-206.361	-0.331	-380.089
		Panel I - G3			
(0, Max]	$100{\times}10^{-1}$	-0.417	-9.442	-0.284	-11.119
(0, Max]	$100{\times}10^{-2}$	-0.388	-16.816	-0.239	-20.241
(0, Max]	$100{\times}10^{-3}$	-0.504	-42.665	-0.350	-64.578
(0, Max]	$100{\times}10^{-4}$	-0.520	-70.909	-0.375	-139.564
(0, Max]	$100{\times}10^{-5}$	-0.636	-141.437	-0.401	-283.683
(0, Max]	$100{\times}10^{-6}$	-0.687	-205.156	-0.331	-380.099
		Panel I - G4			
(0, Max]	$100{\times}10^{-1}$	-0.419	-9.679	-0.282	-11.033
(0, Max]	$100{\times}10^{-2}$	-0.382	-16.928	-0.239	-20.235
(0, Max]	$100{\times}10^{-3}$	-0.502	-44.244	-0.351	-64.662
(0, Max]	$100{\times}10^{-4}$	-0.512	-72.470	-0.375	-139.434
(0, Max]	$100{\times}10^{-5}$	-0.625	-145.073	-0.401	-283.450
(0, Max]	$100{\times}10^{-6}$	-0.660	-200.848	-0.331	-379.654
		Panel I - G5			
(0, Max]	$100{\times}10^{-1}$	-0.359	-12.412	-0.285	-10.355
(0, Max]	$100{\times}10^{-2}$	-0.294	-21.912	-0.217	-17.044
(0, Max]	$100{\times}10^{-3}$	-0.407	-63.696	-0.345	-57.717
(0, Max]	$100{\times}10^{-4}$	-0.386	-114.377	-0.359	-121.469
(0, Max]	$100{\times}10^{-5}$	-0.421	-228.292	-0.401	-261.995
(0, Max]	$100{\times}10^{-6}$	-0.446	-376.874	-0.307	-314.131

This table repeats the exercise in Table 5 to examine whether frequencies of trades at round sizes are higher than those at unrounded sizes within the wash trading subsample and the non-wash trading subsample, respectively, with alternative wash trading definitions. In Panels I and II, we follow Cong, Li, Tang and Yang (2020) and carry out two sets of t-tests with different rounded size and surrounding window definitions: Multiples of 100 units with a window radius 50 (100X-50, 100X+50), and multiples of 500 units with a window radius 100 (500X-100, 500X+100), respectively, where one unit ranges from $\times 10^{-6}$ to $\times 10^{-1}$ BTC. Specifically, for each rounded size (e.g. $200{\times}10^{-4}$ BTC) and its nearby window (e.g. between $(200-50)\times 10^{-4}$ and $(200+50)\times 10^{-4}$ BTC, excluding $200{\times}10^{-4}$ BTC), we compare the frequency of those trades at the rounded size versus the highest frequency among all other non-rounded trades within the window. The differences between the rounded and the highest unrounded frequencies are then carried out over all rounded sizes, with ***, **, and * denote *positive* differences (which indicate that frequencies at round sizes are higher than the rest within the observation window, suggesting trade-size clustering) per t-statistics at significance levels of 1%, 5%, and 10%, respectively.

Table A1: Size Clustering Patterns (Alternative Wash Trading Definitions, Continued)

