Crypto wash trading

Article

Accepted Version

Cong, W. L., Li, X. ORCID: https://orcid.org/0000-0003-1034-4539, Tang, K. and Yang, Y. (2023) Crypto wash trading. Management Science, 69 (11). pp. 6427-6454. ISSN 1526-5501 doi: https://doi.org/10.1287/mnsc.2021.02709 Available at https://centaur.reading.ac.uk/113166/

Crypto Wash Trading*

Lin William Cong   Xi Li   Ke Tang   Yang Yang

First draft: Nov 2019; current draft: July 2023

Abstract

We present a systematic approach to detect fake transactions on cryptocurrency exchanges by exploiting robust statistical and behavioral regularities associated with authentic trading. Our sample consists of 29 centralized exchanges, among which the regulated ones feature transaction patterns consistently observed in financial markets and nature. In contrast, unregulated exchanges display abnormal first-significant-digit distributions, size rounding, and transaction tail distributions, indicating widespread manipulation unlikely driven by specific trading strategy or exchange heterogeneity. We then quantify the wash trading on each unregulated exchange, which averaged over 70% of the reported volume. We further document how these fabricated volumes (trillions of dollars annually) improve exchange ranking, temporarily distort prices, and relate to exchange characteristics (e.g., age and user base), market conditions, and regulation. Overall, our study cautions against potential market manipulations on centralized crypto exchanges with concentrated power and limited disclosure requirements, and highlights the importance of FinTech regulation.

Keywords: Bitcoin; CeFi; Cryptocurrency; Forensic Finance; Fraud Detection; Regulation

JEL Classification: G18, G23, G29.

*The authors are especially grateful to Agostino Capponi, Deeksha Gupta, Kose John, Tao Li, Evgeny Lyandres, and Fahad Saleh for repeated discussions and detailed feedback. We also thank Marlene Amstad, Mykola Babiak, Kevin Dowd, Valeria Ferrar, Itay Goldstein, Hanna Halaburda, Angel Hernando-Veciana, Andrew Karolyi, Dongyongp Lee, Minhyuk Lee, Jiasun Li, Laura Xiaolei Liu, Roger Loh, Emmanouil Platanakis, Amin Shams, Donghwa Shin, Rajeev Singhal, Baolian Wang, Shang-jin Wei, Wei Xiong, Scott Yonker and seminar and conference participants and reviewers at the Alibaba Group Luohan Academy Webinar, 17th Asia-Pacific Association of Derivatives Annual Conference, Australasian Banking and Finance Conference, Behavioral Finance/Corporate Finance/Digital Finance (BF/DF/CF) Seminar Group, Cornell University, China International Conference in Finance 2021, Cowles Foundation for Research In Economics Conference on the Economics of Cryptocurrencies, 1st Crypto and Blockchain Economics Research Conference, 11th CSBF Conference (National Taiwan University), Durham University Department of Economics and Finance, 2021 Eastern Finance Association Annual Meeting, Econometric Society World Congress (Bocconi University), Economic Club of Memphis, European Financial Management Association Annual Meeting 2021, European Securities and Markets Authority, Federal Reserve System-Wide Webinar, 2021 Financial Management Association Annual Meeting, 11th Financial Markets and Corporate Governance Conference, 2021 Global AI Finance Research Conference, IIF International Research Conference & Award Summit, 3rd International Conference on Blockchain Economics (Tokenomics 2021), 13th International Risk Management Conference, Inaugural KAIST Digital Finance Conference, Inaugural Machine Lawyering Conference: "Human Sovereignty and Machine Efficiency in the Law," 18th Paris December Finance Meeting, Paris FinTech and Crypto Webinar, Ripple Labs London Onsite (Markets Team), 2023 SCU Crypto Conference, 60th Southwestern Finance Association Meeting, Sun Yat-sen University, 3rd Toronto FinTech Conference, Tsinghua University PBC School of Finance, UCSB-ECON Defi Seminar Series, University of Central Florida, University of Technology Sydney (UTS), University of New South Wales (UNSW Sydney), University of Zurich (UZH) Blockchain Center, 3rd UWA Blockchain and Cryptocurrency Conference, USAO-N.D. Cal. / U.S. DOJ Fraud Section / National Cryptocurrency Enforcement Team Cryptocurrency Fraud Seminar, 1st Virtual Symposium on Web3 Financing and Inclusivity, Inaugural Wolfram ChainScience Conference (Boston), World Finance Conference 2021, Xi'an Jiaotong University, and the Zhongnan University of Economics and Law for helpful comments. This research was partly funded by the Ewing Marion Kauffman Foundation, the National Natural Science Foundation of China (Grant No. 72192802, 72192800 and 72192801), Ripple Labs, and the FinTech at Cornell Initiative. The authors have no affiliation with or research support from any cryptocurrency exchange. The contents of this publication are solely the responsibility of the authors. Cong (will.cong@cornell.edu) is at the Cornell University SC Johnson College of Business and NBER; Li (xi.li@icmacentre.ac.uk) at ICMA Centre, Henley Business School, University of Reading; Tang (ketang@tsinghua.edu.cn) is at the Institute of Economics, School of Social Sciences, Tsinghua University and Yanqi Lake Beijing Institute of Mathematical Sciences and Applications; Yang (yang.yang@bristol.ac.uk) is at the School of Computer Science, University of Bristol.

1 Introduction

The combined market capitalization of all cryptocurrencies reached a peak of U.S. $3 trillion in late 2021 and, despite recent market crashes, still surpassed U.S. $1.2 trillion as of July 2023. The monthly crypto trading volume amounted to trillions of U.S. dollars in 2020, multiplying that of equity markets (Helms, 2020). Both financial institutions and retail investors have had substantial exposure to the cryptocurrency industry (Bogart, 2019; FCA, 2019; Fidelity, 2019; Henry, Huynh, and Nicholls, 2019). Meanwhile, crypto exchanges, arguably the most profitable players in the ecosystem, remain mostly unregulated until recently. As of mid-2022, regulated exchanges (Coinbase, Bitstamp, Gemini, BitFlyer, itBit, etc.) only cover less than 3% of spot market transactions. In the process of vying for dominance in this lightly regulated market, crypto exchanges became increasingly vertically integrated in the absence of proper regulation and disclosure, resulting in incidents such as the FTX fraud (Q.ai, 2022). Some exchanges also attempted to gain an advantage in unethical and legally questionable ways (Rodgers [Forbes], 2019; Vigna [WSJ], 2019; BTI, 2019). One such market manipulation is wash trading, whereby investors simultaneously sell and buy the same assets to create artificial transactions, distorting price and hurting investor confidence and participation, as seen in other financial markets (Aggarwal and Wu, 2006; Cumming, Johan, and Li, 2011; Imisiker and Tas, 2018).

Against this backdrop, we provide the first systematic and rigorous study of misreporting and wash trading on cryptocurrency exchanges. Our goal when we conducted the investigation in 2019 was to rigorously establish that wash trading is a widespread and systemic issue for the entire industry, and to warn that centralized crypto exchanges can garner much market power and engage in harmful activities in the absence of regulatory scrutiny. Both issues have now occupied public attention, given what has transpired in the crypto market over the past year. By inspecting the distribution of trade size, we document wash trading on most unregulated exchanges in late 2019, which by our estimate inflates reported transaction volume by over 70% on average.[1] Furthermore, such misreporting and volume faking appear to improve the ranking and prominence of the exchanges, create short-term price dispersion across exchanges, occur more on newly established exchanges with smaller user bases, and have implications for long-term industrial organization, development,

[1] Wash trading is defined by the U.S. Commodity Exchange Act 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 other words, wash trading occurs when someone fabricates trades and acts as the counterparty on both sides to inflate volume. The definition of wash trading from the U.S. Commodity Exchange Act can be found at: *https://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/glossary_wxyz.html.*

and regulations. In fact, our research has contributed to the broad awareness of and actions for addressing the wash trading problem: regulators have increased scrutiny on wash trading and several ranking websites such as CoinMarketCap and CoinGecko also changed their matrices from purely volume-based to more sophisticated ranking models, allowing filtering fake volumes using methodologies similar or derived from ours.[2] The most recent SEC allegation against Binance includes a wash trading accusation consistent with our findings.

Wash trading on centralized crypto exchanges warrants our attention for several reasons. First, crypto exchanges play an essential role in the industry (e.g., Amiram, Lyandres, and Rabetti, 2022) by providing liquidity and facilitating price discovery just like traditional exchanges. Many crypto exchanges have expanded into upstream (e.g., mining) and downstream (e.g., payment) sectors, consequently wielding significant influence as a complex of trading platforms, custodians, banks, and clearinghouses. Second, because liquidity begets liquidity, crypto exchanges have strong economic incentives to inflate trading volumes to increase brand awareness and ranks on third-party aggregator websites or media (e.g., CoinMarketCap, CoinGecko, Bitcointalk, and Reddit), which in turn increases the exchanges' profits from transaction fees.[3] Third, while wash trading is largely prohibited in most financial markets and developed economies (IOSCO, 2000), cryptocurrencies are particularly susceptible to wash trading under limited regulatory oversight. Online Appendix A contains more institutional details of crypto exchanges.

While media and industry reports in 2018-2019 constituted whistle-blowers, they were often imprecise and speculative (Fadilpasic, 2019). Opinions on wash trading were divided, making practitioners and regulators unsure whether wash trading only concerned a few specific legal cases or was widespread.[4] We not only use rigorous statistical tools and intuitive behavioral benchmarks to demonstrate the existence of wash trading as an industry-wide phenomenon, but also contributes suggestive evidence of the efficacy of regulation in this industry, which has implications for investor protection and financial stability, not to mention that the findings are also relevant for ongoing lawsuits and empirical research on cryptocurrencies, which frequently rely on transaction volumes.

[2] Our research was consulted to assist federal investigations and legislation by the U.S. Department of Justice, the Securities Exchange Commission, New York State Office of Attorney General, and the Federal Bureau of Investigation.

[3] Individuals could wash trade as well. It is documented that traders use cryptocurrency and Non-Fungible Token wash trading to net millions of profits (Cong, Landsman, Maydew, and Rabetti, 2022; Quiroz-Gutierrez, 2022).

[4] For example, Bitwise Asset Management suggested to the SEC in 2019 potential wash trading on crypto exchanges (Fusaro and Hougan, 2019), but the allegations were denied by the exchanges (see, https://cryptonews.net/news/market/235179/ and https://blokt.com/news/alameda-research-bitwise-report-on-fake-bitcoin-trading-volume-inaccurate).

Finally, our research expands the applications of statistical and behavioral principles in forensic finance (Griffin and Kruger, 2023), with regulatory implications for FinTech and beyond.

Our first key finding is that wash trading is prevalent on unregulated exchanges but absent on regulated exchanges. To this end, we employ multiple methodologies that have been successfully applied in natural and social science fields and are unlikely to be affected by specific trading strategies, exchange characteristics, or specificities of the asset class. We also advocate combining the various approaches for noise reduction and robust manipulation detection.

Specifically, we examine the distribution of the first significant digit for transactions on each exchange against Benford's law—a well-known statistical benchmark in natural and social sciences, widely used to detect fraud in fields such as macroeconomics, accounting, and engineering (e.g., Durtschi, Hillison, and Pacini, 2004; Li, Cong, and Wang, 2004). We next utilize a behavioral regularity in trading: clustering at transaction sizes at round numbers. Transactions cluster at round numbers, such as multiples of 10 in the decimal system, because they are cognitive reference points in decision-making (Rosch, 1975). Rounding is frequently observed in finance (Kandel, Sarig, and Wohl, 2001; Chen, 2018; Kuo, Lin, and Zhao, 2015), analysts' forecasts (Clarkson, Nekrasov, Simon, and Tutticci, 2020; Roger, Roger, and Schatt, 2018), currency trading (Osler and Savaser, 2011), or LIBOR submissions (Hernando-Veciana and Tröge, 2020). Our third test explores whether the trade size distributions have fat tails characterized by power law, as found in traditional financial markets and other economic settings (e.g., Gabaix, Gopikrishnan, Plerou, and Stanley, 2003a). We consistently find anomalous trading patterns on unregulated exchanges, with higher ranked exchanges failing more than 20% of the tests and lower ranked exchanges failing more than 60%. Our findings remain robust in joint hypothesis tests.

We further quantify the fraction of fake volume by taking advantage of the rounding regularity. Illicit traders routinely employ program-generated fake orders with random order sizes to achieve scale without drawing attention (e.g., Vigna and Osipovich, 2018; Rodgers, 2019). Therefore, wash trades primarily generated by automated programs are likely to have low levels of roundness, i.e., a larger effective number of decimals for trades. Authentic trades can be unrounded due to algorithmic trading or other transaction needs. We thus establish a benchmark ratio (based on calculations from the regulated exchanges) of unrounded trades to authentic trades with round sizes. We then consider any unrounded trades on unregulated exchanges beyond this ratio as wash trades.

We find that the volume of wash trading is, on average, as high as 77.5% of the total trading volume on unregulated exchanges, with a median of 79.1%. In particular, wash trades on the twelve Tier-2 exchanges are estimated to be more than 80% of the total trade volume, which is still over 70% after accounting for observable exchange heterogeneity. These estimates, combined with the reported volume in Helms (2020), translate into wash trading of over $4.5 trillion in spot markets and over $1.5 trillion in derivatives markets in the first quarter of 2020 alone. To mitigate the influence of heterogeneity of traders and algorithmic trading strategies across various exchanges, we validate the roundness-ratio estimation and conduct several robustness tests to allay selection concerns.

We next study exchange characteristics that correlate with wash trading and investigate the impact of wash trading on market outcomes such as exchange ranking. Through proprietary data on historical rankings and trading volume information from CoinMarketCap, we discover that wash trading influences exchange ranking. Specifically, 70% of wash trading of total reported volume moves an exchange's rank up by 46 positions. An exchange's wash trading positively correlates with its cryptocurrency prices over the short term. Furthermore, exchanges with longer establishment histories and larger user bases wash trade less. Less prominent exchanges, in contrast, have short-term incentives for wash trading without drawing too much attention. Finally, wash trading is positively predicted by returns and negatively by price volatility.

While current business incentives and ranking systems fuel the rampant wash trading on unregulated exchanges, regulated exchanges, which have committed considerable resources towards compliance and license acquisition and face severe punishments for market manipulation (Perez, 2015), do little wash trading. We thus offer a concrete set of tools for regulation and third-party supervision in the crypto market for convincingly exposing wash trading and potentially combating non-compliant companies. The tests we introduce are not exhaustive, and wash traders may adjust their strategies in response to these tests. Nevertheless, our tools can still make transaction fabrications more difficult and regulation or litigation easier.

Our paper contributes to recent studies and regulatory debates on cryptocurrencies.[5] Amiram, Lyandres, and Rabetti (2022) is a closely related study that extends our framework to offer additional detection tools for wash trading, provides lower bounds using more recent data, and

[5] Cong, Li, and Wang (2020, 2022), Lyandres, Palazzo, and Rabetti (2022), Howell, Niessner, and Yermack (2020), and Cong and Xiao (2021) provide further institutional background on cryptocurrencies and ICOs; a number of articles discuss the role of crypto-tokens in fundraising and commitment (e.g., Goldstein, Gupta, and Sverchkov, 2019); studies such as Liu and Tsyvinski (2021) and Shams (2020) document empirical patterns in cryptocurrency returns. With respect to non-financial aspects of cryptocurrencies see Halaburda et al. (2020) for a discussion of cryptocurrencies' design and references therein.

analyzes how competition interacts with exchange operations. Aloosh and Li (2023) is a complementary study that validates our detection methodology by showing individual traders clear their own orders using account-level data leaked from the now-closed Mt. Gox exchange. Victor and Weintraud (2021) find that wash trading worth 159 million U.S. dollars exists on decentralized exchanges such as EtherDelta and IDEX. Two other studies, Le Pennec, Fiedler, and Ante (2021) and Chen, Lin, and Wu (2022), follow our study to analyze crypto wash trading. The former introduces alternative detection tools, utilizing web traffic and wallet data, while the latter develops a matrix combining off-chain and on-chain data to examine five exchanges. Finally, Cong, Landsman, Maydew, and Rabetti (2023) document crypto wash trading by individuals for tax-loss harvesting.

More broadly, our study belongs to the literature on manipulation and misreporting in finance.[6] Concerning cryptocurrency markets, Foley, Karlsen, and Putninš (2019) study the illegal usage of cryptocurrencies; Gandal, Hamrick, Moore, and Oberman (2018) and Griffin and Shams (2020) discuss manipulative behavior in Bitcoin and Tether; Li, Shin, and Wang (2020), among others, document pump-and-dump patterns in various cryptocurrencies; most recently, Choi and Jarrow (2020) discuss crypto bubbles caused by speculation or manipulation. These studies do not examine wash trading, which our unique and comprehensive dataset enables us to do.

Our study is also among the first to discuss the effects of regulation on crypto exchanges, filling a void in the literature and offering new insights into cryptocurrency regulation. We further contribute to the debates on market concentration, collusion, and regulation in the blockchain industry (e.g., Cong and He, 2019; Cong, He, and Li, 2020; Capponi, Olafsson, and Alsabah, 2022; Roşu and Saleh, 2020; Amiram et al., 2022) by highlighting another detriment of vertical concentration of the operational scope of crypto exchanges. Relatedly, Irresberg, John, and Saleh (2020) document that only a few blockchains dominate the public blockchain ecosystem. Without proper regulation and with vertical integration not seen in other markets, crypto exchanges may potentially engage in market manipulation or even outright fraud.

In terms of methodology, we enrich the use of and demonstrate the efficacy of statistical laws and behavioral principles for manipulation detection at scale in accounting and finance. Specifically, we are the first to apply Benford's law, trade-size clustering, and power law in FinTech and

[6] Our paper therefore contributes to forensic finance—the application of economic and financial knowledge to discover or substantiate evidence of criminal wrongdoing that meets standards in a court of law (e.g., Allen and Gale, 1992; Jarrow, 1992; Christie and Schultz, 1994; Ritter, 2008; Zitzewitz, 2012). Recently, blockchain forensics have been applied to on-chain data to study cybercrimes (e.g., Cong, Harvey, Rabetti, and Wu, 2022 and Cong, Gaur, Rabetti and Updegrave, 2023).

cryptocurrency studies. Our use of Pareto–Lévy distribution (instead of Zipf's law, as seen in Mao, Li, and Fu, 2015 and Prandl et al., 2017) for fraud detection is also novel in social sciences. Importantly, our findings imply that researchers using reported volumes by exchanges also need to heed the presence of wash trading and test the robustness of their conclusions.

