



Determinants of wash trading in major cryptoexchanges

Jan Sila^{ID}, Evzen Kocenda^{ID}, Ladislav Kristoufek, Jiri Kukacka^{ID}*

Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, 182 00 Prague 8, Czechia
Charles University, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00 Prague 1, Czechia

ARTICLE INFO

Dataset link: <http://www.tardis.dev>, <http://www.coinmetrics.io>, <http://www.fred.stlouisfed.org>, <http://www.investing.com>, <http://www.spglobal.com>, <http://trends.google.com>, <http://pageviews.wmcloud.org>, <http://www.policyuncertainty.com>

JEL classification:

C58
E44
G12
G23
O35

Keywords:

CeFi
Cryptocurrency
Fake trading volume
Fraud detection
Market manipulation
Regulation

ABSTRACT

We investigate a crypto-market-wide phenomenon of wash trading that affects exchange integrity and the accuracy of liquidity claims. We examine four leading cryptocurrencies using a dataset that spans from November 15, 2020, to January 31, 2022. We employ two detection approaches to assess the extent of wash trading: a trade-size roundness metric and a Benford's law-based deviation metric. We examine more than 40 different explanatory variables, including blockchain and cryptocurrency measures, as well as financial and macroeconomic factors. Variable selection is conducted using a robust combination of the variance inflation factor and Bayesian model averaging. Our findings indicate that market volatility and public sentiment are robustly associated with wash trading, with its volume tending to rise under volatile conditions on which exchanges may capitalize. The models in our study provide valuable insights for regulators and market participants in identifying and mitigating such practices, thereby enhancing market integrity and investor confidence.

1. Introduction

Cryptocurrency markets have experienced immense growth in the last two decades, evolving from an individual initiative launched by Nakamoto (2008) to specialized trading platforms and global payment ecosystems (Corbet et al., 2019). In particular, El Salvador's adoption of Bitcoin as a legal tender in September 2021 presents a quasi-natural economic experiment, demonstrating the applicability of digital currencies in real-world monetary practice (Alvarez et al., 2023). However, this rapid development has been accompanied by numerous fraud

cases, malicious behaviors, and market manipulation, with wash trading being particularly widespread. Our analysis focuses on wash trading on centralized exchanges, its dynamics, and the patterns behind it.

Wash trading occurs when a single entity or a coordinated group buys and sells a financial asset either simultaneously or within a short time frame without facing any actual market risk, with the intent of artificially inflating trading volume and creating a misleading impression of market activity or asset's liquidity.¹ Although illegal on conventionally regulated asset exchanges,² it has been widely documented

* Corresponding author at: Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, 182 00 Prague 8, Czechia.
E-mail addresses: jan.sila@fsv.cuni.cz (J. Sila), evzen.kocenda@fsv.cuni.cz (E. Kocenda), ladislav.kristoufek@fsv.cuni.cz (L. Kristoufek), jiri.kukacka@fsv.cuni.cz (J. Kukacka).

¹ The Commodity Futures Trading Commission provides the following precise legal definition of wash trading: "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." It is also referred to as 'round-trip trading,' 'self-trading,' or 'wash sales.' Another useful definition comes from the Chicago Mercantile Exchange Group Market Regulation Advisory Notice RA1308-5 from 2013, suggesting that "A wash trade requires: (a) that the transaction or series of transactions produces a wash result - meaning the purchase and sale of the same instrument at the same price, or a similar price, for accounts with [...] common beneficial ownership; and (b) that the party(ies) intended to achieve a wash result. [...]"