Panel II

trade size (BTC)	Round size unit (BTC)	Wash Trade Difference (1)	Wash Trade t-stat (2)	Non-Wash Trade Difference (3)	Non-Wash Trade t-stat (4)
		Panel II - G1			
(0, Max]	$500{\times}10^{-2}$	-0.372	-9.566	-0.313	-15.888
(0, Max]	$500{\times}10^{-3}$	-0.431	-22.621	-0.269	-29.062
(0, Max]	$500{\times}10^{-4}$	-0.518	-51.130	-0.387	-89.941
(0, Max]	$500{\times}10^{-5}$	-0.574	-91.608	-0.388	-175.455
(0, Max]	$500{\times}10^{-6}$	-0.676	-166.364	-0.418	-358.921
(0, Max]	$500{\times}10^{-7}$	-0.694	-211.432	-0.340	-442.020
		Panel II - G2			
(0, Max]	$500{\times}10^{-2}$	-0.413	-11.467	-0.307	-15.526
(0, Max]	$500{\times}10^{-3}$	-0.433	-23.358	-0.269	-28.964
(0, Max]	$500{\times}10^{-4}$	-0.518	-52.008	-0.386	-89.770
(0, Max]	$500{\times}10^{-5}$	-0.564	-91.052	-0.387	-175.137
(0, Max]	$500{\times}10^{-6}$	-0.665	-166.381	-0.418	-358.801
(0, Max]	$500{\times}10^{-7}$	-0.688	-215.769	-0.341	-442.021
		Panel II - G3			
(0, Max]	$500{\times}10^{-2}$	-0.413	-11.468	-0.307	-15.527
(0, Max]	$500{\times}10^{-3}$	-0.434	-23.459	-0.269	-28.964
(0, Max]	$500{\times}10^{-4}$	-0.518	-52.298	-0.386	-89.773
(0, Max]	$500{\times}10^{-5}$	-0.564	-91.365	-0.387	-175.094
(0, Max]	$500{\times}10^{-6}$	-0.664	-166.223	-0.418	-358.807
(0, Max]	$500{\times}10^{-7}$	-0.686	-214.503	-0.341	-442.027
		Panel II - G4			
(0, Max]	$500{\times}10^{-2}$	-0.416	-11.736	-0.306	-15.463
(0, Max]	$500{\times}10^{-3}$	-0.434	-24.014	-0.269	-28.969
(0, Max]	$500{\times}10^{-4}$	-0.514	-53.884	-0.386	-89.795
(0, Max]	$500{\times}10^{-5}$	-0.553	-92.986	-0.387	-174.952
(0, Max]	$500{\times}10^{-6}$	-0.652	-169.259	-0.418	-358.519
(0, Max]	$500{\times}10^{-7}$	-0.655	-207.395	-0.340	-441.650
		Panel II - G5			
(0, Max]	$500{\times}10^{-1}$	-0.338	-15.065	-0.303	-14.283
(0, Max]	$500{\times}10^{-2}$	-0.312	-29.005	-0.254	-25.078
(0, Max]	$500{\times}10^{-3}$	-0.436	-85.548	-0.375	-78.950
(0, Max]	$500{\times}10^{-4}$	-0.396	-140.930	-0.372	-154.271
(0, Max]	$500{\times}10^{-5}$	-0.460	-297.048	-0.422	-332.622
(0, Max]	$500{\times}10^{-6}$	-0.452	-421.922	-0.324	-373.409

Table A1: Size Clustering Patterns (Alternative Wash Trading Definitions, Continued)

trade size (BTC)	Round size unit (BTC)	Wash Trade Difference (1)	Wash Trade t-stat (2)	Non-Wash Trade Difference (3)	Non-Wash Trade t-stat (4)
		Panel III			
(50, 500]	100×10^{-1}	-0.079	-1.602	0.113***	3.243
(5, 50]	100×10^{-2}	0.042	1.477	0.230***	7.995
(0.5, 5]	100×10^{-3}	0.004	0.142	0.128***	3.805
(0.05, 0.5]	100×10^{-4}	0.064***	2.906	0.206***	5.923
(0.005, 0.05]	100×10^{-5}	0.002	0.052	0.072**	2.186
(0.0005, 0.005]	100×10^{-6}	-0.055	-4.192	0.128***	6.203
		Panel III - G2			
(50, 500]	100×10^{-1}	-0.029	-0.638	0.113***	3.242
(5, 50]	100×10^{-2}	0.078***	2.589	0.230***	7.991
(0.5, 5]	100×10^{-3}	0.033	0.965	0.128***	3.806
(0.05, 0.5]	100×10^{-4}	0.062***	2.891	0.206***	5.924
(0.005, 0.05]	100×10^{-5}	0.022	0.560	0.072**	2.185
(0.0005, 0.005]	100×10^{-6}	-0.054	-4.124	0.128***	6.199
		Panel III - G3			
(50, 500]	100×10^{-1}	-0.026	-0.557	0.113***	3.242
(5, 50]	100×10^{-2}	0.078**	2.555	0.230***	7.991
(0.5, 5]	100×10^{-3}	0.033	0.986	0.128***	3.806
(0.05, 0.5]	100×10^{-4}	0.062***	2.894	0.206***	5.924
(0.005, 0.05]	100×10^{-5}	0.022	0.560	0.072**	2.185
(0.0005, 0.005]	100×10^{-6}	-0.054	-4.124	0.128***	6.199
		Panel III - G4			
(50, 500]	100×10^{-1}	-0.002	-0.057	0.113***	3.244
(5, 50]	100×10^{-2}	0.090***	2.927	0.230***	7.994
(0.5, 5]	100×10^{-3}	0.049	1.432	0.128***	3.806
(0.05, 0.5]	100×10^{-4}	0.074***	2.963	0.206***	5.928
(0.005, 0.05]	100×10^{-5}	0.028	0.721	0.072**	2.184
(0.0005, 0.005]	100×10^{-6}	-0.051	-3.858	0.127***	6.194
		Panel III - G5			
(50, 500]	100×10^{-1}	0.096**	2.533	0.120***	3.567
(5, 50]	100×10^{-2}	0.212***	7.241	0.239***	8.316
(0.5, 5]	100×10^{-3}	0.085**	2.447	0.146***	4.360
(0.05, 0.5]	100×10^{-4}	0.206***	5.764	0.204***	5.873
(0.005, 0.05]	100×10^{-5}	0.046	1.264	0.086***	2.789
(0.0005, 0.005]	100×10^{-6}	0.124***	4.883	0.105***	5.858