The paper proceeds as follows. Section 2 describes our data and provides summary statistics. Section 3 presents the methodologies of wash trading detection and reports our empirical findings. Section 4 quantifies wash trading and details an array of tests to validate the methodology and demonstrate the robustness of the results. Section 5 relates wash trading to exchange characteristics, cryptocurrency returns, and exchange ranking, before discussing its implications for regulation and industry practice. Section 6 concludes. The online appendices provide supplementary results and discussion, development and regulatory status of cryptocurrency exchanges, a theoretical model of wash trading, further explanation of Benford's law as a forensic tool (available at https://www.dropbox.com/s/tg9bdms0fhlzz2g).

2 Data and Summary Statistics

We collect cryptocurrency transaction information on 29 major centralized exchanges from the proprietary database maintained by TokenInsight (*www.tokeninsight.com*), a company specializing in consulting, rating, and research reports for cryptocurrency-related businesses. The selection of these exchanges was based on their publicity (rank on third-party websites), representativeness, and API compatibility, including well-known ones like Binance, Coinbase, and Huobi, as well as many obscure ones.[7] Our data cover the period from 00:00 July 9th, 2019 to 23:59 November 3rd, 2019, and each transaction contains the exchange information, unique transaction ID, timestamp, price, amount of cryptocurrency traded, and trade pair symbol.[8] For each exchange, we focus on the four most widely recognized and heavily traded cryptocurrencies, Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), and Ripple (XRP), which represent over 60% of the volume and are available on almost all exchanges. The final sample contains 448,475,535 transactions. Other exchange-related variables such as aggregated trading volume, reputation metrics, and exchange characteristics (e.g., exchange age,

---

[7] There may be concerns that our data could disproportionately represent exchanges with a higher prevalence of wash trading. However, the dataset spans a wide range of ranks (1-300 among crypto exchanges), and as we will later illustrate, even lower-ranked exchanges are highly incentivized to engage in wash trading. Moreover, TokenInsight evaluates representativeness beyond mere rankings when choosing exchanges to analyze. Some exchanges that became prominent in subsequent years, such as FTX, had not yet been established and are not covered in our data.

[8] Since U.S. dollars (USD) are only allowed to exchange in three US regulated exchanges (Bitstamp, Coinbase and Gemini), digital dollars (symbol USDT, also known as 'stablecoins', which are designed to be pegged to the U.S. dollar) are commonly used as substitutes and widely accepted by most trading platforms. We treat cryptocurrency-USD pairs and cryptocurrency-USDT pairs as being the same in this study.

ranking, web traffic, etc.) are collected from official exchange websites and various data tracking and analysis platforms including SimilarWeb, Alexa, and CoinMarketCap.

The New York State Department of Financial Services (NYSDFS), a regulatory entity in New York State, was one of the first agencies to establish regulation over cryptocurrencies and led the world in developing the regulatory framework for the cryptocurrency industry. Hence, we categorize the three exchanges (Bitstamp, Coinbase, and Gemini) with BitLicense issued and supervised by NYSDFS as regulated exchanges.[9] The other 26 non-compliant exchanges are classified as unregulated and divided into 10 Tier-1 (including Binance) and 16 Tier-2 exchanges based on their web traffic, which reflects an exchange's user base and reputation and plays essential roles in customer acquisition and competition. Specifically, Tier-1 unregulated exchanges are the ones in the top 700 of the "SimilarWeb" website traffic ranking of the investment category during the sample period. [10] Our main findings are robust to using alternative regulatory frameworks around the world.[11]

[Insert Table 1]

Table 1 summarizes the characteristics of exchanges, including age, trading volume, and ranks from different metrics. The age for exchanges refers to the period from their establishment to July 2019. In Table 1, all the regulated exchanges have survived for at least five years. However, most unregulated Tier-2 exchanges were launched in 2017 and 2018, while Tier-1 exchanges are generally older.

In general, trade volume shows little correlation with our classification of exchanges: Some unregulated exchanges have much larger trading volumes compared with regulated exchanges. For example, Coinbene, an unregulated Tier-2 exchange, has a $50,944 million volume, while Coinbase's volume is only $15,212 million. The trading volume of different unregulated exchanges varies significantly. For instance, Exmo has only dozens of millions, while many unregulated exchanges exceed tens of billions.

[9] BitLicense requires an exchange to build a sophisticated compliance system, an anti-money laundering program, a capital control and custodian system, a record-keeping and customer identity system, an information security team, and a disaster recovery system, as well as to submit necessary documents for routine checks, which cost between 20,000 to 100,000 U.S. dollars even for compliant exchanges (Perez, 2015).
[10] Tier-2 unregulated exchanges in our sample all ranked lower than 960. This distinction of tiers does not affect any of our results since they are mostly at the exchange level..
[11] The Singaporean authority integrated crypto exchanges into the existing systems by requiring crypto exchanges to comply with the new Payment Services Act (https://www.mas.gov.sg/regulation/guidelines/ps-g02-guidelines-on-provision-of-digital-payment-token-services-to-the-public). The Swiss Financial Market Supervisory Authority (FINMA) issued several guidelines and ordinances to regulate Distributed Ledger Technology trading facilities and ICOs (www.finma.ch/en/authorisation/fintech/). The Japanese Financial Services Agency (FSA) the British Financial Conduct Authority (FCA) have also established their on crypto regulations.

Finally, we find regulated exchanges, especially Bitstamp and Gemini, rank behind many unregulated Tier-1 exchanges based on web traffic. Coinbase has the highest trading volume among regulated exchanges and a better rank under both ranking algorithms. Regarding CoinMarketCap ranks based on trading volumes, seven unregulated Tier-2 exchanges rank Top 20 and outperform the majority of unregulated Tier-1 and regulated exchanges. Although trading-volume ranks cannot fully represent the quality and liquidity of exchanges, it is used by most ranking agencies. Thus, cryptocurrency investors are likely to choose an exchange according to these trading-volume-based ranks. One would anticipate that unregulated exchanges, especially those launched later, are motivated to engage in wash trading to achieve higher rankings and acquire more customers.

3 Empirical Evidence of Wash Trading

We present empirical evidence of crypto wash trading entailing four major trading pairs (BTC/USD, ETH/USD, LTC/USD, and XRP/USD).[12] Specifically, we examine the properties of trade sizes on each exchange and test them against three well-established statistical and behavioral benchmarks. The use of multiple tests at the exchange level demonstrates the presence of wash trading on unregulated exchanges robustly. As these tests are grounded in fundamental principles, they are least likely to be influenced by heterogeneous (yet authentic) trading specific to individual traders and exchanges. We further control this when quantifying the extent of wash trading in the subsequent section. It is important to note that each exchange may engage in wash trading using its unique approach (if they do so). Our implicit assumption is that large-scale wash trades during our sample period were not specifically designed to comply with all three or even some patterns. Since wash traders can learn from our work and adjust, we do not claim that these detection tools will remain effective indefinitely. However, without extensive coordination, it may be challenging for traders to fabricate transactions that pass all three tests simultaneously.

3.1 Distribution of First Significant Digits

Benford's law describes the distribution of the first significant digit in various naturally generated data sets, deriving from the intuition that many systems follow multiplicative processes (e.g., Li, Cong, and Wang, 2004). According to Benford (1938):

$$\text{Prob}(N\ is\ the\ first\ significant\ digit) = \log_{10}\left(1 + N^{-1}\right),\ N \in \{1,2,3,4,5,6,7,8,9\} \tag{1}$$

[12] LTC/USD data is not available on Liquid, Bgogo, Lbank, and Exmo. XRP/USD data is not available on Gemini, Bgogo and Lbank.

The probability of 1 being the first significant digit is 30.10%. Digits 2 and 3 have probabilities of 17.60% and 12.50%, respectively. The probabilities of the rest (9.7%, 7.9%, 6.7%, 5.8%, 5.1%, and 4.6%, respectively) decrease as the digit increases. In Online Appendix B, we provide a theoretical derivation of the distribution of Benford's law and validate it for detecting wash trading through simulation.

In this section, we report whether the first-significant-digit distribution of transactions (denominated in the cryptocurrencies in question) conforms to the pattern implied by Benford's law (as shown in equation 1) on the 29 exchanges.[13] Inconsistency with Benford's law suggests potential data manipulations.

Figure 1 illustrates the first-significant-digit distributions for four cryptocurrencies with one regulated exchange and four unregulated exchanges. The five exchanges are the ones that fail the most tests in their categories and are consistently chosen throughout the paper for concise illustration. The distributions for the rest of the exchanges exhibit similar patterns and are shown in Online Appendix C. Bars show the fraction of transactions in which the trade size has integer *i* as the first significant digit, and dots represent the frequency distribution implied by Benford's law.

[Insert Figure 1]

For Coinbase, 32.75% of BTC trades and 32.73% of ETH trades have "1" as the leading digit, consistent with the benchmark frequency of 30.10% in Benford's law. Unregulated exchanges such as Fcoin and Exmo violate Benford's law, with some first significant digits occupying a disproportionally large fraction, fitting our assumption of incentivized wash trading campaign, which Fcoin and Exmo both offered in different formats.[14] Violations of another assumption, i.e., exchange faking trade orders, can also be found in unregulated exchanges, such as Biki.

The first-significant-digit distributions of all regulated exchanges comply with Benford's law regardless of the type of cryptocurrency. Half of the unregulated exchanges, including Tier-1 and Tier-2, exhibit apparent discrepancies with Benford's law in at least one type of cryptocurrency. Disconformity with Benford's law is observed on nine unregulated Tier-2 exchanges, among which seven violate the law in at least two cryptocurrencies.

[13] Benford's law is most widely known and applied to examine the first-significant-digit distribution of a data set. The law also makes predictions about the distribution of second digits, third digits, digit combinations, and so on. Here in this research, only the first-significant-digit part of Benford's law is applied to avoid interference from other behavioral biases.

[14] https://info.exmo.com/en/platform-features/what-is-exmo-coin/ and https://www.coindesk.com/markets/2018/06/22/fcoin-crypto-exchange-draws-fire-for-controversial-business-model/

To quantitatively assess whether first-significant-digit distributions conform with Benford's law, we employ Pearson's Chi-squared test. The value of the test statistic is calculated as

$$\chi^2 = \sum_1^9 \frac{(O_i - E_i)^2}{E_i} \tag{2}$$

in which O_i is the observed distribution frequency, and E_i represents the value from Benford's law. In our study, trading data contains millions of observations in each trading pair of each exchange. The standard Chi-square test will not work properly with a large sample size (Bergh, 2015). We utilize the nominal value of distribution frequencies of the first significant digits to form a contestant approach through different subsamples. The null hypothesis is that the first-significant-digit distribution observed in an exchange's trading data is consistent with that of Benford's law.

[Insert Table 2]

As seen in Table 2, trades of regulated exchanges follow Benford's law, and so do those on most of the unregulated Tier-1 exchanges. However, patterns for Bitfinex are inconsistent with Benford's law in BTC and XRP trades, with a significance level of 1%. Moreover, five Tier-2 exchanges (DragonEX, Mxc, Fcoin, Exmo, and Coinegg) significantly diverge from Benford's law in most cryptocurrencies. Other unregulated exchanges show sizable differences in several cryptocurrencies. For example, Liquid violates Benford's law in BTC at a 5% level; Biki and Coinmex fail in BTC and XRP at a 1% confidence level, respectively; Biki and BitZ fail at a 5% confidence level in ETH.

Overall, all regulated exchanges have transactions described by Benford's law. Meanwhile, 20% of unregulated Tier-1 exchanges violate Benford's law in at least one cryptocurrency at a 5% significance level, and 50% of Tier-2 exchanges fail to follow Benford's law in at least one cryptocurrency. The extent of violation observed here for a widely recognized forensic rule such as Benford's law may be surprising. However, there are a few reasons for this. First, unregulated exchanges might not consider the violation of Benford's law a serious issue in a largely unregulated industry. Faced with more pressing accusations, such as market manipulation and fraud, newly established small exchanges are unlikely to allocate resources to cover up wash trading traces. Some unregulated exchanges even publicly promoted incentivized wash trading programs. Second, these exchanges may not be prepared for financial forensic tools like traditional institutions are. At the time we collected the data in 2019, Benford's law had not been extensively applied in the cryptocurrency industry. In the years since our initial draft, the situation has changed, and some exchanges might have incorporated Benford's law into their wash trading strategies.

The clustering effect in trade-size distribution histograms on sample exchanges. Two sets of observation ranges are applied for each trading pair: 0-1 BTC, 0-10 BTC, 0-10 ETH, 0-100 ETH, 0-100 LTC, 0-1,000 LTC, 0-10,000 XRP, and 0-100,000 XPR. In each histogram, we highlight every 5th and 10th bin to illustrate the clustering effect.

3.2 Trade Size Clustering

We next examine the presence of clustering—traders' tendencies to use round trade sizes and round prices—within crypto exchanges. Clustering is a classic behavioral regularity frequently observed in financial markets. Grossman et al. (1997) propose a theory that the clustering in competitive markets with assets valued with great precision, such as NYSE, AMEX, and LSE is due to traders' attempts to minimize price negotiation and transaction costs. Clustering is also explained by psychology: authentic traders use round numbers as cognitive reference points (Rosch, 1975) to simplify and save effort in the decision-making and evaluation (Ikenberry and Weston, 2008; Kuo et al., 2015; Lacetera, Pope, and Sydnor, 2012), distinguishing authentic trades from algorithmic trades (Mahmoodzadeh and Gençay, 2017; O'Hara, Yao, and Ye, 2014). Wash traders typically use automated trading programs, particularly when fake orders feature small transaction size yet substantial aggregate amounts (Vigna and Osipovich, 2018; Rodgers, 2019). As a result, wash trading inherently reduces the proportion of authentic volume and, thus clustering.

As most cryptocurrencies can be traded in fractions, and certain currencies possess larger unit values (particularly BTC), we establish, for the remainder of this paper, the smallest unit (base unit) as one unit in a specific decimal place value in proximity to one U.S. dollar. For instance, with the price of Bitcoin fluctuating between $8,000-$10,000 in our sample period, most BTC/USD orders are below 1 BTC. In this context, round numbers in traditional financial markets such as 10, 100, or 10,000 are too substantial for individual traders. Considering the value of 10^{-4} BTC is within the order of magnitude of one U.S. dollar, it is deemed the base unit in this study. Similarly, the base units of other currencies are 0.001 ETH, 0.01 LTC, and 1 XRP, respectively.

We now examine whether trade-size clustering appears at multiples of 100 base units for each cryptocurrency.[15]

[15] We focus on clustering in terms of round numbers in the number of tokens instead of dollar amounts because our data contains the number of tokens traded and its product with token price is typically not equal to the actual dollar amount traders use in their orders due to exchange fees. For a few exchanges that we can obtain the time series of fees, we verify that our results to be robust to the alternative specification using dollar amounts.

The clustering effect in trade-size distribution histograms on sample exchanges. Two sets of observation ranges are applied for each trading pair: 0-1 BTC, 0-10 BTC, 0-10 ETH, 0-100 ETH, 0-100 LTC, 0-1,000 LTC, 0-10,000 XRP, and 0-100,000 XPR. In each histogram, we highlight every 5th and 10th bin to illustrate the clustering effect.

Figure 2 depicts trade size distributions of representative exchanges in two observation ranges for BTC, ETH, LTC, and XRP, highlighting the clustering effect at round sizes. Online Appendix D presents the histograms of the remaining exchanges.

[Insert

Figure 2*]*

Trade data from regulated exchanges display a downward-sloping curve with prominent peaks at multiples of 5,000 base units in the range of 0-10 BTC (e.g., 0.5 BTC, 1 BTC, 1.5 BTC, 2 BTC, etc.). Similar patterns also appear in distributions of ETH, LTC, and XRP. These findings suggest the presence of trade size clustering on regulated crypto exchanges, consistent with the trade patterns in regulated financial markets, which display a downward trend because large orders are less frequently placed and executed, as well as a trade size clustering effect (e.g., Moulton, 2005; Alexander and Peterson, 2007; ap Gwilym and Meng, 2010; Mahmoodzadeh and Gençay, 2017; Verousis and ap Gwilym, 2013).

On the other hand, trade size distributions of unregulated exchanges demonstrate some abnormal patterns. Taking KuCoin in

Figure 2 as an example, it does not show a clear clustering pattern in round number trades. Besides, most trades of KuCoin are concentrated at small sizes and display an anomalous drop in frequency, especially in LTC and XRP trades. Moreover, clustering patterns for different assets vary across crypto exchanges and have shown no overall pattern. On unregulated Tier-2 exchanges, we observe less apparent clustering at round sizes, and trade patterns vary dramatically, deviating from the typical downward distribution. For instance, when examining larger ranges, trade frequency on Fcoin does not monotonically change with the increase in trade size for all cryptocurrency trades. Similar issues are observed on other exchanges (refer to Online Appendix D, e.g., DragonEX, Mxc, and Digifinex in BTC trades; BitZ, Mxc, Bibox, and Digifinex in ETH trades). Additionally, abnormal patterns such as gaps, cliffs, scarce peaks, and uniform distributions can be observed in unregulated exchanges, which are contrary to the behavioral regularity in financial markets.

To quantify the effect of trade-size clustering, we conduct the Student's *t*-test for each crypto exchange by comparing trade frequencies at round trade sizes with the highest frequency of nearby unrounded trades. The value of the test statistic is calculated as:

The clustering effect in trade-size distribution histograms on sample exchanges. Two sets of observation ranges are applied for each trading pair: 0-1 BTC, 0-10 BTC, 0-10 ETH, 0-100 ETH, 0-100 LTC, 0-1,000 LTC, 0-10,000 XRP, and 0-100,000 XPR. In each histogram, we highlight every 5th and 10th bin to illustrate the clustering effect.

$$t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}} \tag{3}$$

where $\overline{x}$ is the average of rounded trade frequencies minus unrounded trade frequencies, s is the sample standard deviation, and n is the sample size. The null hypothesis of the test is that the difference between frequencies at rounded numbers and nearby unrounded trades is zero.

For each trading pair, we set up two sets of observation windows: windows centered on multiples of 100 units (100X) with a radius of 50 units (100X-50, 100X+50), and windows centered on multiples of 500 units (500Y) with a radius of 100 units (500Y-100, 500Y+100). Trade frequency is calculated as the number of transactions with size *i* over total transaction numbers in the observation window. For example,

Figure 3 illustrates that in BTC transactions on Bitstamp, the observation window of around 200 units (0.02 BTC) ranges from 150 units to 250 units (0.015 - 0.025 BTC). Orders at 0.02 BTC constitute 16.42% of the total within the entire observation window, while the highest trade frequency of unrounded orders is only 2.54%. The apparent difference indicates that orders within the 0.015-0.025 BTC window cluster at 0.02 BTC.