Table A1: Size Clustering Patterns (Alternative Wash Trading Definitions, Continued)

trade size (BTC)	Round size unit (BTC)	Wash Trade Difference (1)	Wash Trade t-stat (2)	Non-Wash Trade Difference (3)	Non-Wash Trade t-stat (4)
Panel IV					
(25, 250]	500×10^{-2}	0.014	0.383	0.129***	4.071
(2.5, 25]	500×10^{-3}	0.040	1.562	0.197***	5.465
(0.25, 2.5]	500×10^{-4}	-0.002	-0.070	0.129***	4.377
(0.025, 0.25]	500×10^{-5}	0.042	1.617	0.145***	3.738
(0.0025, 0.025]	500×10^{-6}	-0.015	-0.453	0.039	1.279
(0.00025, 0.0025]	500×10^{-7}	-0.093	-4.610	0.092***	4.136
Panel IV - G2					
(25, 250]	500×10^{-2}	0.027	0.698	0.129***	4.073
(2.5, 25]	500×10^{-3}	0.105***	2.962	0.197***	5.464
(0.25, 2.5]	500×10^{-4}	0.016	0.619	0.129***	4.376
(0.025, 0.25]	500×10^{-5}	0.054*	1.828	0.145***	3.736
(0.0025, 0.025]	500×10^{-6}	-0.013	-0.387	0.039	1.277
(0.00025, 0.0025]	500×10^{-7}	-0.092	-4.559	0.092***	4.130
Panel IV - G3					
(25, 250]	500×10^{-2}	0.026	0.677	0.129***	4.073
(2.5, 25]	500×10^{-3}	0.107***	3.004	0.197***	5.464
(0.25, 2.5]	500×10^{-4}	0.016	0.636	0.129***	4.376
(0.025, 0.25]	500×10^{-5}	0.054*	1.828	0.145***	3.736
(0.0025, 0.025]	500×10^{-6}	-0.013	-0.387	0.039	1.277
(0.00025, 0.0025]	500×10^{-7}	-0.092	-4.559	0.092***	4.130
Panel IV - G4					
(25, 250]	500×10^{-2}	0.040	1.074	0.128***	4.017
(2.5, 25]	500×10^{-3}	0.118***	3.244	0.197***	5.463
(0.25, 2.5]	500×10^{-4}	0.033	1.236	0.129***	4.376
(0.025, 0.25]	500×10^{-5}	0.060*	1.893	0.145***	3.738
(0.0025, 0.025]	500×10^{-6}	-0.008	-0.249	0.039	1.274
(0.00025, 0.0025]	500×10^{-7}	-0.090	-4.482	0.092***	4.119
Panel IV - G5					
(25, 250]	500×10^{-2}	0.115***	3.423	0.138***	4.494
(2.5, 25]	500×10^{-3}	0.170***	4.641	0.215***	6.029
(0.25, 2.5]	500×10^{-4}	0.104***	3.520	0.139***	4.639
(0.025, 0.25]	500×10^{-5}	0.141***	3.347	0.143***	3.799
(0.0025, 0.025]	500×10^{-6}	0.020	0.591	0.054*	1.858
(0.00025, 0.0025]	500×10^{-7}	0.070**	2.439	0.072***	3.946