[Insert

Figure 3 *and Table 3]*

Table 3 presents the results of *t*-test for size clustering on regulated exchanges (Panel A), unregulated Tier-1 (Panel B), and Tier-2 exchanges (Panel C). As anticipated, trade frequency at round sizes is significantly higher than at unrounded sizes across all three regulated exchanges, regardless of the cryptocurrencies and observation ranges examined. This finding aligns with the observations in

Figure 2. Additionally, in terms of difference and *t*-statistics, size clustering is more evident at multiples of 500 units. For example, when looking at BTC trades on Bitstamp, the difference in frequency is 9.1% in trade size of multiples of 100 units (e.g., 0.01 BTC, 0.02 BTC, and 0.03 BTC) while the difference is 20.3% at sizes that are the common multiples of 500 units (e.g., 0.05 BTC, 0.10 BTC, 0.15 BTC).

Like regulated exchanges, three unregulated Tier-1 exchanges (Bitfinex, Liquid, and Poloniex) show positive and significant differences at a 1% level in trades of all available cryptocurrencies (except for XRP on Poloniex, which is significant at 5%). Trade clustering also appears more frequently at multiples of 500 units. For example, six Tier-1 exchanges (Binance, Bitfinex, Huobi, Liquid, Okex, and Poloniex) display noticeable clustering effects at multiples of 500 units for all four cryptocurrencies. However, KuCoin and Zb show insignificant differences in frequencies between round and unrounded trades.

In contrast, clustering at round sizes is largely absent on unregulated Tier-2 exchanges. Half exchanges exhibit no sign of clustering for all cryptocurrencies in both observation windows. Except for Bitmax, all Tier-2 exchanges have no clustering in at least one cryptocurrency. Besides, on some exchanges, trade clustering becomes less evident at a higher level of roundness (multiples of 500 units). For example, on BitZ and DragonEX, frequencies at multiples of 100 units are higher (significantly at a 1% level), but clusters at multiples of 500 units are insignificant.

We also regress the (logit) percentage of trades at certain sizes on various dummy variables, which are set to one at round sizes. Online Appendix E reports the consistent findings.

In summary, we observe that regulated exchanges display an evident clustering effect in transaction size, whereas unregulated Tier-1 and Tier-2 exchanges exhibit little clustering, with 30% and 50% of exchanges showing no trade-size clustering in all cryptocurrencies, respectively. It is important to note that clustering involves rounding off the last non-trivial digits and affects little the distribution of the first significant digit. As we are only applying the plain-vanilla Benford's law concerning the distribution of the first significant digits (not the first two or three significant digits), there is no interference with the clustering tests.[16]

3.3 Tail Distribution

We next examine the tails of trade-size distributions on each crypto exchange. In economics and finance, power law has been found to capture the "fat tails" of many distributions.[17] Mathematically, power-law distribution has a cumulative density function (CDF):

$$\mathrm{P}(\mathrm{X} > \mathrm{x}) \sim \mathrm{x}^{-\alpha}, \tag{4}$$

[16] Rounding could lead to violations of Benford's law for later digits, yet the first significant digits still follow Benford's law, as seen in multiple forensic applications (e.g., Carslaw, 1988; Thomas, 1989).

[17] For example, power-law distribution of tails can be found in Pareto distribution of income (Pareto, 1896), the distribution of stock returns (Gopikrishnan et al., 1999), trade size (Gopikrishnan et al., 2000), and share volume (Plerou et al., 2000), fluctuations in foreign exchange markets (Da Silva, Matsushita, Gleria, and Figueiredo, 2007; Ohnishi et al., 2008; Vandewalle, Ausloos, and Boveroux, 1997), and cryptocurrency transactions (Li et al., 2019; Schnaubelt et al., 2019)..

where α is known as the power-law exponent or the tail exponent. When using the probability density function (PDF), the relevant parameter is $\alpha_{pdf} = \alpha + 1$.

One explanation for the power-law tails in financial data is the trading behavior of large investors who try to avoid large price impacts in the markets (Gabaix, Gopikrishnan, Plerou, and Stanley, 2003a). Other studies attribute them to the investors' limited information on the value of assets (Kostanjčar and Jeren, 2013; Nirei, Stachurski, and Watanabe, 2020) and herding (Nirei, Stachurski, and Watanabe, 2020). Specifically, for trade volume, several previous studies (Maslov and Mills, 2001; Gabaix et al., 2006; Plerou and Stanley, 2007) found that in stock markets, trading volume distribution follows the power law with exponent $\alpha \cong 1.5$. For theory, Gabaix et al. (2006) propose a model derived from trading strategies by large institutional investors. Intuitively, institutional investors trade as much as possible while avoiding price impact to their robustness constraint. Given that fund sizes follow Zipf distribution, presumably from the random growth of funds, trade size conforms to the power-law distribution with an exponent of 1/2. These conditions likely apply to cryptocurrency markets too, i.e., cryptocurrency transaction sizes are highly likely to conform to a power law.

To examine trade size distribution tails, we use two widely adopted techniques: The first is to take the logarithm of the empirical probability density function and fit the log-log data to power-law distribution by Ordinary Least Square (OLS). The second is to apply the Maximum Likelihood Estimation approach (MLE) and use the Hill estimator $\widehat{\alpha}_{Hill}$ for the data fitting. Hill estimator is asymptotically normal and calculated as follows (Clauset, Shalizi, and Newman, 2009; Hill, 1975):

$$\widehat{\alpha}_{Hill} = 1 + n\left(\sum_{i=1}^{n} ln\frac{x_i}{x_{min}}\right)^{-1} \quad , \tag{5}$$

where n is the number of observations and x_{min} is the cut-off threshold. The distribution yields to power law after x_{min}. The cut-off x_{min} , which signifies the start of the tails, is set as the top 10% of the largest trades during the sampling period. In this study, trade size distributions are constructed for empirical probability density functions using logarithmic spacing, and the Python package "powerlaw" (Alstott, Bullmore, and Plenz, 2014) is applied to fit the data and calculate the exponent.

Gabaix, Gopikrishnan, Plerou, and Stanley (2003b) show that theoretically and empirically, stock trade size follows a half cubic law ($\alpha = 1.5$). Various studies on trading volumes or sizes have shown that the vast majority of tail exponents lie in the Pareto–Lévy regime ($1 < \alpha < 2$) for traditional financial assets and bitcoins (Li et al., 2019; Schnaubelt et al., 2019).[18] We thus check whether the values of exponent α in the fitted results fall within the Pareto–Lévy range.

Table 4 presents the results from OLS and MLE fittings for four cryptocurrency trades. We can visually inspect the goodness of fit and identify whether crypto exchanges display a power-law tail in trade size distribution, as shown in Figure 4.

[Insert Table 4 and Figure 4]

As anticipated, both scaling estimators $\widehat{\alpha}_{OLS}$ and $\widehat{\alpha}_{Hill}$ lie in the Pareto–Lévy regime on regulated exchanges, indicating a stable power-law decay in all cryptocurrency transactions. Similar patterns are observed on half of the unregulated Tier-1 exchanges. HitBTC and KuCoin have estimators that fall outside the Pareto–Lévy range for all cryptocurrencies, suggesting inconsistency with power-law exponents for trade sizes in traditional markets. Furthermore, tail exponents for Liquid, Okex, and Zb display inconsistency for one cryptocurrency data.

On unregulated Tier-2 exchanges, only three exchanges show estimated exponents within the Pareto–Lévy range, whereas 62.5% show statistical evidence of disconformity to parameters of empirical regularity in all cryptocurrencies. Among the remaining samples, Bitmart shows fitted exponents within the ranges for both LTC and ETH transactions. Coinegg displays a pattern consistent with the range for LTC, while Gateio does so for ETH.

Figure 4 displays the probability density for trade size and the fitted power-law distributions on log-log plots, with one regulated and four unregulated exchanges as representatives for brevity. Online Appendix F contains figures for the rest.

As in mainstream financial markets, transactions from regulated exchanges display a downward linear trend in log-log plots and appear visually fit with power-law distributions. For instance, in Panel Coinbase of Figure 4, empirical data points fall around the fitted lines without obvious outliners, implying that trades in regulated exchange generally follow power law in all four listed cryptocurrencies. In general, the OLS line fits equally in the whole range, while MLE estimation weighs more at the start of the tail, where the probability value is higher. Consistent with regulated exchanges, 90% of unregulated Tier-1 exchanges resemble power-law tails in trade size distributions.

[18] Gopikrishnan et al. (2000) find that the power-law exponent of trade volume is around 1.5 in U.S. equity market. Plerou and Stanley (2007) investigate trades in New York Stock Exchange, London Stock Exchange and Paris Bourse and show that trade size in all three markets display the power-law decay with exponent in the range from 1 to 2. Moreover, value of exponents is not affected by industry and market capitalization. Note that Mandelbrot (1960) propose that income follows the stable "Pareto–Lévy" distributions with $1 < \alpha < 2$.

Straight lines estimated by OLS and MLE are roughly fitted to the data. Conversely, KuCoin (shown in Figure 4) shows a curvy shape in tails and fails to show the power-law distribution in the trade size.

On unregulated Tier-2 exchanges, tail distributions exhibit a variety of behaviors and reveal irregular patterns across exchanges and cryptocurrencies. Four Tier-2 exchanges (Lbank; Bitmax; Digifinex; Gateio) show a linear decrease in the tail zones and comply with the power-law tail. Exmo (as shown in Figure 4) displays a good linear fit but is inconsistent with the MLE method. On Fcoin, data points are dispersed in the tails of BTC, ETH, and LTC trades; additionally, a curvy shape is observed on the logarithm scale in BTC and XRP trades. In Coinegg's BTC samples, the tail appears to be flat, with some outliers. The ETH, LTC, and XRP graphs of Coinegg show step-like decay patterns.

The findings suggest that regulated exchanges behave as power law predicts, with estimators consistent with the Pareto–Lévy exponents in mainstream financial markets. But only half of Tier-1 exchanges display a power-law tail with exponents characterized by the Pareto–Lévy regime in all cryptocurrencies. In contrast, 75% of Tier-2 exchanges fail to follow.

3.4 Selection Concern, Multi-hypothesis Testing, and Conclusive Evidence

So far, we have conducted three independent statistical analyses for each cryptocurrency on every crypto exchange, including the Chi-squared test for Benford's law distribution, the *t*-test for trade-size clustering, and the linear fit for power law. The results are consistent for each category (regulated, unregulated tier-1, and tier-2) and for most exchanges. Overall, more than half of the unregulated exchanges fail at least half of all tests at the 5% significance level. Except for Bitmax, Tier-2 exchanges fail at least 30% of the tests, with ten exchanges failing more than 65% of all the tests. At the cryptocurrency level, unregulated exchanges as a whole fail more than 40% of the tests for each cryptocurrency. In contrast, regulated exchanges pass all the tests. These findings align with Prediction 1 of the theoretical model (Online Appendix G), indicating that exchanges under regulations are less likely to engage in wash trading than those without regulation. Furthermore, consistent with Prediction 2, companies in jurisdictions with higher fines (i.e., U.S.) are less likely to wash trade compared to ones with a lower level of fine (i.e., Japan). Notably, the U.S. is a major country that regularly establishes enforcement actions and issues fines against cryptocurrency exchanges (Blandin et al., 2019; Robinson, 2021).

Because multiple statistical tests may increase the possibility of Type I error and raise the concern of *p*-hacking, we perform a multiple (global) hypothesis test for robustness. The details and the results of the test can be found in Online Appendix H. The results are consistent with our findings in previous subsections. Trade patterns of all regulated exchanges show insignificant differences from those of traditional financial markets. Tier-1 unregulated exchanges have lower proportions in rejecting null hypotheses than Tier-2 ones in all cryptocurrencies. 75% of the Tier-2 unregulated exchanges fail to follow the universal law or trade patterns of traditional financial markets. In addition, BTC has the highest failure rates, followed by XRP. Furthermore, more unregulated exchanges fail the joint tests than individual tests in all cryptocurrency pairs. Some fraudulent exchanges may "luckily" display similar trade distribution as traditional markets in certain aspects but fail to satisfy all regularities, therefore leading to higher failed percentages in multiple hypothesis tests.

We also advocate combining the various detection approaches because every exchange may engage in wash trading differently and just due to randomness, an individual test may contain Type I and Type II errors. For example, Aloosh and Li (2023) find that using Power-law tail distribution does not detect wash trading on Mt. Gox during an earlier episode. This does not invalidate the approach and may be explained by the fact that in the early days wash trades are characterized by a large quantity of small transactions as the authors document or the high volatility of Bitcoin prices.

One might be concerned that traders and algorithms are unique or select-selected on unregulated exchanges. However, it has been documented that trading algorithms are generally exchange-agnostic (Alameda, 2019). Furthermore, according to PwC (2022), institutional investors choose trading venues primarily based on exchange liquidity and opportunities, rather than their regulatory status. We find no significant difference regarding the volume and distribution of transactions on regulated exchanges compared to unregulated exchanges when they became regulated. For example, Coinbase received its BitLicense in 2017 with no exodus of traders. In fact, its trading volume increased significantly since.

While these may not completely rule out traders' self-selecting into trading on regulated versus unregulated exchanges, we further appeal to the power of Benford's law and power law to allay our concern. If institutional investors or algorithmic traders systematically prefer regulated or unregulated exchanges, it would bias our findings towards seeing different tail distributions and less rounding on those exchanges. However, power-law distributions for the tail can have different cut-offs, so having disproportionally large trades is unlikely to affect the general shape of the tail (it could affect the exponent parameter). Moreover, whether one transacts large amounts or uses algorithms should not affect the first-significant-digit distribution because Benford's law is robust to changing accounting units and rounding behavior. Therefore, the three tests complement one another.

4 Quantifying Wash Trading

We now quantify the extent of wash trading by directly estimating wash trading volume. We also conduct several robustness and validation tests and provide alternative wash trading metrics.

4.1 Trade-size Roundness and Benchmark Roundness Ratio

Authentic human trades tend to have round sizes. In contrast, unrounded trades typically relate to programmed trading for various purposes such as market marking, high-frequency arbitration, and in particular, wash trading. Strong evidence suggests that most wash trading is done by bots, which can be easily added to the trading structure scripted by simple Python programs (e.g., Vigna and Osipovich, 2018; Rodgers, 2019), making round/unrounded trades reasonable proxies for authentic/fake orders.

We first show that levels of roundness for trade sizes differ across unregulated and regulated exchanges. The level of roundness is a qualitative parameter describing the decimal or integer places of the last non-zero digit. For instance, 1.01 BTC has a higher level of roundness than 2.123 BTC; 100 ETH has a higher level of roundness than 1234 ETH.[19] Authentic trades should display a higher level of roundness in size than artificial ones. We thus expect regulated exchanges to present a higher level of roundness in trade sizes compared with unregulated exchanges. For each crypto exchange, we analyze the trade-size distribution over levels of roundness (ten thousands, thousands, hundreds, tens, ones, tenths, hundredths, etc. base units). We then compare the regulated versus unregulated exchanges.

[Insert Table 5]

Table 5 shows that Tier-1 exchanges have significantly large Chi-squared statistics in at least one cryptocurrency. Unregulated Tier-2 exchanges, except for Mxc in BTC trades, show different roundness distributions from regulated exchanges with a 1% significance level for nearly all cryptocurrencies. Evidently, unregulated exchanges, especially unregulated Tier-2 exchanges, have a lower level of roundness in trade size relative to the regulated exchanges.

[19] For 1.01 BTC, the place value of last non-zero digit (1) is hundredths, while the place value of last non-zero digit (3) is thousandths in 2.123 BTC. In 100 ETH, the place value of last non-zero digit (1) is hundreds while the place value of last non-zero digit (4) is ones in 1234 ETH.

Assuming that the computer-based legitimate (non-wash) trades on unregulated exchanges have the same sensitivity to the authentic trading strategies and exchange characteristics as those on regulated exchanges, we estimate the legitimate number of unrounded trades for unregulated exchanges. The difference between the observed unrounded and legitimate trading volume is then a reasonable proxy for wash-trading volume. Because one can rarely label wash trades at an exchange without detailed information about the traders, our method provides a general way of estimating systematic wash trading that can be time-varying, therefore serving as a first-order benchmark.

From our earlier analysis, we detect no systematic wash trading on regulated exchanges. This is corroborated by the fact that round trades constitute around 30% of total trades on regulated crypto exchanges, which is consistent with patterns in the U.S. equity markets that are approximately free of wash trading due to regulation (Gomber, Gsell, Pujol, and Wranik, 2009; Tabb, Iati, and Sussman, 2009). We carry out a "cross-validation" test using any two regulated exchanges as the no-wash-trading benchmark to estimate the wash trading amount on the remaining regulated exchange. We found the wash trades estimated on average constitute less than 5% of the reported volumes.

4.2 Estimated Volume of Wash Trades

We estimate the wash-trading volume by calculating the abnormal proportion of unrounded trades. Specifically, we categorize trading volumes into round and unrounded ones by checking if the last non-zero digit of a certain trade size is less than 100 basis units. We then perform a pooled regression to estimate the ratio of (log) unrounded volume to (log) round volume at a weekly frequency:

$$\ln\left(V_{Unrounded_{it}}\right) = \alpha + \beta * \ln\left(V_{Round_{it}}\right) + \gamma * X_{it} + \epsilon_{it}\,, \tag{6}$$

where $V_{Unrounded_{it}}$ and $V_{Round_{it}}$ are unrounded and round trading volumes of regulated exchange *i* at week *t*, respectively. In the baseline, we exclude exchange-level controls by setting X_{it} to zero. To mitigate the concern that heterogeneous authentic algorithmic trading on various exchanges drives the estimates, we include a vector of exchange characteristics, X_{it} including age, rank, CoinMarketCap web traffic percentage, and unique visitors, in an alternative specification. We employ the parameters in (6) to calculate the legitimate (non-wash) unrounded trades of unregulated exchanges using their corresponding round trades. Wash trade volume is thus calculated as the non-negative amount by which the total unrounded trades exceed legitimate unrounded trades.

[Insert Table 6]

Table 6 presents the simple averaged and volume-weighted wash trading percentage for each exchange category and the exchange-level wash trading percentage by four cryptocurrency pairs. The results using models with or without controls are similar. Since some exchanges have missing data on the control variables and the residual standard errors in the model without controls are comparable to those with controls (meaning out-of-sample predictability is comparable), we only report the results using estimates from the model without controls for simplicity in subsequent analyses on price impacts, ranking, and so on. Standard deviations of wash trading volumes from bootstrapping the sample 1000 times are also included in the table.

On average, wash trades account for over 70% of the total trading volume on each unregulated exchange and approximately 61% even after controlling for exchange characteristics. Wash trades represent 53.4% of Tier-1 and 81.7% of Tier-2 exchanges' volume. Given that the four cryptocurrencies we examine dominate transaction volumes on all the exchanges, these figures are informative even without including all cryptocurrencies. It is also noteworthy that for all unregulated exchanges, an estimated 77.5% of the total reported volume appears to be wash trades. Our estimates are in the same order of magnitude as the estimates from the Wall Street Journal and industry reports (Rodgers, 2019; BTI, 2019), which are in the range of 67% to 99%. For example, the BTI Summary of Market Surveillance report discovered that, as of April 2019, 17 of the top 25 exchanges listed on CoinMarketCap contained over 99% fake volume. Our estimates are lower because exchanges might have taken actions since those earlier estimates were published---the Lucas critique applies.

4.3 Further Validation and Robustness Tests

Some may be concerned that heterogenous traders and thus their strategies across crypto exchanges could distort our estimation of wash trade. To alleviate the concerns, we use Benford's law and power law to test if our estimation (Section 4.2) predominantly captures wash trading. The results in Online Appendix I indicate the roundness-based estimation to be unaffected by authentic algorithmic trades.

That said, we provide complementary metrics that should help convince the readers that wash trading on unregulated exchanges is rampant and economically significant. One shows the extent of an exchange's wash trading by summarizing results of statistical tests in Section 4, grouped by exchanges and cryptocurrencies separately. Details can be found in Online Appendix J. In addition, we compare the trade size distribution of unregulated exchanges to regulated exchanges for robustness (Online Appendix K). Then, we examine an alternative method to gauge the extent of wash trading using Benford's law in Online Appendix L. At last, we discuss existing industry reports and why our methodologies are likely to be more robust and superior in Online Appendix M.

5 Wash Trading Incentives, Impacts, and Implications

We now discuss the potential drivers and implications of crypto wash trading, starting with the incentives for wash trading and how it affects the ranking of crypto exchanges. We then analyze the characteristics of exchanges that portend wash trading, and explore wash trading's impacts on crypto asset prices, before examining its regulatory and industrial ramifications.

In traditional markets, wash trading is typically conducted by individual traders rather than platforms. However, individual wash traders alone cannot fully explain the differences observed between regulated and unregulated exchanges. While the cost of wash trading for individuals should be associated with fees and bid-ask spreads, there is no systematic correlation found between the extent of wash trading and these variables. In contrast, evidence abounds that exchanges wash trade either directly or indirectly. For instance, top executives at some exchanges are known to trade on their own platforms while managing cryptocurrency hedge funds (e.g., Bitfinex'ed, 2017). Additionally, multiple companies have also pleaded guilty to direct wash trading (Sinclair, 2020). Moreover, exchanges can facilitate wash trading indirectly through fee rebate programs that incentivize their customers to engage in such activities. For example, Fcoin rewards platform tokens for trade mining, where more FT tokens are earned by trading more.

5.1 Wash Trading and Exchange Ranking

Exchanges' profit crucially depends on brand awareness and website traffic for customer acquisition, both of which heavily rely on public rankings. We utilize the proprietary, high-frequency data on exchange ranks and reported trading volumes from CoinMarketCap.com, which most exchanges rely on for referral traffic.[20] To study the incentives for wash trading by crypto exchanges, we first verify the ranking rule of CoinMarketCap using the daily rankings and reported volumes of more than 260 crypto exchanges. Spearman rank-order correlation coefficient is estimated to measure the rank correlation between trade volume and ranking in the CoinMarketCap. The coefficient is -0.995, approaching -1, indicating that ranks and volume are perfectly negatively related (see Figure 5). The

[20] For instance, according to SimilarWeb reports, CoinMarketCap contributes 65% of web traffic to one regulated exchange in our sample. It serves as the leading referral website and contributes most of traffics to 20 unregulated exchanges. Furthermore, 17 of these unregulated exchanges receive more than 30% of their total web traffic from CoinMarketCap referrals.

rankings of CoinMarketCap are determined by the trade volume of crypto exchanges. Exchanges with larger volumes would rank higher and gain more visibility and visits.

[Insert Figure 5]

Our findings support the intuition that to survive the fierce competition, many crypto exchanges naturally wash trade to gain prominence and market share so that the exchange can generate higher profits.[21] Indeed, from Figure 6, we observe that a 70% wash trading volume can move the rank of an exchange up by more than 25 positions relative to its rank in a world without wash trading.

[Insert Figure 6]

5.2 Price Impacts of Wash Trading

In Table 7, we examine the effect of wash trading on cryptocurrency prices. Panel A illustrates the relationship between wash trading volumes and weekly returns. Panel B further reports whether wash trading makes the price listed on unregulated exchanges deviate from "fair" prices on regulated exchanges. For each unregulated exchange, price deviation is measured as the log difference between its weekly close price and the average price from regulated exchanges (whose prices are very similar). In both panels, we regress these price indicators on logarithms of estimated wash trade volumes and control for features of exchanges both in contemporaneous and predictive regression specifications. The random-effect model is adopted in all regressions based on the Hausman test, with robust stand errors clustered at the exchange-currency level. We also include the currency fixed effect as robustness in both panels.

[Insert Table 7]

As shown in Panel A of Table 7, wash trade volume exhibits a positive and significant correlation with the weekly return in the same week. This result supports Prediction 4 as outlined in our theoretical model (Online Appendix G), which posits that wash trading is proportional to cryptocurrency prices. However, this relation reverses in the following week. The coefficients are statistically and economically significant, as shown in Models 3 and 7 of Table 7, Panel A). A one-standard-deviation

[21] Because crypto exchanges are not listed, we do not observe exchanges' revenues and profits. But we can estimate exchanges' profit for the ones that issue their own tokens with utility and dividend functions. Such exchanges periodically use a portion of their operating profit to buyback and destroy tokens from the secondary market (monthly or quarter). We manually collect all available buyback reports and token white papers from exchanges' website to compute the value of the tokens bought back or burned. Then with the buyback/profit ratio the exchanges promise (typically described) in the exchange tokens' white papers, we calculate the exchanges' profits. In our sample, Binance, Bitfinex, Huobi, KuCoin, Okex, Zb, Bgogo, BitZ, Mxc, Bibox, and Bitmart issue exchange tokens and have data available. We find an exchange's profit is positively correlated with both the reported volume and our estimated real volume. In an unreported pooled regression controlling for week fixed effect, the coefficient of log profit on log real volume is 0.85 and significant at a 1% level. We also find that reported CoinMarketCap volume positively and significantly predicts the subsequent week's non-wash-trading volume, consistent with the intuition and empirical findings in Amiram, Lyandres, and Rabetti (2022).

increase in wash trade volume(log) leads to a 0.63% increase in concurrent weekly return (annualized 32.76%), followed by a 0.42% decrease in the subsequent week (annualized 21.8%). The reverse relation with return suggests that higher wash trade volume drives up the contemporaneous price, but the wash-trade effect on price does not last long, and the price reverses in the following week. What we observe is intuitive: Faking transactions at higher prices can attract more investors who like to chase returns, but arbitrageurs close the pricing gap across exchanges over the next week.

To substantiate this intuition, we treat prices on regulated exchanges as “fair” price benchmarks and examine the price deviation of unregulated exchanges against this benchmark. Panel B shows a strong and positive relationship between wash trading volume and price deviations while controlling exchange characteristics. Considering the average wash trade percentage of 70% for unregulated exchanges, such an increase in wash trade volume leads to a 3.22% higher price on unregulated exchanges compared to their regulated counterparts, reflecting its significant economic implications. This finding also corresponds to Prediction 4. In addition, the price deviation converges to a marginal difference of 0.014% in the following period (as shown in Models 1 and 3 of Table 7, Panel B). This observation aligns with the idea that arbitrageurs take advantage of price differences across various exchanges in the following week, thereby reducing price deviations.

5.3 Determinants of Wash Trading

We first investigate which types of exchanges are more likely to engage in wash trading. We run a cross-sectional regression of the overall fraction of wash trades on an exchange against its characteristics, as shown in

Table 8. Robust standard errors are calculated to tackle heteroskedasticity. In the regressions, we include the age of the exchange and all three traffic indicators derived from a series of SimilarWeb reports. The number of unique visitors refers to the number of distinct individuals visiting a webpage, which is a close indicator of the user number or the “real” traders in the exchanges. A smaller number also implies that more visitors may have accessed the exchanges through third-party aggregators or referrals of the ranking websites. The other two indicators are based on each exchange’s top 5 traffic geographical origins. We rank all traffic countries based on GDP and Financial Access.[22] The number of countries ranked in the bottom 15 is counted if they appear in the top 5 traffic countries for crypto exchange.

[22] We extract 2016-2018 GDP and financial access data from the World Bank Databank. The measurement of finance access includes the number of commercial bank branches (per 100,000 adults), account ownership at a financial institution, and the number of ATM (per 100,000 adults). The average value of GDP and financial access measurement is used to rank all traffic countries in our sample.

[Insert

Table 8*]*

Table 8 demonstrates that the number of unique visitors is negatively correlated with wash trading, suggesting that exchanges with fewer unique visitors have a higher proportion of wash trades. From a *t*-test grouped by number of unique visitors, platforms with over 100,000 unique users on average engage in wash trading for 60.21% of the reported volume, which is significantly less than 82.69% for those with fewer than 100,000 users. These results align with the economic incentives behind wash trading. The prevailing notion among practitioners is that exchanges with a larger number of real users are subject to greater scrutiny, leading to stronger reputational concerns and a motivation to maintain transparency and accuracy (Rodgers, 2019). This observation is also following Prediction 3 outlined in our theoretical framework.

In addition, we observe a negative relationship between the age of exchange and the fraction of wash trades, statistically significant at a 1% level. The adjusted R^2 is 28.4% in Model 1, implying that the age of exchange is one leading factor correlated with the decision to wash trade. Newly established exchanges are more eager to wash trade since it is a shortcut to increase brand awareness and acquire clients. In fact, unregulated exchanges more than five years old on average wash trade 47.83% of the reported volume, a significant lower percentage compared to 81.32% for exchanges with less than five years.

The insignificant relationship with traffic country indicators implies that the extent of exchanges' wash trading may not vary across countries. We expect exchanges that rely more on referral traffic to have more incentives for wash trading, but this does not show up in our data due to either the short sampling period or the fact that many exchanges may not actively monitor web traffic sources.

Next, we investigate how market dynamics affect wash trading. Table 9 presents a panel regression of wash trade volumes on lagged "true" cryptocurrency weekly return and volatility obtained from the third-party composite price index on CoinMarketCap.[23] Standard errors are clustered at an exchange-currency level.

[Insert Table 9]

[23] Note that the weekly volatility is calculated using daily returns in the week. All regressions employ random effects with robust errors.

In Table 9, lagged cryptocurrency returns positively predict wash trade volume, while lagged volatility shows a strong negative prediction. In other words, misbehaving crypto exchanges tend to increase wash trading volumes when the market experiences recent positive returns or decreases in volatility in the past one or two weeks. Price increases could draw retail investors' attention and encourage speculation. Therefore, crypto exchanges are incentivized to pump up volumes to vie for better ranking and more clients. In addition, decreased volatility reduces the potential costs of wash trading (wash trading risks of capital loss in a volatile market). Therefore, lower volatility can lead to higher wash trading activities.

5.4 Suggestive Effects of Regulation and Implications for Policy and Industry Practice

Considering the substantial evidence and prevalent scale of wash trading in the crypto market, it is crucial to address its regulatory implications. Despite the decentralized ideal of crypto ecosystems, they remain heavily influenced by centralized exchanges that are not only vulnerable to cyberattacks but also prone to manipulative behavior. This casts doubt on the industry's progress and supports the skepticism raised by critics about the technology's limitations and the industry's fraudulent aspects (Roubini, 2018). Our findings add new insights concerning the role of regulation by demonstrating the vastly divergent trading patterns between regulated and unregulated exchanges. Without claiming causality, we offer three potential interpretations of the results.

First, regulated exchanges are required to follow the regulation, and violations are severely punished (Section 23 CRR-NY 200.3 and 200.6 of the New York Codes, Rules and Regulations, BitLicense, 2015). The centralized nature of these exchanges does make direct inspections and the enforcement of regulation on crypto exchanges more feasible than on other (often anonymous) agents. For example, faking trading records are nearly impossible because regulated exchanges are required to regularly submit data 'for each transaction, the amount, date, and precise time of the transaction, any payment instructions, the total amount of fees and charges received and paid to, by, or on behalf of the licensee' (23 CRR-NY 200.12, New York Codes, Rules and Regulations).[24]

Second, it is possible that compliance with regulation is costly but does not affect wash trading incentives directly. Some firms simply get a license to signal their quality (e.g., Spence, 1978). This is inconsistent with the observation that after acquiring the license, regulated exchanges still do not

[24] Regulators, with basic forensic tools, can easily find wash trading evidence with this level of trading records. For example, the exchange account carrying wash trading activities are likely to exhibit abnormally large volume as well as unusual behavior patterns (like Willy and Markus in Mt. Gox). Besides, wash trading volume do not bring trading fee revenues, so exchanges need to work hard to cover the tells in the balance sheet. Regulators can also find suspicious wash trading from the trading volume and user scale and custodian assets size.

wash trade. Third, some unobserved exchange characteristics may cause the exchange to refrain from wash trading and acquire licenses simultaneously. Such a screening function is plausible and would imply that by observing which exchanges are regulated, traders can tell whether wash trading takes place on a particular exchange.

Contrary to popular belief, the five regulated spot exchanges under BitLicense only constitute less than 3% of the total transaction volume in the cryptocurrency market based on CoinMarketCap data (October 2022). This implies that wash trading on unregulated exchanges is a first-order issue that demands more regulatory attention. To address this, we provide an initial set of tools to effectively uncover wash trading and combat non-compliant and unethical behaviors. It is essential for regulatory tools and policies to be adaptive, as our statistical tests may become outdated when sophisticated wash traders incorporate them into their strategies. Nevertheless, the benefits of transparency, proper regulation, and close public monitoring that we touch upon are enduring.

6 Conclusion

The nascency of the cryptocurrency industry provides a unique setting in which we observe both regulated and unregulated exchanges that are influential. We demonstrate that most major unregulated crypto exchanges feature excessive wash trading and warn that centralized (and vertically integrated) exchanges absent proper regulation can be problematic, as seen in the collapse of FTX Trading. Specifically, we find that first-digit distributions of trade size follow Benford's law for regulated exchanges, whereas nearly 30% of unregulated exchanges show violations. Regulated exchanges show apparent trade clustering at round sizes and a high level of transaction roundness; while for unregulated exchanges, the levels of roundness are generally low, and the trade-size clustering phenomenon is less prominent. Furthermore, regulated exchanges display the power-law distributions with exponents in the Pareto–Lévy range, consistent with other financial markets; in contrast, 20% of Tier-1 and 75 % of Tier-2 unregulated exchanges fail to follow in any cryptocurrency.

We estimate the average wash trading to be 53.4% of trading on unregulated Tier-1 exchanges and 81.8% on Tier-2 exchanges and provide several robustness and validation tests. We further provide suggestive evidence that wash trading inflates exchange rankings and cryptocurrency prices and is being predicted by market signals such as past cryptocurrency prices and volatility and exchange characteristics such as exchange age and user base. As the first comprehensive study of the pervasive crypto wash trading, our paper not only provides a cautionary tale to policymakers around the globe concerning centralized crypto exchanges but also reminds the readers of the disciplining or screening effects of regulation in emerging industries, the importance of using wash-trading-adjusted volume in certain empirical studies, and the utility of statistical tools and behavioral benchmarks for forensic finance and fraud detection. Going forward, our approaches can be further adapted to constructing wash trading metrics using publicly available data but at a lower frequency, or to detect wash trading in the new Non-Fungible Token (NFT) markets.

Our study provides compelling evidence that centralized exchanges, due to their opacity, vertical integration, and lack of regulation, create ample opportunities for market manipulation, particularly by exchanges themselves. In response, consumers might be inclined to seek alternative trading venues, such as DEXs. However, DEXs come with their own set of unique liquidity, legal, and security risks (e.g., Capponi, Kaplan, and Sarkar, 2022). Additionally, many decentralized finance (DeFi) platforms possess suboptimal designs in fee mechanisms (Cong, Tang, Wang, and Zhao, 2022) or interest functions (Rivera, Saleh, and Vandeweyer, 2023) that need to be improved. Future research should empirically compare market manipulation and other limitations in both centralized finance institutions (CeFi) like centralized exchanges and DeFi platforms. Moreover, whether centralized or decentralized, financial services should not be outside the law. Appropriate and clear regulations, as our study suggests, prove crucial for the long-term development of the industry.

References

Aggarwal, R. K., and Wu, G. (2006). Stock market manipulations. *The Journal of Business, 79*(4), 1915-1953.

Alameda. (2019). *Investigation into the Legitimacy of Reported Cryptocurrency Exchange Volume*. Retrieved from *https://ftx.com/volume-report-paper.pdf*

Alexander, G. J., and Peterson, M. A. (2007). An analysis of trade-size clustering and its relation to stealth trading. *Journal of Financial Economics, 84*(2), 435-471.

Allen, F., and Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies, 5*(3), 503-529.

Aloosh, A., and Li, J. (2023). Direct evidence of Bitcoin wash trading. *Management Science,* Forthcoming

Alstott, J., Bullmore, E., and Plenz, D. (2014). powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS one, 9*(1), e85777.

Amiram, D., Lyandres, E., and Rabetti, D. (2022). Cooking the Order Books: Information Manipulation and Competition among Crypto Exchanges . *Available at SSRN 3745617*.

ap Gwilym, O., and Meng, L. (2010). Size clustering in the FTSE100 index futures market. *Journal of Futures Markets: Futures, Options, and Other Derivative Products, 30*(5), 432-443.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, 551-572.

Bergh, D. (2015). Sample size and chi-squared test of fit—a comparison between a random sample approach and a chi-square value adjustment method using Swedish adolescent data. In Pacific rim objective measurement symposium (PROMS) 2014 conference proceedings (pp. 197-211). Springer, Berlin, Heidelberg.

Bitfinex'ed. (2017). Wash Trading Bitcoin Part II: Who and why is someone wash trading on Bitfinex? *Medium.com* Oct 22. Retrieved from https://medium.com/@bitfinexed/wash-trading-bitcoin-part-ii-who-and-why-is-someone-wash-trading-on-bitfinex-e1c7b5e0b3bb

BitLicense. (2015). Regulations of the Superintendent of Financial Services, Part 200: Virtual Currencies*. New York State Department of Financial Services.* Retrieved from https://govt.westlaw.com/nycrr/Browse/Home/NewYork/NewYorkCodesRulesandRegulations?guid=I7444ce80169611e594630000845b8d3e&originationContext=documenttoc&transitionType=Default&contextData=(sc.Default)

Blandin, A., Cloots, A. S., Hussain, H., Rauchs, M., Saleuddin, R., Allen, J. G., and Cloud, K. (2019). Global cryptoasset regulatory landscape study. *University of Cambridge Faculty of Law Research Paper*(23).

Bogart, S. (2019). Blockchain Capital Bitcoin Survey. Retrieved from *https://medium.com/blockchain-capital-blog/bitcoin-is-a-demographic-mega-trend-data-analysis-160d2f7731e5*?