Table A2: Power-law Fitting (Alternative Wash Trading Definitions)

wash trading definition	wash sample			non-wash sample		
	$\hat{\alpha}_{OLS}$	$\hat{\alpha}_{Hill}$	Pareto-Lévy ($1 < \alpha < 2$)	$\hat{\alpha}_{OLS}$	$\hat{\alpha}_{Hill}$	Pareto-Lévy ($1 < \alpha < 2$)
G0	1.883	1.580	Y	2.674	2.054	N
G1	1.834	1.582	Y	2.658	2.054	N
G2	1.880	1.735	Y	2.668	2.061	N
G3	1.883	1.732	Y	2.668	2.061	N
G4	1.908	1.729	Y	2.669	2.061	N
G5	2.458	2.007	N	2.649	2.083	N

This table presents the results of power-law fitting on our sample, with various wash trading definitions. Ordinary Least Square (OLS) and Maximum Likelihood Estimation (MLE) are applied for the estimation of scaling parameters $\hat{\alpha}_{OLS}$ and $\hat{\alpha}_{Hill}$, respectively. We also check whether the estimated parameters are within the Pareto–Lévy range ($1 < \alpha < 2$) and mark "Y" if both exponents lie within the Pareto–Lévy range. Overall, we find that the α coefficients estimated from wash trades tend to fall within the Pareto-Lévy range, while those estimated from non-wash trades tend to fall out of the Pareto-Lévy range. This finding stands in contrast with the argument in Cong, Li, Tang and Yang (2020) that normal trades tend to have tails in order sizes that follow power-law distributions with α coefficients within the Pareto-Lévy range. Therefore, our direct evidence suggests that the power law test may not be an accurate method for testing the presence of wash trades in a sample.

Table A3: Deviations from log-normal distributions (Alternative Wash Trading Definitions)

Panel A: Trading volume

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	0.139	0.071	0.045	0.075	0.147	0.203	0.254
	non-wash	0.051	0.020	0.013	0.039	0.056	0.062	0.091
G1	wash	0.139	0.071	0.045	0.073	0.143	0.203	0.254
	non-wash	0.051	0.020	0.013	0.039	0.056	0.062	0.091
G2	wash	0.137	0.066	0.048	0.074	0.127	0.191	0.276
	non-wash	0.051	0.020	0.013	0.038	0.057	0.062	0.091
G3	wash	0.138	0.065	0.048	0.074	0.127	0.191	0.276
	non-wash	0.051	0.020	0.013	0.038	0.057	0.062	0.091
G4	wash	0.138	0.068	0.046	0.073	0.127	0.194	0.276
	non-wash	0.051	0.020	0.013	0.038	0.057	0.062	0.091
G5	wash	0.078	0.023	0.034	0.065	0.079	0.095	0.130
	non-wash	0.061	0.023	0.016	0.045	0.066	0.076	0.099

Panel B: #Trades

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	0.223	0.071	0.106	0.163	0.219	0.293	0.337
	non-wash	0.035	0.010	0.020	0.028	0.036	0.041	0.057
G1	wash	0.223	0.071	0.106	0.163	0.219	0.29	0.342
	non-wash	0.035	0.010	0.020	0.028	0.036	0.041	0.057
G2	wash	0.225	0.084	0.106	0.158	0.197	0.291	0.427
	non-wash	0.036	0.010	0.020	0.028	0.036	0.042	0.056
G3	wash	0.225	0.084	0.106	0.159	0.197	0.283	0.427
	non-wash	0.036	0.010	0.020	0.028	0.036	0.042	0.056
G4	wash	0.224	0.081	0.103	0.166	0.197	0.283	0.398
	non-wash	0.036	0.010	0.021	0.027	0.036	0.042	0.056
G5	wash	0.063	0.022	0.027	0.047	0.064	0.075	0.112
	non-wash	0.042	0.011	0.019	0.034	0.04	0.05	0.065

This table presents the summary statistics of the monthly sample of KS distances between the empirical cumulative density function of the natural logarithm of trading volume (in Panel A) and the number of trades (Panel B) aggregated into ten-minute bins within a month, and the cumulative distribution function (c.d.f.) of a normal distribution with the same mean and variance. Overall, as KS measures take the entire sample's properties, augmenting the wash trading sample with "hidden" wash trades does not significantly affect monthly summary statistics.