BTI. (2019). Blockchain Transparency Institute April Summary of Market Surveillance Report. Retrieved from

Capponi, A., Kaplan, N., and Sarkar, A. (2022). Can Decentralized Finance Provide More Protection for Crypto Investors? *Liberty Street Economics*. Dec 21.

Capponi, A., Olafsson, S., and Alsabah, H. (2022). Proof-of-Work Cryptocurrencies: Does Mining Technology Undermine Decentralization? *Management Science,* Forthcoming.

Carslaw, C. A. (1988). Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review,* 63(2), 321–327

Chen, J., Lin, D., and Wu, J. (2022). Do cryptocurrency exchanges fake trading volumes? An empirical analysis of wash trading based on data mining. *Physica A: Statistical Mechanics and its Applications*, 586, 126405.

Chen, T. (2018). Round-number biases and informed trading in global markets. *Journal of Business Research,* 92, 105-117.

Choi, S.H., and Jarrow, B. (2020). Testing the Local Martingale Theory of Bubbles using Cryptocurrencies. *Available at SSRN 3701960*.

Christie, W. G., and Schultz, P. H. (1994). Why do NASDAQ market makers avoid odd-eighth quotes? *The Journal of Finance, 49*(5), 1813-1840.

Clarkson, P., Nekrasov, A., Simon, A., and Tutticci, I. (2020). Target price forecasts: Fundamental and non-fundamental factors. *Journal of Business Finance & Accounting,* Forthcoming

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review, 51*(4), 661-703.

CoinMarketCap. (2020). Exchange Ranking. *Available at: https://support.coinmarketcap.com/hc/en-us/articles/360052030111-Exchange-Ranking*

Cong, L. W., Harvey, C., Rabetti, D., and Wu, Z.Y. (2022). An Anatomy of Crypto-Enabled Cybercrimes. NBER Working Paper.

Cong, L. W., and He, Z. (2019). Blockchain Disruption and Smart Contracts. *Review of Financial Studies, 32(*5), 1754-1797.

Cong, L. W., Grauer, K., Rabetti, D., and Updegrave, H. (2023). Blockchain Forensics and Crypto-Related Cybercrimes. *Available at SSRN 4358561*.

Cong, L. W., Li, Y., and Wang, N. (2020). Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies, 34*(3), 1105-1155.

Cong, L. W., Li, Y., and Wang, N. (2022). Token-based platform finance. *Journal of Financial Economics*, *144*(3), 972-991.

Cong, L. W., Landsman, W., Maydew, E., & Rabetti, D. (2023). Tax-loss harvesting with cryptocurrencies. *Journal of Accounting and Economics*, 101607.

Cong, L. W., Tang, K., Wang, Y., and Zhao, X. (2022). Inclusion and Democratization Through Web3 and DeFi? Initial Evidence from the Ethereum Ecosystem. *Available at SSRN 4229308.*

Cong, L. W., and Xiao, Y. (2021). Categories and functions of crypto-tokens. The Palgrave Handbook of FinTech and Blockchain, 267-284.

Cumming, D., Johan, S., and Li, D. (2011). Exchange trading rules and stock market liquidity. *Journal of Financial Economics, 99*(3), 651-671.

Da Silva, S., Matsushita, R., Gleria, I., and Figueiredo, A. (2007). Hurst exponents, power laws, and efficiency in the Brazilian foreign exchange market. *Economics Bulletin, 7*(1), 1-11.

Durtschi, C., Hillison, W., and Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting, 5*(1), 17-34.

Fadilpasic, S. (2019). OKEx Defends Itself from Wash Trading Accusations With a BTC 100 Bet. *CryptoNews*. Sept 24.

FCA. (2019). Cryptoassets: Ownership and attitudes in the UK. (Financial Conduct Authority). Retrieved from *https://www.fca.org.uk/publication/research/cryptoassets-ownership-attitudes-uk-consumer-survey-research-report.pdf.*

Fidelity. (2019). Institutional Investments in Digital Assets. Retrieved from *https://s2.q4cdn.com/997146844/files/doc_news/archive/59439969-390c-4354-94a9-772219d0b8b9.pdf*

Foley, S., Karlsen, J. R., and Putniņš, T. J. (2019). Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies, 32*(5), 1798-1853.

Fusaro, T., and Hougan, M. (2019). Presentation to the US Securities and Exchange Commission. Retrieved from https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003a). A theory of power-law distributions in financial market fluctuations. *Nature, 423*(6937), 267-270.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003b). Understanding the cubic and half-cubic laws of financial fluctuations. *Physica A: Statistical Mechanics and its Applications, 324*(1-2), 1-5.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2006). Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, *121*(2), 461-504.

Gandal, N., Hamrick, J., Moore, T., and Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem. *Journal of Monetary Economics, 95*, 86-96.

Goldstein, I., Gupta, D., & Sverchkov, R. (2019). Utility Tokens as a Commitment to Competition. *Available at SSRN 3484627*.

Gomber, P., Gsell, M., Pujol, G., and Wranik, A. (2009). Regulation and Technology in Equity Trading–The Impact of Regulatory and Technological Changes on Order Execution and the Trading Venue Landscape in Europe. *Facing New Regulatory Frameworks in Securities Trading in Europe*, 31-54.

Gopikrishnan, P., Plerou, V., Amaral, L. A. N., Meyer, M., and Stanley, H. E. (1999). Scaling of the distribution of fluctuations of financial market indices. *Physical Review E, 60*(5), 5305.

Gopikrishnan, P., Plerou, V., Gabaix, X., and Stanley, H. E. (2000). Statistical properties of share volume traded in financial markets. *Physical Review E, 62*(4), R4493.

Grinblatt, M. and Keloharju, M. (2004). Tax-loss trading and wash sales. *Journal of Financial Economics*, *71*(1), 51-76

Griffin, J. M., and Kruger, S. (2023). What is Forensic Finance?. Available at SSRN 4490028.

Griffin, J. M., and Shams, A. (2020). Is Bitcoin really untethered? *The Journal of Finance, 75*(4), 1913-1964.

Grossman, S. J., Miller, M. H., Cone, K. R., Fischel, D. R., and Ross, D. J. (1997). Clustering and competition in asset markets. *The Journal of Law and Economics*, *40*(1), 23-60.

Halaburda, H., Haeringer, G., Gans, J., and Gandal, N. (2022). The microeconomics of cryptocurrencies. *Journal of Economic Literature*, 60(3), 971-1013.

Henry, C. S., Huynh, K. P., and Nicholls, G. (2019). Bitcoin Awareness and Usage in Canada: An Update. *The Journal of Investing, 28*(3), 21-31.

Helms, K. (2020). $8.8 Trillion Traded in Cryptocurrency Spot and Futures Markets in Q1: Reports. Retrieved from *https://news./trillion-traded-cryptocurrency-spot-futures-markets/*

Hernando-Veciana, Á., and Tröge, M. (2020). Cheap Talk and Strategic Rounding in LIBOR Submissions. *The Review of Financial Studies, 33*(6), 2585-2621.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 1163-1174.

Howell, S., Niessner, M., and Yermack, D. (2020). Initial Coin Offerings: Financing Growth with Cryptocurrency Tokens. *Review of Financial Studie*s 33 (9), 3925–3974

Huillet, M. (2019). OKEx Slams New Wash Trading Allegations as 'Inaccurate and Misleading'. *CoinTelegraph*, Sept. 23. Retrieved from https://cointelegraph.com/news/okex-slams-new-wash-trading-allegations-as-inaccurate-and-misleading

Ikenberry, D. L., and Weston, J. P. (2008). Clustering in US stock prices after decimalisation. *European Financial Management, 14*(1), 30-54.

Imisiker, S., and Tas, B. K. O. (2018). Wash trades as a stock market manipulation tool. *Journal of behavioral and experimental finance, 20*, 92-98.

IOSCO. (2000). A Resolution on IASC Standards. In: Presidents' Committee of IOSCO Madrid, Spain.

Irresberg, F., John, K., and Saleh, F. (2020). The public blockchain ecosystem: an empirical analysis. *Working paper.*

Jarrow, R. A. (1992). Market manipulation, bubbles, corners, and short squeezes. *Journal of financial and Quantitative Analysis, 27*(3), 311-336.

Kandel, S., Sarig, O., and Wohl, A. (2001). Do investors prefer round stock prices? Evidence from Israeli IPO auctions. *Journal of Banking and Finance, 25*(8), 1543-1551.

Kostanjčar, Z., and Jeren, B. (2013). Emergence of power-law and two-phase behavior in financial market fluctuations. *Advances in Complex Systems, 16*(01), 1350008.

Kuo, W.Y., Lin, T. C., and Zhao, J. (2015). Cognitive limitation and investment performance: Evidence from limit order clustering. *The Review of Financial Studies, 28*(3), 838-875.

Lacetera, N., Pope, D. G., and Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review, 102*(5), 2206-2236.

Le Pennec, G., Fiedler, I., and Ante, L. (2021). Wash trading at cryptocurrency exchanges. *Finance Research Letters*, Feb 19:101982

Li, M., Cai, Q., Gu, G., and Zhou, W. (2019). Exponentially decayed double power-law distribution of Bitcoin trade sizes. *Physica A: Statistical Mechanics and its Applications, 535*, 122380.

Li, T., Shin, D., and Wang, B. (2020). Cryptocurrency pump-and-dump schemes. *Available at SSRN 3267041*.

Li, Z., Cong, L., and Wang, H. (2004). Discussion on Benford's Law and its Application. *arXiv preprint math/0408057*.

Liu, Y., and Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *Review of Financial Studies, 34*(6), 2689-2727.

Lyandres, E., Palazzo, B., & Rabetti, D. (2022). Initial coin offering (ICO) success and post-ICO performance. *Management Science*, *68*(12), 8658-8679.

Mahmoodzadeh, S., and Gençay, R. (2017). Human vs. high-frequency traders, penny jumping, and tick size. *Journal of Banking & Finance, 85*, 69-82.

Mandelbrot, B. (1960). The Pareto–Lévy law and the distribution of income. *International Economic Review, 1*(2), 79-106.

Mao, R., Li, Z., and Fu, J. (2015). Fraud transaction recognition: A money flow network approach. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1871-1874).

Maslov, S., and Mills, M. (2001). Price fluctuations from the order book perspective—empirical facts and a simple model. *Physica A: Statistical Mechanics and its Applications*, *299*(1-2), 234-246.

Moulton, P. C. (2005). You can't always get what you want: Trade-size clustering and quantity choice in liquidity. *Journal of Financial Economics, 78*(1), 89-119.

Nirei, M., Stachurski, J., and Watanabe, T. (2020). *Trade Clustering and Power Laws in Financial Markets*. *Theoretical Economics*, *15*(4), 1365-1398.

O'Hara, M., Yao, C., and Ye, M. (2014). What's not there: Odd lots and market data. *The Journal of Finance, 69*(5), 2199-2236.

Ohnishi, T., Takayasu, H., Ito, T., Hashimoto, Y., Watanabe, T., and Takayasu, M. (2008). Dynamics of quote and deal prices in the foreign exchange market. *Journal of Economic Interaction and Coordination, 3*(1), 99.

Osler, C., and Savaser, T. (2011). Extreme returns: The case of currencies. *Journal of banking & finance*, *35*(11), 2868-2880.

Pareto, V. (1896). Cours d'Économie Politique Professé a l'Université de Lausanne (in French). Lausanne, Rouge.

Perez, Y. B. (2015). The Real Cost of Applying for a New York BitLicense. Coindesk, Aug 13, Retrieved from *https://www.coindesk.com/real-cost-applying-new-york-bitlicense*

Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Gabaix, X., and Stanley, H. E. (2000). Economic fluctuations and anomalous diffusion. *Physical Review E, 62*(3), R3023.

Plerou, V., and Stanley, H. E. (2007). Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets. *Physical Review E, 76*(4), 046109.

Prandl, S., M. Lazarescu, D. S. Pham, S. T. Soh and S. Kak. (2017). An investigation of power law probability distributions for network anomaly detection, IEEE security and privacy workshops (SPW), pp. 217-222.

PwC. (2022). 4th Annual Global Crypto Hedge Fund Report 2022. Retrieved from https://www.pwc.com/gx/en/financial-services/pdf/4th-annual-global-crypto-hedge-fund-report-june-2022.pdf

Q.ai. (2022). What Happened To Crypto Giant FTX? A Detailed Summary Of What We Actually Know So Far. *Forbes,* Dec 12*.*

Quiroz-Gutierrez, M. (2022). A handful of NTF users are making big money off of a stealth scam. Here's how `wash trading' works. *Fortune* Feb 4

Ritter, J. R. (2008). Forensic finance. *Journal of Economic Perspectives, 22*(3), 127-147.

Rivera, T. J., Saleh, F., and Vandeweyer, Q. (2023) Equilibrium in a DeFi Lending Market. *Available at SSRN 4389890*.

Robinson, T. (2021). Crypto Enforcement Actions by US Regulators Reach $2.5 Billion. Elliptic Blog. Retrieved from https://www.elliptic.co/blog/https/www.elliptic.co/blog/crypto-enforcement-actions-by-us-regulators-reach-2.5-billion

Rodgers, T. (2019). 95% Of Volume Could Be Wash trading As Bitcoin Price Surges. *Forbes*, Apr 4.

Roger, T., Roger, P., and Schatt, A. (2018). Behavioral bias in number processing: Evidence from analysts' expectations. *Journal of Economic Behavior and Organization, 149*, 315-331.

Rosch, E. (1975). Cognitive reference points. *Cognitive psychology, 7*(4), 532-547.

Roşu, I, and Saleh, F (2021). Evolution of Shares in a Proof-of-Stake Cryptocurrency. *Management Science, 67*(2), 661-672.

Roubini, N. (2018). Testimony for the Hearing of the US Senate Committee on Banking, Housing and Community Affairs On "Exploring the Cryptocurrency and Blockchain Ecosystem. Retrieved from *https://www.banking.senate.gov/imo/media/doc/Roubini%20Testimony%2010-11-18.pdf*

Schnaubelt, M., Rende, J., and Krauss, C. (2019). Testing stylized facts of bitcoin limit order books. *Journal of Risk and Financial Management, 12*(1), 25.

Shams, A. (2020). The Structure of Cryptocurrency Returns. *Charles A. Dice Center Working Paper*(2020-11)

Sinclair, S. (2020). Canada Crypto Exchange Coinsquare Accused of Wash Trading by Watchdog. *Coindesk*, July 20, Policy and Regulation

Spence, M. (1978). Job Market Signaling. *Uncertainty in economics*, pp. 281-306. Academic Press.

Tabb, L., Iati, R., and Sussman, A. (2009). US equity high frequency trading: Strategies, sizing and market structure. *TABB Group report*(1), 2.

Thomas, J. K. (1989). Unusual patterns in reported earnings. *Accounting Review*, 773-787.

Vandewalle, N., Ausloos, M., and Boveroux, P. (1997). *Detrended fluctuation analysis of the foreign exchange market.* Paper presented at the Econophysic Workshop, Budapest, Hungary.

Verousis, T., and ap Gwilym, O. (2013). Trade size clustering and the cost of trading at the London Stock Exchange. *International Review of Financial Analysis, 27*, 91-102.

Victor, F., and Weintraud, A. M. (2021). Detecting and Quantifying Wash Trading on Decentralized Cryptocurrency Exchanges. Papers 2102.07001, arXiv.org.

Vigna, P. (2019). Most Bitcoin Trading Faked by Unregulated Exchanges, Study Finds. *The Wall Street Journal*, March 22

Vigna, P., and Osipovich, A. (2018). Bots Are Manipulating Price of Bitcoin in `Wild West of Crypto. *The Wall Street Journal*, Oct 2

Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature, 50*(3), 731-769.

Table 1. Information on Crypto Exchanges

Table 1 demonstrates several characteristics of crypto exchanges in the data set. The exchange age is the duration from an exchange's establishment date to July 2019. Exchanges are categorized into three groups based on their length of survival: "more than 5 years," "between 2 and 5 years," and "less than 2 years". Trade volume is calculated as the sum of all transactions involving the four selected cryptocurrency pairs, i.e., BTC, ETH, LTC, and XRP, all against U.S. dollars. SimilarWeb rankings are based on the SimilarWeb report over the period from Aug. 2019 to Oct. 2019, available at https://www.similarweb.com/. Alexa's ranking is accessed through https://www.alexa.com/siteinfo in Nov. 2019. CoinMarketCap ranking is based on daily trade volume, calculated as daily average using proprietary data acquired from https://www.coinmarketcap.com during the sample period.

Exchange Code	Exchange Age	Trade Volume ($mil)	Ranking by Web Traffic: SimilarWeb Average Rank in the Investment Section	Ranking by Web Traffic: SimilarWeb Average Number of Monthly Visits (millions)	Ranking by Web Traffic: Alexa Average Rank among all Websites	Ranking by Trade Volume: CoinMarketCap
Panel A Regulated exchanges						
Bitstamp	A ≥ 5	1466	473	1.872	14297	63.7
Coinbase	A ≥ 5	15212	17	20.678	2254	50.3
Gemini	A ≥ 5	1568	1418.5	0.487	23950	99.2
Panel B Unregulated Tier-1 exchanges						
Binance	2 ≤ A< 5	41936	21	18.770	1630	10.5
Bittrex	A ≥ 5	434	276	2.983	5960	89.9
Bitfinex	A ≥ 5	11175	345	2.57	9683	59.5
HitBTC	A ≥ 5	34157	498.5	1.363	9815	27.9
Huobi	A ≥ 5	38789	285.5	1.673	8379	22.7
KuCoin	A < 2	4005	255.5	1.879	8663	55.2
Liquid	A ≥ 5	545	699	0.394	13357	53.3
Okex	A ≥ 5	24646	633	1.224	3636	14.5
Poloniex	A ≥ 5	975	38	2.146	768	95.6
Zb	A ≥ 5	18452	517.5	1.449	5231	30.0
Panel C Unregulated Tier-2 exchanges						
Bgogo	A < 2	7805	17322	0.032	81142	29.9
Biki	A < 2	30997	N/A	0.260	3684	19.0
BitZ	2 ≤ A< 5	3464	4926.5	0.096	19860	16.1
Coinbene	A < 2	50944	2594	0.234	30210	10.2
DragonEX	A < 2	14534	5928.5	0.031	363745	46.6
Lbank	2 ≤ A < 5	52741	6735	0.092	6422	16.0
Mxc	A < 2	34624	2770	0.265	6306	11.9
Fcoin	A < 2	21848	1818.5	0.092	100223	15.0
Exmo	2 ≤ A < 5	52	961.5	0.919	37634	90.0
Coinmex	A < 2	2756	11567	0.007	1684659	6.6
Bibox	A < 2	32305	3403.5	0.190	1714	16.8
Bitmart	A < 2	16035	3243	0.313	22780	30.8
Bitmax	A < 2	2612	2316.5	0.342	28739	30.4
Coinegg	2 ≤ A <5	16668	10350.5	0.032	53000	21.3
Digifinex	A < 2	23525	3061.5	0.188	1858	16.0
Gateio	A ≥ 5	2013	1096.5	1.065	2808	73.7

Table 2. Chi-squared Test for Conformity with Benford's Law

The results in this table show whether trade-size distributions of exchanges are consistent with the distribution of Benford's law. χ^2 statistics and *p*-value are reported in the table. ***, **, and * denote the statistical significance levels at 1%, 5% and 10%, respectively.

Exchange	BTC/USD		ETH/USD		LTC/USD		XRP/USD	
	χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
Panel A Regulated exchanges								
Bitstamp	1.647	0.990	1.639	0.990	4.905	0.768	11.487	0.176
Coinbase	2.736	0.950	2.767	0.948	3.218	0.920	2.189	0.975
Gemini	3.304	0.914	0.698	1.000	1.969	0.982	NA	NA
Panel B Unregulated Tier-1 exchanges								
Binance	2.495	0.962	4.113	0.847	4.645	0.795	7.205	0.515
Bittrex	1.464	0.993	2.620	0.956	6.117	0.634	0.748	0.999
Bitfinex	29.501***	0.000	5.349	0.720	7.157	0.520	47.121***	0.000
HitBTC	6.329	0.610	3.833	0.872	7.641	0.469	1.482	0.993
Huobi	6.832	0.555	3.104	0.928	1.094	0.998	0.468	1.000
KuCoin	5.969	0.651	4.100	0.848	7.386	0.496	7.790	0.454
Liquid	17.223**	0.028	4.823	0.776	NA	NA	3.644	0.888
Okex	2.601	0.957	1.956	0.982	3.724	0.881	4.230	0.836
Poloniex	3.228	0.919	7.886	0.445	2.454	0.964	14.219*	0.076
Zb	2.815	0.945	0.069	1.000	0.813	0.999	0.541	1.000
Panel C Unregulated Tier-2 exchanges								
Bgogo	0.548	1.000	0.949	0.999	NA	NA	NA	NA
Biki	24.261***	0.002	16.677**	0.034	6.505	0.591	4.371	0.822
BitZ	4.660	0.793	19.569**	0.012	3.396	0.907	4.490	0.810
Coinbene	1.360	0.995	2.468	0.963	0.673	1.000	0.723	0.999
DragonEX	50.614***	0.000	8.254	0.409	124.881***	0.000	39.69***	0.000
Lbank	0.399	1.000	0.064	1.000	NA	NA	NA	NA
Mxc	5.088	0.748	23.086***	0.003	60.516***	0.000	15.300*	0.054
Fcoin	114.788***	0.000	141.768***	0.000	31.068***	0.000	57.021***	0.000
Exmo	63.022***	0.000	122.298***	0.000	NA	NA	71.949***	0.000
Coinmex	10.771	0.215	4.662	0.793	12.325	0.137	26.135***	0.001
Bibox	2.430	0.965	7.140	0.522	4.115	0.847	7.602	0.473
Bitmart	0.544	1.000	0.122	1.000	1.042	0.998	14.676*	0.066
Bitmax	1.157	0.997	2.583	0.958	11.614	0.169	4.815	0.777
Coinegg	0.678	1.000	23.351***	0.003	109.944***	0.000	26.835***	0.001
Digifinex	2.240	0.973	0.536	1.000	0.703	1.000	2.249	0.972
Gateio	1.695	0.989	0.924	0.999	1.317	0.995	0.577	1.000

Table 3. Student's *t*-tests for Trade-size Clustering

In the table, two sets of *t*-test results with different testing points and observation windows are demonstrated: 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). A positive difference indicates that frequency at round size is higher than the rest within the observation window, i.e., trade-size clustering. ***, **, and * denote statistical significance levels at 1%, 5%, and 10%, respectively.

Observation range: Multiples of 100 units (100X-50, 100+50)

Exchange	BTC/USD		ETH/USD		LTC/USD		XRP/USD	
	Difference	*t* statistics	Difference	*t* statistics	Difference	*t* statistics	Difference	*t* statistics
Panel A Regulated exchanges								
Bitstamp	0.091***	14.490	0.112***	12.280	0.160***	10.767	0.063***	6.726
Coinbase	0.089***	14.875	0.135***	15.647	0.109***	8.945	0.032***	2.955
Gemini	0.125***	13.655	0.119	9.713	0.203***	8.284	NA	NA
Panel B Unregulated Tier-1 exchanges								
Binance	0.188***	16.993	0.226***	20.740	0.179***	9.310	0.005	0.540
Bittrex	0.026*	1.926	0.039**	2.327	0.065***	2.943	0.076***	3.952
Bitfinex	0.100***	12.654	0.078***	8.655	0.110***	6.696	0.076***	5.681
HitBTC	0.005	1.073	-0.002	-0.568	0.004	0.644	-0.005	-0.556
Huobi	0.128***	16.895	0.083***	14.442	0.104***	8.003	0.010	1.116
KuCoin	-0.015	-2.668	-0.001	-0.081	-0.003	-0.089	-0.014	-1.379
Liquid	0.088***	6.854	0.057***	3.685	NA	NA	0.132***	6.498
Okex	0.082***	12.620	0.067***	10.614	0.047***	5.289	0.009	0.903
Poloniex	0.084***	10.192	0.060***	5.782	0.101***	4.018	0.054**	2.570
Zb	-0.013	-4.119	-0.016	-18.635	-0.030	-9.173	-0.020	-16.206
Panel C Unregulated Tier-2 exchanges								
Bgogo	-0.016	-86.208	-0.022	-7.374	NA	NA	NA	NA
Biki	-0.015	-24.733	-0.014	-12.297	-0.017	-27.701	-0.017	-34.675
BitZ	0.030***	7.110	0.029***	3.687	-0.002	-0.131	-0.083	-2.264
Coinbene	-0.008	-5.629	-0.015	-5.415	-0.012	-2.601	-0.008	-1.019
DragonEX	0.073***	6.573	-0.027	-7.279	-0.015	-13.844	-0.014	-11.199
Lbank	-0.020	-33.174	-0.022	-52.875	NA	NA	NA	NA
Mxc	0.019*	1.952	0.096***	9.019	0.058***	9.982	-0.017	-15.221
Fcoin	-0.001	-0.341	0.035***	6.552	-0.005	-0.804	-0.008	-1.207
Exmo	0.106**	2.313	0.032	1.038	NA	NA	-0.022	-0.450
Coinmex	-0.004	-5.622	-0.015	-11.549	-0.016	-12.730	-0.015	-22.775
Bibox	0.259***	20.279	0.123***	31.466	0.111***	15.258	-0.017	-16.156
Bitmart	-0.015	-13.164	-0.014	-15.846	-0.021	-15.304	-0.035	-3.158
Bitmax	0.034***	3.411	0.061***	8.316	0.094***	5.662	0.083***	6.503
Coinegg	-0.032	-22.436	-0.021	-33.123	-0.036	-16.175	-0.033	-2.149
Digifinex	-0.015	-8.266	-0.015	-8.765	-0.018	-35.684	-0.017	-30.582
Gateio	0.243***	20.575	0.019**	2.354	0.018*	1.753	0.004	0.333

Observation range: Multiples of 500 units (500X-100, 500X +100)

Exchange	BTC/USD		ETH/USD		LTC/USD		XRP/USD	
	Difference	*t* statistics	Difference	*t* statistics	Difference	*t* statistics	Difference	*t* statistics
Panel A Regulated exchanges								
Bitstamp	0.203***	15.193	0.271***	15.533	0.248***	7.904	0.166***	7.849
Coinbase	0.195***	16.758	0.290***	18.503	0.206***	9.965	0.137***	5.893
Gemini	0.266***	13.145	0.310***	13.376	0.331***	7.750	NA	NA
Panel B Unregulated Tier-1 exchanges								
Binance	0.354***	25.223	0.391***	35.160	0.393***	16.171	0.083***	3.529
Bittrex	0.096***	3.000	0.102***	2.898	0.114	1.691	0.137***	3.544
Bitfinex	0.221***	13.626	0.193***	12.202	0.236***	7.838	0.197***	6.004
HitBTC	0.039***	2.978	0.033***	3.572	0.039**	2.086	0.035	1.602
Huobi	0.257***	24.010	0.147***	19.769	0.198***	10.850	0.059***	3.018
KuCoin	-0.018	-2.342	0.024	0.889	0.069	0.960	-0.030	-1.427
Liquid	0.185***	5.603	0.171***	4.938	NA	NA	0.247***	5.746
Okex	0.139***	16.418	0.105***	13.011	0.077***	5.647	0.035**	2.012
Poloniex	0.163***	6.312	0.159***	7.099	0.239***	4.518	0.096***	2.768
Zb	-0.010	-2.025	-0.009	-6.041	-0.029	-3.679	-0.013	-7.457
Panel C Unregulated Tier-2 exchanges								
Bgogo	-0.008	-45.062	-0.014	-2.571	NA	NA	NA	NA
Biki	-0.007	-18.615	-0.002	-0.596	-0.009	-10.838	-0.009	-12.036
BitZ	0.007	1.122	0.041**	2.366	-0.055	-1.133	-0.070	-0.843
Coinbene	-0.005	-3.509	-0.001	-0.142	0.006	0.451	-0.001	-0.096
DragonEX	-0.009	-3.261	-0.014	-4.028	-0.006	-3.890	-0.006	-8.531
Lbank	-0.014	-11.815	-0.012	-17.525	NA	NA	NA	NA
Mxc	0.079**	2.078	0.246***	15.485	0.018*	2.008	-0.009	-7.708
Fcoin	0.006	1.333	0.030***	3.498	0.000	-0.022	0.003	0.415
Exmo	0.182**	2.880	0.070	1.154	NA	NA	0.059	0.602
Coinmex	-0.002	-6.491	-0.007	-16.342	NA	NA	NA	NA
Bibox	0.369***	11.156	0.061***	9.883	0.062***	5.522	-0.008	-13.686
Bitmart	-0.001	-0.743	-0.008	-12.134	-0.012	-8.184	NA	NA
Bitmax	0.150***	5.935	0.098***	6.720	0.054***	2.845	0.155***	6.923
Coinegg	-0.020	-11.980	-0.012	-13.575	-0.022	-9.611	0.001	0.120
Digifinex	-0.004	-0.622	-0.001	-0.185	-0.009	-10.539	-0.008	-15.631
Gateio	0.219***	8.589	0.080***	4.489	0.051**	2.499	0.036	1.442

Table 4. Power-law Fitting

This table shows the fitting results of the tails of trade size distribution. Ordinary Least Square (OLS) and Maximum Likelihood Estimation (MLE) are applied for the estimation of scaling parameters $\widehat{\alpha}_{OLS}$ and $\widehat{\alpha}_{Hill}$, respectively. We apply the probability density function to estimate the scaling exponents 1+ α. We also check whether the estimated parameters are within the Pareto–Lévy range (1<α<2) and mark “Y” if both exponents lie within the Pareto–Lévy range.

Exchange	BTC/USD			ETH/USD			LTC/USD			XRP/USD		
	$\widehat{\alpha}_{OLS}$	$\widehat{\alpha}_{Hill}$	Pareto–Lévy (1<α<2)	$\widehat{\alpha}_{OLS}$	$\widehat{\alpha}_{Hill}$	Pareto–Lévy (1<α<2)	$\widehat{\alpha}_{OLS}$	$\widehat{\alpha}_{Hill}$	Pareto–Lévy (1<α<2)	$\widehat{\alpha}_{OLS}$	$\widehat{\alpha}_{Hill}$	Pareto–Lévy (1<α<2)
Panel A Regulated exchanges												
Bitstamp	1.806	1.279	Y	1.696	1.374	Y	1.510	1.849	Y	1.748	1.338	Y
Coinbase	1.763	1.191	Y	1.745	1.308	Y	1.857	1.309	Y	1.809	1.257	Y
Gemini	1.668	1.297	Y	1.762	1.425	Y	1.673	1.835	Y	NA	NA	NA
Panel B Unregulated Tier-1 exchanges												
Binance	1.669	1.209	Y	1.795	1.436	Y	1.836	1.411	Y	1.960	1.430	Y
Bittrex	1.911	1.671	Y	1.582	1.880	Y	1.807	1.497	Y	1.798	1.722	Y
Bitfinex	1.680	1.277	Y	1.719	1.425	Y	1.815	1.397	Y	1.948	1.430	Y
HitBTC	0.620	0.663	N	0.785	0.790	N	0.692	0.879	N	0.552	0.803	N
Huobi	1.750	1.089	Y	1.842	1.505	Y	1.871	1.447	Y	1.966	1.651	Y
KuCoin	3.325	1.656	N	3.014	1.609	N	4.563	5.865	N	5.976	5.579	N
Liquid	1.406	0.905	N	1.494	1.358	Y	NA	NA	NA	1.282	1.231	Y
Okex	1.680	0.949	N	1.675	1.020	Y	1.863	1.320	Y	1.812	1.212	Y
Poloniex	1.629	1.008	Y	1.615	1.816	Y	1.662	1.428	Y	1.804	1.470	Y
Zb	1.479	1.095	Y	1.841	1.417	Y	1.546	0.932	N	1.634	1.194	Y
Panel C Unregulated Tier-2 exchanges												
Bgogo	1.333	2.760	N	3.345	3.941	N	NA	NA	NA	NA	NA	NA
Biki	5.197	7.155	N	10.428	7.076	N	1.739	2.046	N	2.194	1.469	N
BitZ	2.374	2.702	N	2.035	1.546	N	2.014	4.005	N	2.202	4.452	N
Coinbene	4.546	2.724	N	4.716	3.573	N	7.165	4.137	N	6.356	4.157	N
DragonEX	2.269	1.701	N	4.367	1.773	N	0.641	1.299	N	8.689	4.863	N
Lbank	1.760	1.638	Y	1.998	1.622	Y	NA	NA	NA	NA	NA	NA
Mxc	7.660	7.063	N	3.598	11.444	N	14.815	11.706	N	12.439	6.862	N
Fcoin	1.020	0.952	N	1.157	0.874	N	1.241	0.765	N	0.656	0.650	N
Exmo	1.370	3.770	N	1.520	3.087	N	NA	NA	NA	1.486	6.373	N
Coinmex	4.292	7.578	N	7.384	7.966	N	5.049	8.802	N	10.697	13.863	N
Bibox	5.829	6.384	N	3.639	5.961	N	3.676	4.877	N	7.116	5.027	N
Bitmart	2.854	1.728	N	1.926	1.880	Y	1.572	1.226	Y	1.831	2.691	N
Bitmax	1.509	1.022	Y	1.669	1.191	Y	1.479	1.193	Y	1.434	1.180	Y
Coinegg	0.718	1.261	N	2.031	1.237	N	1.077	1.056	Y	6.551	10.524	N
Digifinex	1.537	1.038	Y	1.618	1.117	Y	1.679	1.129	Y	1.548	1.001	Y
Gateio	2.048	1.631	N	1.925	1.954	Y	2.173	2.430	N	2.175	2.074	N

Table 5. Chi-squared Test for Trade-size Roundness of Unregulated Exchanges

Table 5 presents the results of Pearson's Chi-squared test on the roundness of unregulated exchanges, with regulated exchanges as a benchmark. Results of unregulated Tier-1 and Tier-2 exchanges are shown in Panel A and Panel B, respectively. The level of roundness is a parameter describing the decimal or integer places of the last non-zero digit. Test results, χ^2 statistics and *p*-values, reveal the difference in distributions between regulated and unregulated exchanges. ***, **, and * denote the statistical significance levels at 1%, 5%, and 10%, respectively.

Exchange	BTC/USD		ETH/USD		LTC/USD		XRP/USD	
	χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
Panel A Unregulated Tier-1 exchanges								
Binance	9.545	0.145	15.013**	0.020	12.18**	0.032	11.993***	0.007
Bittrex	3.100	0.796	11.455*	0.075	9.222	0.101	13.387***	0.004
Bitfinex	92.104***	0.000	8.086	0.232	5.616	0.345	51.094***	0.000
HitBTC	17.224***	0.008	13.387**	0.037	7.547	0.183	11.393***	0.010
Huobi	115.48***	0.000	11.01*	0.088	14.311**	0.014	9.5**	0.023
KuCoin	7.909	0.245	17.469***	0.008	24.886***	0.000	16.603***	0.001
Liquid	182.435***	0.000	16.518**	0.011	NA	NA	49.766***	0.000
Okex	4.384	0.625	15.649**	0.016	19.46***	0.002	12.18***	0.007
Poloniex	3.247	0.777	5.427	0.490	11.906**	0.036	14.268***	0.003
Zb	1461.8***	0.000	692.292***	0.000	21.797***	0.001	18.032***	0.000
Panel B Unregulated Tier-2 exchanges								
Bgogo	18.774***	0.005	32.402***	0.000	NA	NA	NA	NA
Biki	60.923***	0.000	62.726***	0.000	28.101***	0.000	19.651***	0.000
BitZ	828.828***	0.000	85.86***	0.000	22.242***	0.000	19.593***	0.000
Coinbene	1670.819***	0.000	31.158***	0.000	32.097***	0.000	19.747***	0.000
DragonEX	1668.236***	0.000	20.761***	0.002	27.753***	0.000	19.109***	0.000
Lbank	1639.493***	0.000	24.944***	0.000	NA	NA	NA	NA
Mxc	9.569	0.144	15.481**	0.017	18.705***	0.002	19.688***	0.000
Fcoin	740.835***	0.000	157.443***	0.000	86.741***	0.000	18.59***	0.000
Exmo	15.455**	0.017	26.838***	0.000	NA	NA	19.182***	0.000
Coinmex	1719.65***	0.000	23.694***	0.001	32.242***	0.000	19.796***	0.000
Bibox	439.322***	0.000	101.26***	0.000	14.106**	0.015	19.458***	0.000
Bitmart	18.605***	0.005	28.754***	0.000	22.785***	0.000	19.768***	0.000
Bitmax	26.08***	0.000	130.687***	0.000	41.623***	0.000	34.596***	0.000
Coinegg	1310.242***	0.000	34.176***	0.000	30.144***	0.000	19.728***	0.000
Digifinex	1546.727***	0.000	23.247***	0.001	29.609***	0.000	19.592***	0.000
Gateio	535.379***	0.000	55.367***	0.000	13.247**	0.021	15.288***	0.002

Table 6. Estimating the Fraction of Wash Trades

Table 6 reports the pooled regression results of the fraction of wash trading for unregulated exchanges. The regression equation below specifies the relationship between round and unrounded trade volumes.

$$\ln\left(V_{Unrounded_{it}}\right) = \alpha + \beta * \ln\left(V_{Round_{it}}\right) + \gamma * X_{it} + \epsilon_{it}$$

where $\ln\left(V_{Round_{it}}\right)$ and $\ln\left(V_{Unrounded_{it}}\right)$ are the logarithms of round trade volume and unrounded trade volume, respectively, for exchange *i* at week *t*. X_{it} is a vector of exchange characteristics and ϵ_{it} is an error term. We categorize trading volume into round and unrounded ones by checking if the mantissa of a particular transaction volume is less than 100 base units or not. Exchange characteristics such as age, rank, CoinMarketCap web traffic percentage, and unique visitors are used as control variables. Exchange Biki and Mxc do not have data on control variables. The regression coefficients are used as a benchmark to calculate the expected unrounded trading volume, then the fraction of wash trading for each unregulated exchange. Fractions of wash trading are estimated for each cryptocurrency of each exchange (Panel B and C for unregulated Tier 1 and 2 exchanges, respectively) and then aggregated amount (Panel A) using equal- and volume-weighted averages. A thousand bootstrapped samples are used to calculate the standard deviation of wash trading estimates, which we report in brackets.