Table A4: Structural breaks (Alternative Wash Trading Definitions)

Panel A: log(Trading volume)

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	7.318	6.869	0	2	6	10	21
	non-wash	0	0	0	0	0	0	0
G1	wash	7.318	6.986	0	2	6	10	22
	non-wash	0	0	0	0	0	0	0
G2	wash	4.682	6.462	0	0	1.5	7.25	21
	non-wash	0	0	0	0	0	0	0
G3	wash	4.773	6.941	0	0	1.5	5.75	22
	non-wash	0	0	0	0	0	0	0
G4	wash	5.773	8.011	0	0	1.5	9.5	23
	non-wash	0	0	0	0	0	0	0
G5	wash	2.682	5.785	0	0	0	0	23
	non-wash	0	0	0	0	0	0	0

Panel B: log(#Trades)

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	21.995	28.429	0	5.25	11.5	25.2	107
	non-wash	1.182	2.130	0	0	0	1.5	6
G1	wash	21.818	28.458	0	5.25	11	25	107
	non-wash	1.182	2.130	0	0	0	1.5	6
G2	wash	22.545	31.831	0	2.50	12.5	17.5	108
	non-wash	1.273	2.374	0	0	0	1.5	8
G3	wash	22.136	32.264	0	2.50	12	17.5	109
	non-wash	1.273	2.374	0	0	0	1.5	8
G4	wash	18.273	25.663	0	2.50	13.5	16	108
	non-wash	1.273	2.374	0	0	0	1.5	8
G5	wash	77.591	87.923	0	11.8	39.5	118	286
	non-wash	1.682	3.524	0	0	0	1.5	14

This table presents the summary statistics of the monthly sample of EDM structural breaks within the log trading volume (in Panel A) and the log of the number of trades (Panel B) aggregated into ten-minute bins within a month. Only full months (that is, from July 2011 to May 2013) are included. Overall, augmenting the wash trading sample with "hidden" wash trades does not significantly affect monthly summary statistic patterns between wash and non-wash trading samples. Specifically, for log-transformed volume or number of trade series, the EDM algorithm identifies a significant number of structural breaks.

Table A4: Structural breaks (Alternative Wash Trading Definitions, Continued)

Panel C: Trading volume

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	0.182	0.664	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G1	wash	0.182	0.664	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G2	wash	0.182	0.664	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G3	wash	0.182	0.664	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G4	wash	0.182	0.664	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G5	wash	0	0	0	0	0	0	0
	non-wash	0	0	0	0	0	0	0

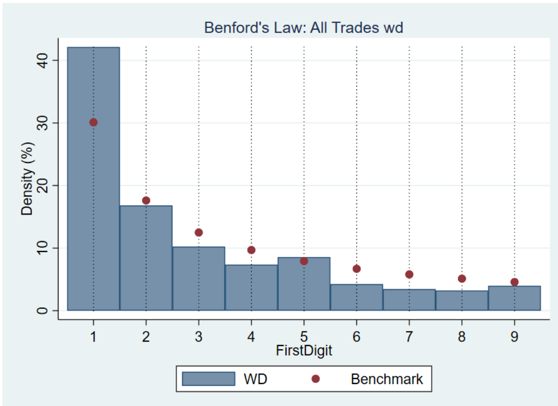
Panel D: #Trades

wash trade definition		Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max.
G0	wash	0.136	0.640	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G1	wash	0.136	0.640	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G2	wash	0.409	1.054	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G3	wash	0.409	1.054	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G4	wash	0.409	1.054	0	0	0	0	3
	non-wash	0	0	0	0	0	0	0
G5	wash	0	0	0	0	0	0	0
	non-wash	0	0	0	0	0	0	0

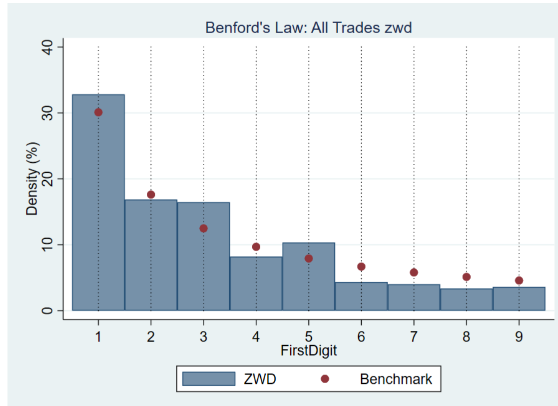
This table presents the summary statistics of the monthly sample of EDM structural breaks within the trading volume (in Panel C) and the number of trades (Panel D) aggregated into ten-minute bins within a month. Only full months (that is, from July 2011 to May 2013) are included. Overall, augmenting the wash trading sample with "hidden" wash trades does not significantly affect monthly summary statistic patterns between wash and non-wash trading samples. Specifically, for (not log-transformed) raw volume or the number of trade series, the EDM algorithm was not sensitive enough to capture many structural breaks.