Panel A: Aggregated Wash Trading Percentage

	Wash Trade Percentage Without Control Variables		**Wash Trade Percentage With Control Variables**	
	Equal-weighted Average	Volume-weighted Average	Equal-weighted Average	Volume-weighted Average
Unregulated	70.85	77.50	60.96	71.43
Unregulated Tier-1	53.41	61.86	46.95	63.62
Unregulated Tier-2	81.76	86.26	70.96	76.96

Panel B: Wash Trading Percentage for Unregulated Tier-1 Exchanges

Exchange	**Wash Trade Percentage No Control**	**Wash Trade Percentage With Control**
Binance	51.76 (1.28)	46.47(1.34)
Bittrex	51.73 (1.65)	18.91(2.34)
Bitfinex	1.87 (0.52)	31.34(2.06)
HitBTC	92.60 (0.66)	89.81(1.93)
Huobi	44.87 (2.08)	57.77(1.69)
KuCoin	74.36 (1.30)	52.96(6.67)
Liquid	19.02 (1.55)	3.02(1.41)
Okex	66.12 (1.52)	72.75(2.02)
Poloniex	37.49 (2.46)	14.94(2.19)
Zb	94.31 (0.54)	81.49(4.20)

Panel C: Wash Trading Percentage for Unregulated Tier-2 Exchanges

Exchange	Wash Trade Percentage No Control	Wash Trade Percentage With Control
Bgogo	99.99 (0.00)	99.93(0.01)
Biki	99.36 (0.13)	NA
BitZ	72.72 (2.41)	72.62(2.18)
Coinbene	95.50 (0.52)	91.64(1.51)
DragonEX	89.71 (0.39)	72.48(2.55)
Lbank	98.13 (0.21)	98.65(0.11)
Mxc	82.00 (3.68)	NA
Fcoin	77.09 (2.17)	48.62(5.32)
Exmo	81.12 (4.21)	64.99(3.85)
Coinmex	98.45 (0.09)	86.12(2.27)
Bibox	34.32 (6.57)	33.63(5.75)
Bitmart	98.10 (1.07)	94.79(2.04)
Bitmax	65.42 (2.12)	61.71(2.21)
Coinegg	96.80 (1.10)	81.24(3.18)
Digifinex	94.36 (0.48)	68.66(5.38)
Gateio	25.04 (4.49)	18.42(4.47)

Table 7. Price Impacts of Wash Trading

Table 7 presents the results of the regression analysis. In Panel A, the dependent variable is the weekly returns for every cryptocurrency on every exchange. In Panel B, the price deviation is calculated as the (log) difference between the close price of each unregulated exchange and averaged close prices of regulated exchanges at the same time. In both panels, wash trading volume is calculated as weekly wash trading percentage times volume. Exchange Age_t is the time span from its establishment to week t for an exchange. Tier-1 Exchange is a dummy variable that equals 1 if the exchange is unregulated Tier-1 exchange, 0 otherwise. The number of unique visitors refers to the number of distinct visitors recorded during the sample period, derived from SimilarWeb August to October 2019 reports. $CoinMarketCapRank_t$ is the rank directly obtained from CoinMarketCap. All models are estimated with random effects based on the Hausman test. Currency fixed effects are included in Model 2, 4, 6, 8, 10, and 12 of Panel A, and Model 2 and 4 of Panel B. Standard errors are clustered at the exchange-currency level. *t*-statistics are reported in the brackets. ***, **, and * denote the statistical significance levels at 1%, 5%, and 10%, respectively.

Panel A: Returns and Wash Trading

	Weekly $return_t$											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(log) wash trade $volume_t$	0.001***	0.001	0.003***	0.002***					0.024***	0.023***	0.024***	0.023***
	(2.61)	(1.33)	(3.11)	(2.96)					(4.75)	(4.68)	(4.66)	(4.61)
(log) wash trade $volume_{t-1}$					-0.001***	-0.002***	-0.002***	-0.003***	-0.024***	-0.024***	-0.024***	-0.024***
					(-2.95)	(-4.80)	(-3.37)	(-4.48)	(-4.83)	(-4.83)	(-4.70)	(-4.71)
Exchange Age_t			-0.000	0.000			-0.000	-0.000			0.000	0.000
			(-0.07)	(0.17)			(-0.29)	(-0.34)			(0.02)	(0.50)
Tier-1 Exchange			0.001	-0.000			0.005	0.004			0.001	-0.001
			(0.29)	(-0.12)			(1.46)	(1.56)			(0.33)	(-0.41)
(log) Number of Unique Visitors			-0.000	0.000			-0.001	-0.001			-0.000	0.000
			(-0.30)	(0.28)			(-1.15)	(-1.25)			(-0.26)	(0.71)
CMC $rank_t$			0.000***	0.000***			-0.000	-0.000**			0.000	0.000
			(3.23)	(3.68)			(-1.24)	(-2.09)			(1.24)	(1.32)
Constant	-0.049***	-0.027***	-0.080***	-0.070***	0.010	0.030***	0.037**	0.067***	-0.008	0.010	-0.016	-0.003
	(-5.15)	(-3.11)	(-3.35)	(-3.60)	(1.28)	(3.49)	(2.26)	(3.50)	(-1.14)	(1.46)	(-0.93)	(-0.23)
Currency Fixed Effects	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Observation	1416	1416	1328	1328	1326	1326	1246	1246	1305	1305	1225	1225
Overall R^2	0.1%	1.0%	0.4%	1.2%	0.1%	1.1%	0.2%	1.2%	3.1%	4.0%	3.3%	4.1%

Panel B: Price Deviations and Wash Trading

	$PriceDeviation_t$		$PriceDeviation_{t+1}$ - $PriceDeviation_t$	
	(1)	(2)	(3)	(4)
(log) wash trade $volume_t$	0.046***	0.040***	-0.048***	-0.051***
	(3.33)	(2.59)	(-3.97)	(-3.31)
Exchange Age_t	-0.000	-0.000	0.000	0.000
	(-0.19)	(-0.00)	(0.96)	(0.89)
Tier-1 Exchange	0.095	0.080	-0.160	-0.155
	(1.25)	(1.12)	(-1.55)	(-1.47)
(log) Number of Unique Visitors	-0.021	-0.017	0.020	0.020
	(-1.15)	(-1.10)	(1.08)	(1.01)
CMC $rank_t$	0.005***	0.005***	-0.004***	-0.004***
	(4.60)	(4.16)	(-3.55)	(-3.17)
Constant	-1.112***	-0.968**	1.080***	1.147**
	(-3.01)	(-2.35)	(3.02)	(2.50)
Currency Fixed Effects	N	Y	N	Y
Observation	1328	1328	1246	1246
Overall R^2	0.7%	1.1%	0.4%	0.5%

Table 8. Wash Trading and Exchange Characteristics

Table 8 reports the cross-sectional regression analysis for the relationship between the fraction of overall wash trading volume for an exchange and its characteristics. Exchange Age is the span between the establishment date and July 2019, the start of our sample period. The remaining indicators are derived from SimilarWeb August to October 2019 reports. The number of unique visitors refers to the number of distinct visitors recorded during the sampling period. Top 5 traffics from lower GDP countries refers to the number of traffic countries ranked at the bottom 15 countries based on GDP. Top 5 traffics from worst finance access countries denotes the number of traffic countries ranked at the bottom 15 countries based on financial access. GDP and financial access data are obtained from the World Bank DataBank. The rank of countries is based on the average value of GDP and financial access over three years from 2016 to 2018. Robust standard errors are calculated**.** *t*-statistics are reported in the brackets. ***, **, and * denote the statistical significance levels at 1%, 5%, and 10%, respectively.

Fraction of wash trades	Unregulated exchange		
	(1)	(2)	(3)
Exchange Age	-0.660***		-0.679***
	(-2.99)		(-3.08)
Number of Unique Visitors		-0.099**	-0.091***
		(-2.12)	(-3.69)
Top 5 Traffics from Lower GDP Countries			3.152
			(0.65)
Top 5 Traffics from Worst Financial Access Countries			4.956
			(0.92)
Constant	94.500***	72.995***	87.263***
	(11.53)	(11.69)	(8.10)
Observations	26	26	26
Adjusted R^2	28.4%	1.0%	30.1%

Table 9. Influence of Returns and Volatility on Wash Trading Volumes

Table 9 presents the panel regression results for the impact of weekly cryptocurrency returns and volatility on wash trading volumes of unregulated exchanges. The weekly returns and volatility are calculated based on the third-party composite price indexes from CoinMarketCap (CMC). CMC Volatility$_{t-1}$ is the standard deviation of daily returns during week *t-1*. Random-effect models with robust errors are used in all regressions. Standard errors are clustered at exchange-currency level. *t*-statistics are reported in the brackets. ***, **, and * denote the statistical significance levels at 1%, 5%, and 10%, respectively.

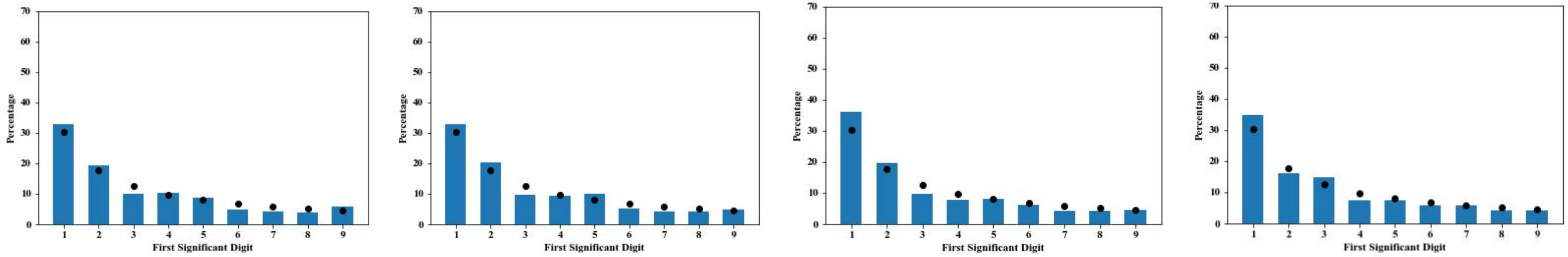
(log) Wash Trade Volume$_t$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Weekly CMC Return$_{t-1}$	1.258***		1.444***				1.415***
	(7.14)		(7.68)				(7.16)
Weekly CMC Return$_{t-2}$		0.318**	0.627***				0.350**
		(2.09)	(3.95)				(2.22)
CMC Volatility$_{t-1}$				-5.717***		-5.636***	-4.116***
				(-6.06)		(-6.03)	(-4.35)
CMC Volatility$_{t-2}$					-2.297**	-2.070**	-3.547***
					(-2.18)	(-2.00)	(-3.15)
(log) Wash Trade Volume$_{t-1}$	0.887***	0.882***	0.886***	0.885***	0.882***	0.884***	0.885***
	(48.67)	(47.61)	(47.93)	(50.07)	(47.86)	(49.38)	(48.56)
Constant	2.304***	2.386***	2.352***	2.543***	2.459***	2.632***	2.619***
	(6.62)	(6.71)	(6.64)	(7.21)	(6.80)	(7.19)	(7.10)
Observation	1305	1305	1305	1305	1305	1305	1305
Overall R^2	92.9%	92.7%	93.0%	92.9%	92.8%	93.0%	93.2%

Figure 1 First-significant-digit Distribution and Benford's Law

Figure 1 displays the first-significant-digit distributions of trading data in bar charts. Black dots represent distributions derived from Benford's law. Coinbase, KuCoin, Fcoin, Exmo, and Coinegg are five exchanges selected from regulated (Panel R), Tier-1 unregulated (Panel UT) and Tier-2 unregulated (Panel U) exchanges, respectively.

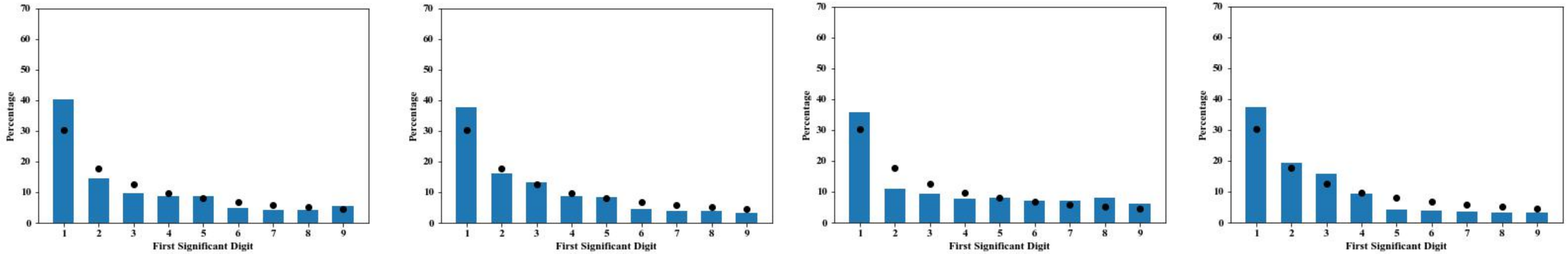
Panel R: Regulated Exchanges
Coinbase

BTC/USD ETH/USD LTC/USD XRP/USD



Panel UT: Unregulated Tier-1 Exchanges
KuCoin

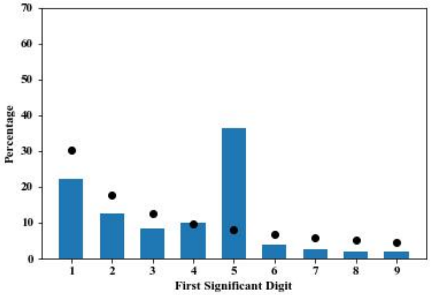
BTC/USD ETH/USD LTC/USD XRP/USD



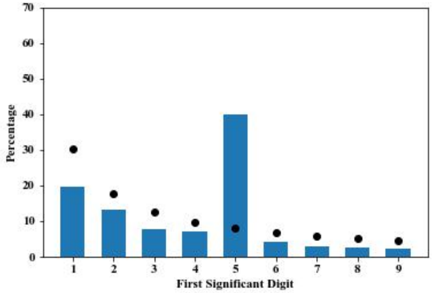
Panel U: Unregulated Tier-2 Exchanges

Fcoin

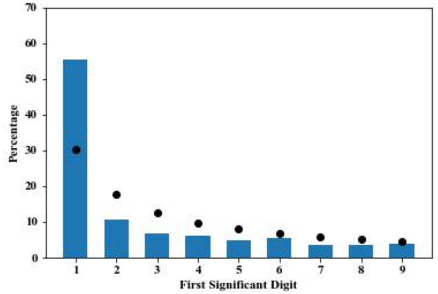
BTC/USD



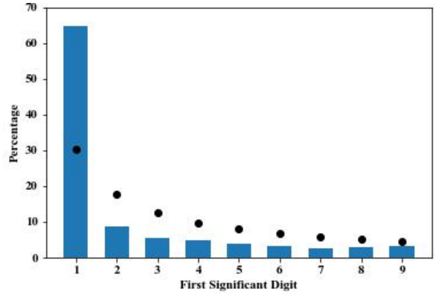
ETH/USD



LTC/USD

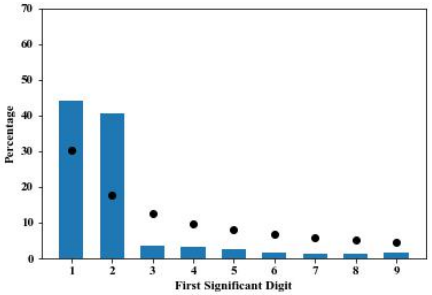


XRP/USD

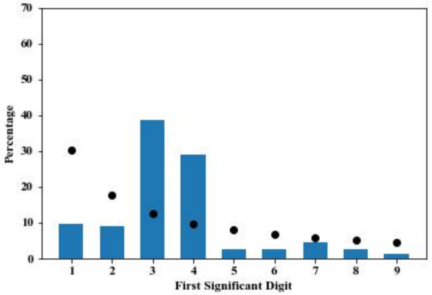


Exmo

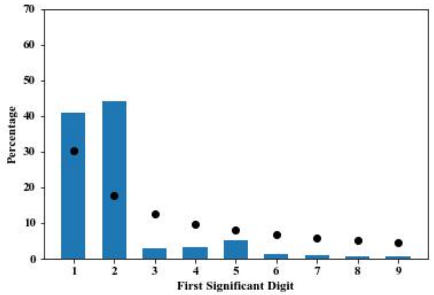
BTC/USD



ETH/USD

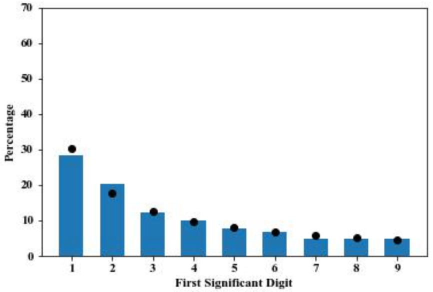


XRP/USD

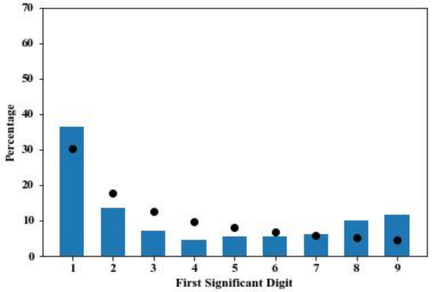


Coinegg

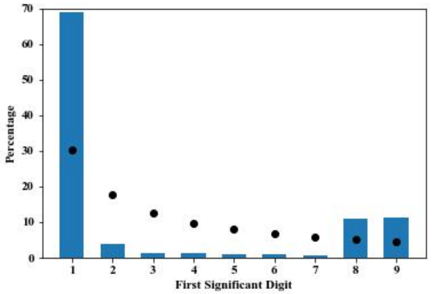
BTC/USD



ETH/USD



LTC/USD



XRP/USD

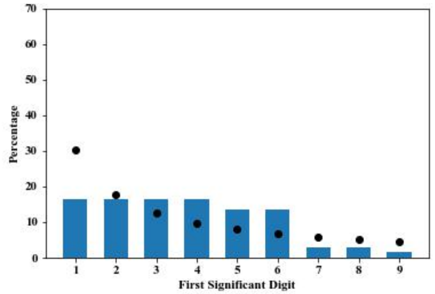
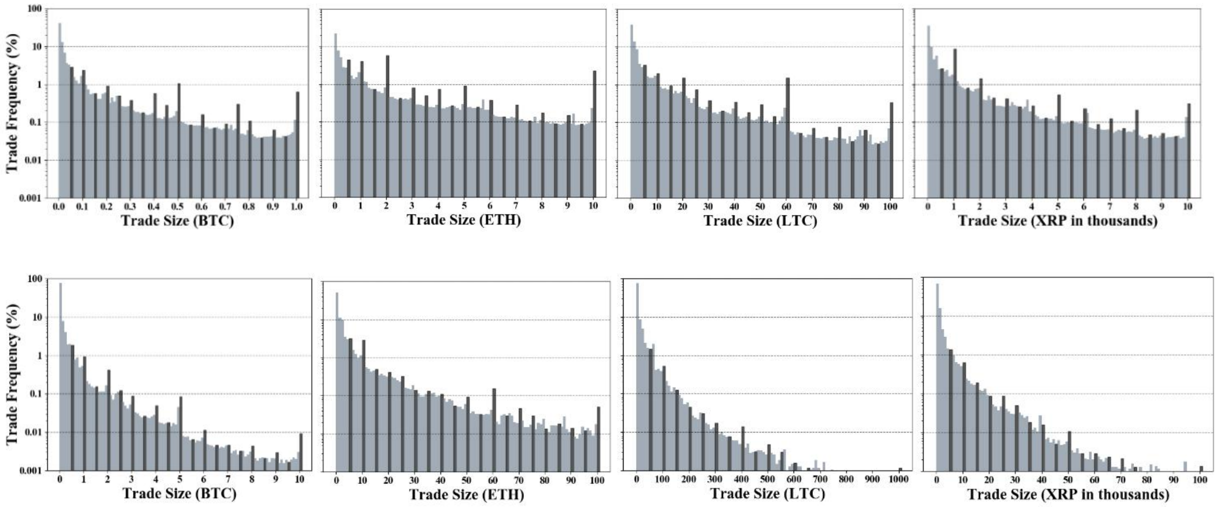