Table A5: Comparing Days with Wash Trading vs. Those without

Panel A: Benford's Law



χ^2 statistics: 8.504; p-value: 0.386



χ^2 statistics: 4.696; p-value: 0.790

Panel B: Trade Size Clustering

trade size (BTC)	Round size unit (BTC)	Wash Trade Difference (1)	Wash Trade t-stat (2)	Non-Wash Trade Difference (3)	Non-Wash Trade t-stat (4)
		Panel I			
(0, Max]	100×10^{-1}	-0.310	-12.463	-0.310	-5.506
(0, Max]	100×10^{-2}	-0.251	-21.402	-0.299	-10.945
(0, Max]	100×10^{-3}	-0.353	-65.270	-0.388	-30.637
(0, Max]	100×10^{-4}	-0.375	-139.808	-0.414	-55.934
(0, Max]	100×10^{-5}	-0.401	-283.565	-0.525	-104.277
(0, Max]	100×10^{-6}	-0.332	-381.742	-0.391	-81.737
		Panel II			
(0, Max]	500×10^{-1}	-0.323	-16.585	-0.302	-6.795
(0, Max]	500×10^{-2}	-0.279	-30.128	-0.336	-15.412
(0, Max]	500×10^{-3}	-0.388	-90.411	-0.438	-43.053
(0, Max]	500×10^{-4}	-0.387	-175.331	-0.462	-72.471
(0, Max]	500×10^{-5}	-0.418	-358.759	-0.541	-115.291
(0, Max]	500×10^{-6}	-0.342	-443.640	-0.372	-78.574
		Panel III			
(50, 500]	100×10^{-1}	0.113***	3.240	-0.124	-1.961
(5, 50]	100×10^{-2}	0.230***	8.004	0.131***	4.273
(0.5, 5]	100×10^{-3}	0.128***	3.809	0.055	1.399
(0.05, 0.5]	100×10^{-4}	0.206***	5.941	0.162***	3.888
(0.005, 0.05]	100×10^{-5}	0.072**	2.180	-0.016	-0.465
(0.0005, 0.005]	100×10^{-6}	0.117***	5.865	-0.213	-6.747
		Panel IV			
(25, 250]	500×10^{-2}	0.129***	4.085	0.029	0.814
(2.5, 25]	500×10^{-3}	0.197***	5.463	0.142***	3.791
(0.25, 2.5]	500×10^{-4}	0.129***	4.372	0.079***	2.589
(0.025, 0.25]	500×10^{-5}	0.145***	3.761	0.085**	1.987
(0.0025, 0.025]	500×10^{-6}	0.039	1.256	-0.053	-2.083
(0.00025, 0.0025]	500×10^{-7}	0.081***	3.807	-0.322	-5.894

In this set of figures/tables, we repeat the exercises in Figure 5 and Table 5, 6, and 7 with a small tweak: instead of comparing wash trades with non-wash trades, we compare the transactions on days with wash trading and those on days without.

Table A5: Comparing Days with Wash Trading vs. Those without (Continued)

Panel C: Deviations from Log-normal Distributions

	Volume			#Trades		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0.051	1	-0.949	0.035	1	-0.965
SD/t	(0.02)	(0)	(-237.162)	(0.01)	(0)	(-457.856)

Panel D: Structural breaks

	Volume (log)			#Trades (log)		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0	0.364	-0.364	0.955	1.727	-0.773
SD/t	(0)	(1.217)	(-1.402)	(1.963)	(4.211)	(-0.780)
	Volume			#Trades		
	wash	non-wash	difference	wash	non-wash	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0	0.136	-0.136	0	0.227	-0.227
SD/t	(0)	(0.64)	(-1)	(0)	(0.752)	(-1.418)

Table A6: External Wash Trading Estimates

Exchanges	2015-16 (1)	2017-18 (2)	2015-18 (3)	Diff (4)	t-stat (5)
Bitstamp	0.168	0.157	0.163	-0.0108	-50.72
Coinbase	0.173	0.159	0.169	-0.0141	-62.73
Itbit	0.159	0.159	0.159	0.0006	3.69
Localbtc	0.177	0.191	0.184	0.0145	82.90
Lake	0.157	0.165	0.162	0.0076	21.64
Cex	0.167	0.160	0.163	-0.0068	-36.97
Kraken	0.158	0.157	0.157	-0.0014	-6.32
Bitkonan	0.129	0.134	0.132	0.0051	24.55
Bitbay	0.144	0.146	0.146	0.0026	11.02

This table presents the raw estimates behind Figure A9. Specifically, we apply our wash trading index to more recent samples from other Bitcoin exchanges, with trades data obtained from http://api.bitcoincharts.com/v1/csv/. We are able to get data from nine exchanges that were active from 2015 to 2018. Since our index is designed for capturing inter-temporal wash trading patterns within each exchange, we compare the average index between the 2015 - 2016 period and the 2017 - 2018 period. As the table shows, the intensity of wash trades on major exchanges (e.g. Coinbase and Bitstamp) have reduced in 2017-2018 compared to in 2015-2016 (in line with Cong, Li, Tang and Yang (2020)) while they have increased on smaller exchanges (consistent with Amiram, Lyandres and Rabetti (2020)).

B Identifying exchange insiders

In this section, we follow the Willy report and define a sequence of increasingly inclusive sets of exchange insiders. We first list the following special trader IDs that have been singled out in the Willy report:

- Willy: Willy is a nickname given to a bot (or several bots) who buys a random number (between 10 and 20) of bitcoins every 5-10 minutes, non-stop, for at least a month until the end of January 2014. A number of traders reportedly first began to suspect the presence of Willy in December 2013, and a brief technical glitch that brought down the Mt.Gox site in early January 2014 finally confirmed Willy's existence (Reddit).

  In our data, all Willy's trader IDs are detected in the following two steps as instructed by Willy report:

  1. IDs 817985, 825654, 832432 are first detected by matching patterns with recognized bot behaviors. These IDs also have unusual User_Country and User_State fields: Normally, these fields contain country/state FIPS codes for verified users or are empty ("`!!`") for unverified users (users who failed verification). However, the three detected IDs all have "`??`" in these fields.
  2. We then single out all trader IDs that have "`??`" for their User_Country and User_State fields.[40] These IDs indeed behave like what the community has identified of Willy: They never performed a single sale, and their trades seamlessly connected to each other — when one user became inactive, the next was created usually within a few hours and actively market-buy coins until some very exact amount of USD ($2,500,000 being the most common) was spent. Such serial trading activities went back all the way to September 27$^{\text{th}}$, 2013. In total, about $112 million was spent to buy close to 270,000 BTC – the bulk of which was bought in November 2013 driving the bitcoin price bubble. The very first "`??`" user is also a "time-traveler" — it has an unusually high ID of 807884 even though regular accounts at that point only went up to around 650000.

Although we will remove Willy from our sample as its active periods are outside of June 26$^{\text{th}}$, 2011 to May 20, 2013 — our interested time period during which wash trading was active (recall that the first appearance of Willy was on Sept. 27$^{\text{th}}$, 2013), Willy is nevertheless important for identifying the following trader IDs that are suspected to be linked to Mt.Gox's owner:

- Markus: Inspired by Willy's "time-traveler" behavior, another time-traveler is singled out (698630, with a registered country and state: "JP", "40" – the FIPS code for Tokyo, Japan, where Mt.Gox is headquartered) and nicknamed Markus. After being active for close to 8 months, Markus became completely inactive 7 hours before the first Willy account became active, and has thus been singled out by Nilsson (2014) and Gandal, Hamrick, Moore and Oberman (2018) to be controlled by the same entity behind Willy.

  Markus also has two peculiar attributes that further strengthen its insider status: First, it always pays zero fees. Second, related to the price correction discussed above in Section

[40] Specifically, 807884, 658152, 659582, 661608, 665654, 683148, 689932, 693122, 694306, 695340, 697722, 698233, 698232, 698234, 698235, 698236, 698237, 698238, 711137, 714565, 716004, 718998, 722068, 724340, 726910, 730861, 734205, 739116, 742432, 746069, 757154, 764692, 769205, 770436, 774567, 783273, 787018, 790503, 790667, 791191, 791965, 793833, 796081, 796083, 804879, 809401, 817985, 825654, and 832432.

1, Markus' fiat spent when buying coins always carries forward from the previous trade's fiat spent, regardless of the actual volume of BTC bought, and thus generating seemingly completely random prices paid per bitcoin. The Willy report conjectured that for Markus, the "Money" spent field is in fact empty, and the program that generates the trading logs simply carries forward whatever latest value was already there before. Finally, like Willy, Markus' trader ID 698630 was also out of place.