Figure 2. Trade-size Clustering

The clustering effect in trade-size distribution histograms on sample exchanges. Two sets of observation ranges are applied for each trading pair: 0-1 BTC, 0-10 BTC, 0-10 ETH, 0-100 ETH, 0-100 LTC, 0-1,000 LTC, 0-10,000 XRP, and 0-100,000 XPR. In each histogram, we highlight every 5th and 10th bin to illustrate the clustering effect.

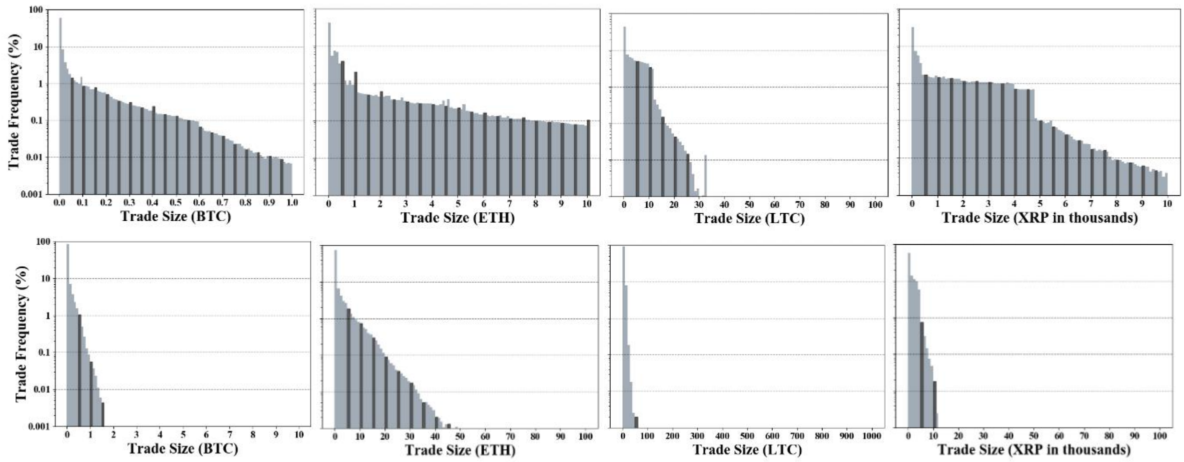
Panel R: Regulated Exchanges

Coinbase



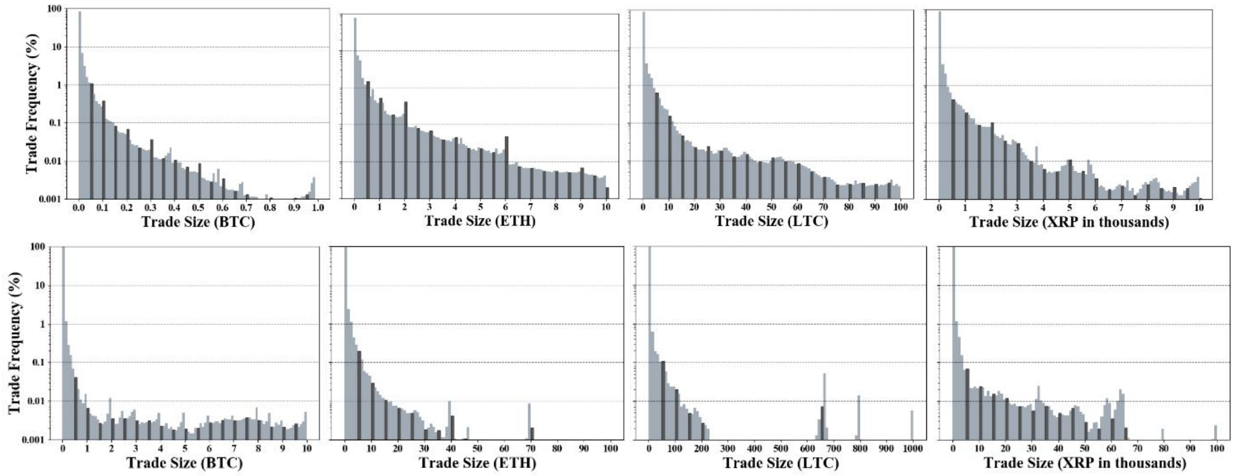
Panel UT: Unregulated Tier-1 Exchanges

KuCoin

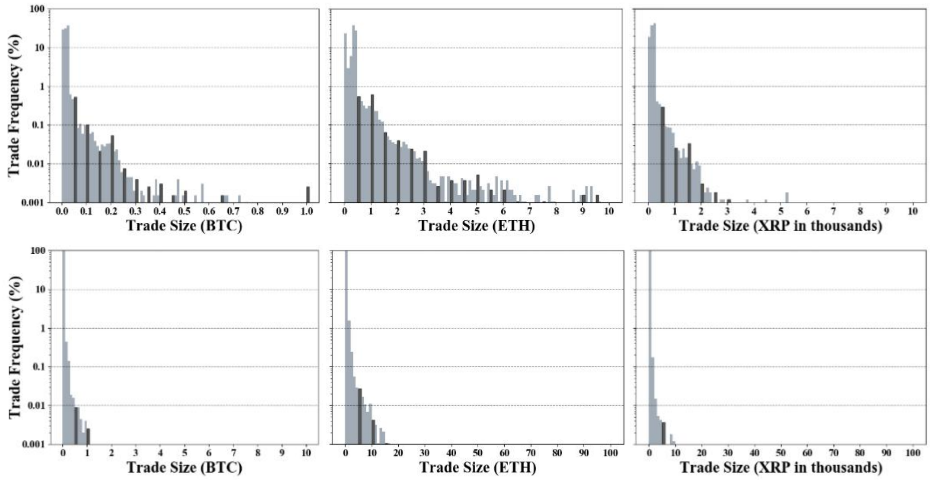


Panel U: Unregulated Tier-2 Exchanges

Fcoin



Exmo



Coinegg

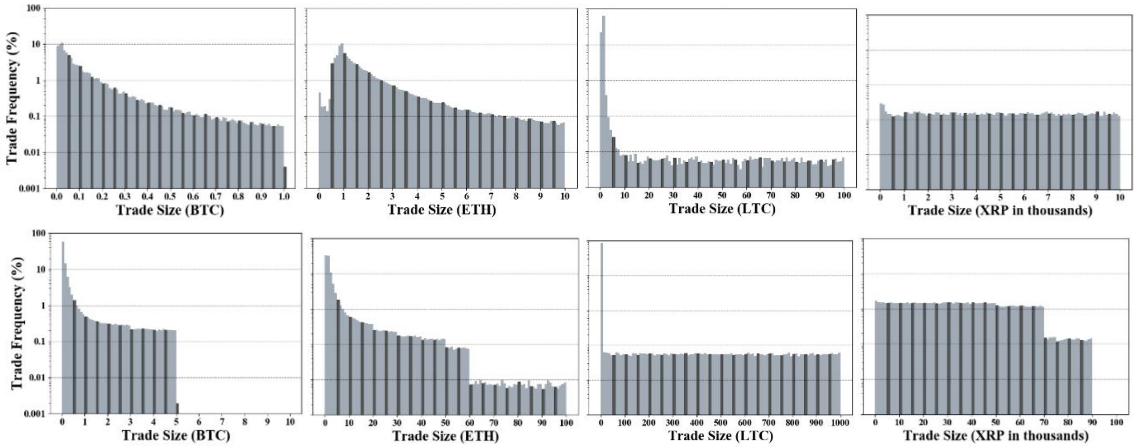


Figure 3. Illustration of the *t*-test for Clusters

Trade frequencies at round trade-sizes are tested against unrounded trade-sizes nearby. Frequency for trade-size *i* is calculated as the number of trades with size *i* over the total number of trades in an observation window (e.g., *i*-50 to *i*+50). Frequencies at round trade sizes (e.g., the 200th unit) and the highest frequencies of nearby unrounded trades (e.g., the 160th unit) are recorded as a pair. The *t*-test on the difference between round and unrounded frequencies in a pair is then carried out over a sample of all pairs.

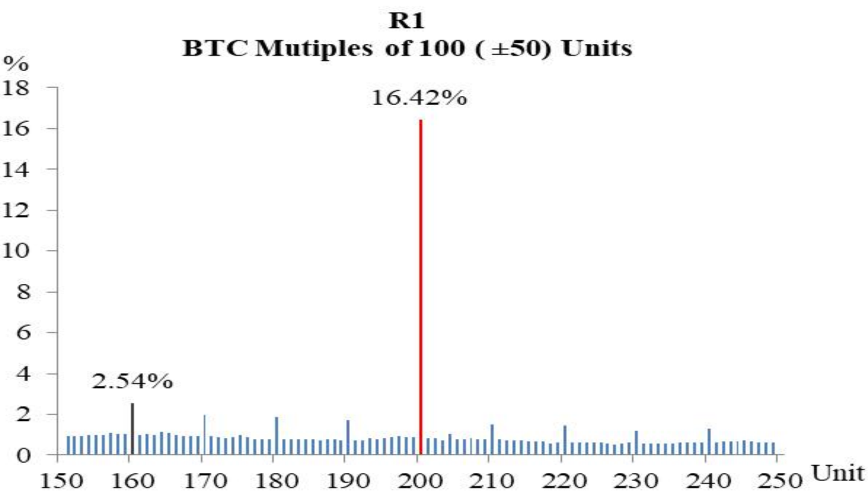
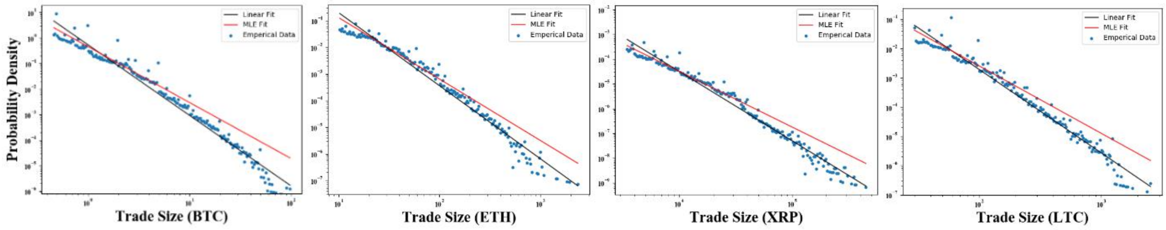


Figure 4. Tail Distribution and Power-law Fitting

Figure 4 displays the tails of trade-size distributions and the fitted power-law lines on log-log plots. The 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.

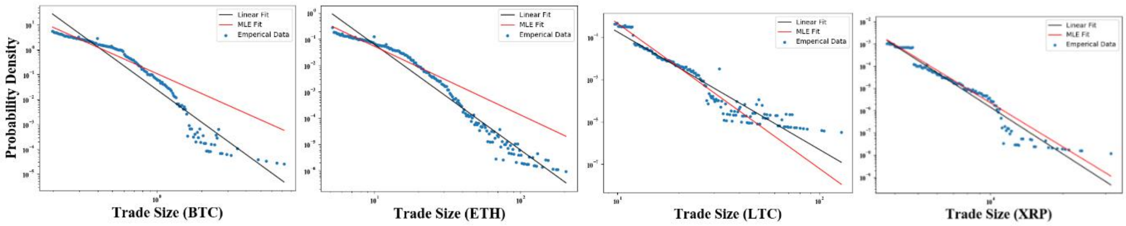
Panel R: Regulated Exchange

Coinbase



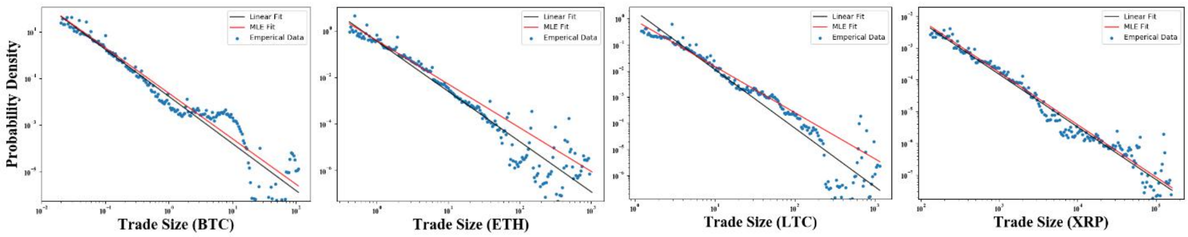
Panel UT: Unregulated Tier-1 Exchanges

KuCoin

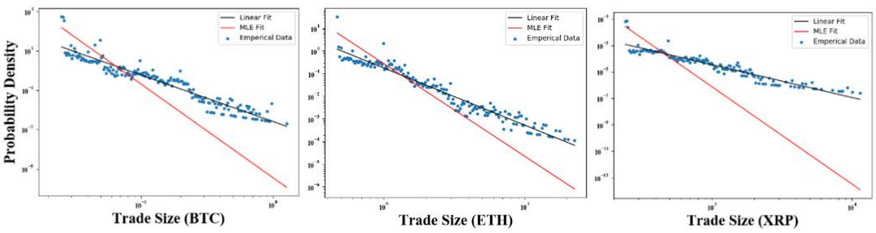


Panel U: Unregulated Tier-2 Exchanges

Fcoin



Exmo



Coinegg

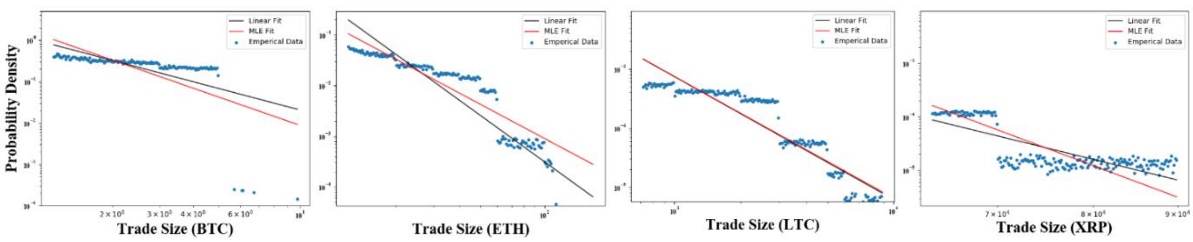


Figure 5. Trading Volumes and Ranks

Figure 5 plots the quantitative relationship between (logarithm) trade volumes and exchange ranks. Data fitting is carried out with Ordinary Least Square (OLS) regression. The estimated coefficients are reported below (t-statistics in brackets) with an adjusted R^2 of 93%.

$$\text{Exchange rank}_i = 416.269 - 19.202 * \log(\text{Volume}_i) + \varepsilon_i$$

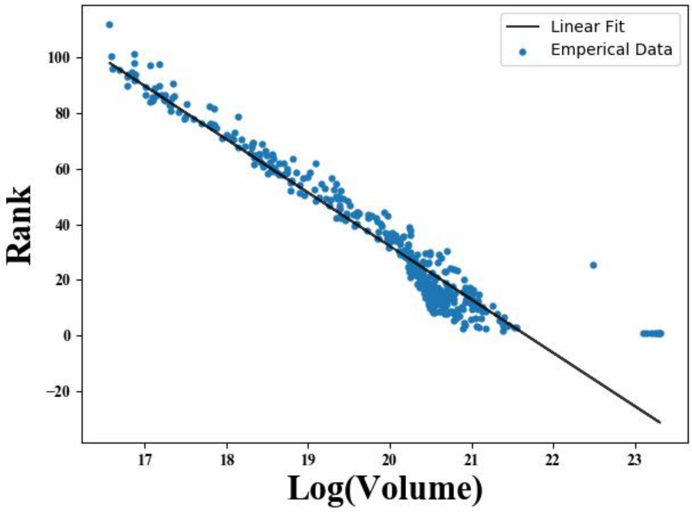


Figure 6. Improvement in Ranks and Wash Trading

Figure 6 plots the relationship between the estimated fraction of wash trading and the improvement in counterfactual ranks. The counterfactual rank is estimated based on the estimated "real" volume for any specific exchange, i.e., the difference between the reported volume in CoinMarketCap and the estimated wash trading volume, using the volume-rank relationship documented in Figure 5. Rank improvement is the difference between the counterfactual rank and the reported rank in CoinMarketCap.