- MagicalTux: MagicalTux refers to the trader ID 634, singled out for the following reason: As we have pointed out in the data description part of Section 1, for some months in 2013 (e.g. April 2013), there are two versions of trading logs in the leaked database: a full log and an anonymized log that differ in two and only two ways:

  1. User hashes and country/state codes in the anonymized log are removed.[41]
  2. More importantly for our analysis, Markus' out-of-place trader ID (698630) in the full log is changed to a small number (634) in the anonymized log, and its strange fixed "Money" values are corrected to the expected values.

  The second point above links ID 634 to Markus. To add to its insider status, from a 2011 leaked account list referenced in Willy report, the user with ID 634 has username "MagicalTux", which is a pseudo-name widely used by Mark Karpelès, the owner and CEO of Mt.Gox, across various venues, including for a now-defunct blog of his, his Reddit account, Bitcointalk.org forum username, and his current Twitter handle.

- #1: Other than the occasional exceptions mentioned above, trader IDs are assigned to new traders in chronicle orders of registration. Therefore, traders with small IDs, and especially trader #1, the first registered trader on Mt.Gox, are most likely to be the exchange owner or at least closely related to the exchange owner.

In the rest of the paper, we will refer to the above three IDs (Markus, MagicalTux, and #1) as "Known Insiders". This category is our most conservative and confident insider designation, as they have been used in various other papers for the same purpose.

The facts that Markus pays zero fees in all his transactions, that Markus has its location in Mt.Gox's headquarter (Japan), and that trader IDs are assigned incrementally as new users register, inspire us to further single out the following special trader IDs.

- Zero-fee Traders: trader IDs that never pay fees, unlike most other trader IDs.

- Hijacker: The Willy report has highlighted a subset of Zero-fee Traders, who in addition to paying zero fees in all transactions, also have location being "Japan" and trader IDs being small (<1000). These 285 IDs are so named by the Willy report as their trades demonstrate patterns suggesting that they are not executed by their original account holders, but rather "hijacked" in some way (see details here).

- Double Users: trader IDs (all of which are smaller than 1000) have the same "User_ID" but different "User" values in the raw data.

[41] The Willy report conjectured that the anonymized log was created to send off to auditors or investors to show some internals.

- Zero-Fee Traders: trader IDs that never pay fees, unlike most other trader IDs.
- All Wash Traders: trader IDs who have ever taken part in self-self wash trades.

These IDs are then used in Section 4.2 to further augment our wash trading sample.

C Quantifying wash trading with trade-size clustering

Compared to other "public" measures, a unique advantage of the trade size clustering approach is that it facilitates the quantification of wash trading. We demonstrate this potential following the framework in Cong, Li, Tang and Yang (2020), but using our suggested improvement to the trading-size clustering method. Specifically, following the heuristic in Table 5, we first define a trade to be rounded if its size is in multiples of 500×10^{-7}, 500×10^{-6}, 500×10^{-5}, 500×10^{-4}, 500×10^{-3}, or 500×10^{-2} BTC when in size buckets of $(0.00025, 0.0025]$, $(0.0025, 0.025]$, $(0.025, 0.25]$, $(0.25, 2.5]$, $(2.5, 25]$, or $(25, 250]$, respectively. We then use the sample of non-wash trades to estimate a benchmark level of trade "roundness", that is, we conduct the following regression:

$$\ln(V_{\text{Unrounded}_t}) = \text{constant} + \beta \times \ln(V_{\text{Rounded}_t}) + \epsilon_t, \quad (4)$$

where $V_{\text{Unrounded}_t}$ and V_{Rounded_t} are the volumes of unrounded and rounded trades during week t, respectively. The regression coefficients are then used to calculate the "benchmark" volume for unrounded trades in our whole sample, and the difference is then interpreted as a rough estimate of wash trading. From this exercise, we estimate a 3.30% of volume being generated by wash trades. This estimate is higher than the true percentage of wash trades in our sample but is at the same order of magnitude. Hence, the direct evidence confirms that using trade size clustering to quantify wash trading could give useful ballpark estimates, although the measure tends to give slightly higher estimates than the actual data.[42]

[42] There could be two reasons to account for the over-estimate: First, as the "public" measure is based on statistical anomalies, it may capture both wash trading and other manipulative trades; Second, as we adopt a narrow definition of wash trading, our "ground truth" may itself be an under-estimate. See discussions in Section 2.1.