² Interestingly, in the U.S., wash trading was prohibited already by the U.S. Securities and Exchange Act and the Commodity Exchange Act from 1934 and 1936, respectively.

on cryptocurrency exchanges, especially on smaller or newly established platforms seeking to expand their user base and operating with minimal regulatory oversight (Amiram et al., 2025; Cong et al., 2023). Wash trading has become a widespread issue in cryptocurrency markets. From the leaked Bitcoin individual transaction data from the Mt. Gox exchange era, Aloosh and Li (2024) traced the origins of crypto wash trading back to June 26, 2011, and provided direct evidence showing that although fabricated transactions represented only more than 2% of trades, they may have accounted for up to 60% of the daily volume in the period between June 2011 and May 2013. Next, Pennec et al. (2021) found that for a group of suspicious centralized exchanges, 96%–98% of the reported trading volume is highly questionable. Similarly, research by Cong et al. (2023) documented that wash trading is common in unregulated exchanges but not in regulated exchanges.³ On average, wash trades represent over 70% of the total trading volume across all unregulated exchanges (up to 77.5% if volume-weighted). These conclusions update the previous finding of Hougan et al. (2019), who was the first to demonstrate a very limited extent of wash trading activities on specific reputable exchanges while estimating that approximately 95% of the overall reported Bitcoin volume consists of wash trading. Finally, the main findings of Sifat et al. (2024) regarding nonhuman and suspicious trading activities in the nonfungible token (NFT) market align with the current research on cryptocurrencies.

The trading volume is a critical metric in cryptocurrency markets. High volume typically signals strong market interest in a given asset, hints at lower transaction fees, and helps ensure that prices accurately reflect underlying value. Recent academic research has persuasively empirically documented trading volume as one of the key determinants of the cryptocurrency pricing dynamics (Balcilar et al., 2017; Kukacka & Kristoufek, 2023; Li & Wang, 2017; Zhang & Zhao, 2023), volatility (Wang et al., 2023), and their spillovers (Andrada-Félix et al., 2020; Bouri & Jalkh, 2023; Charfeddine et al., 2022; Ji et al., 2019; Sila et al., 2024). Wash trading causes substantial negative externalities and misleading public sentiment, leading to misguided investment decisions and financial losses, especially for inexperienced investors and retail traders. It thus introduces significant economic inefficiencies in what could otherwise be a highly efficient market (Hougan et al., 2019). Wash trading generates misleading trading signals, distorts prices, and increases market volatility, which is known to be highly endogenous in cryptocurrency (Mark et al., 2022). As a result, it undermines market integrity and erodes the trust of participants in decentralized finance ecosystems, pushing industry data providers to offer adjusted metrics that reflect the ‘real’ trading volume (Aloosh & Li, 2024) or advanced indices of exchange quality that are not solely reliant on volume data (Amiram et al., 2025).

The decentralized nature of cryptocurrency markets, characterized by imperfect information and limited regulatory oversight, makes them particularly susceptible to wash trading. Using account-level blockchain data, Victor and Weintraud (2021) show that fully anonymous decentralized exchanges are especially vulnerable, with more than 30% of traded tokens exhibiting patterns consistent with wash trading. More recent evidence confirms that such behavior extends beyond traditional token markets. Specifically, Tošić et al. (2025) detect wash trading by exploiting transaction-level network structures in NFT markets, demonstrating that manipulative trading remains pervasive even in fully on-chain environments. Complementing this line of research, Wang et al. (2025) link wash trading activity to insider selling within pump-and-dump schemes, highlighting its role as a core mechanism

underlying financial misconduct in crypto markets. The lack of transparency on centralized exchanges, which currently play a major role in the segment, and the complex tactics used by coordinated groups add an additional layer of difficulty in detecting wash trading effectively. While the impact of wash trading in the past was largely restricted to the relatively small and detached cryptocurrency markets, this dynamic has recently shifted. With the U.S. Securities and Exchange Commission (SEC) approving the first Bitcoin and Ethereum ETFs in early 2024, cryptocurrencies are becoming more interconnected with traditional financial markets. This connection increases the risk of crypto wash trading affecting the global financial system and impacting both institutional and individual investors.

The economic motivations for engaging in wash trading vary considerably across different market participants. While in traditional financial markets, the primary motivation for wash trading typically involves individual investors, often for tax avoidance purposes (Grinblatt & Keloharju, 2004), in cryptocurrency markets, the primary drivers of artificially inflating trading volumes are believed to be the platforms themselves. Cryptocurrency exchanges aiming to establish their position in a highly competitive sector with largely homogeneous products might be tempted to wash trade to attract more users and improve their rankings on exchange comparison platforms such as www.coinmarketcap.com (Cong et al., 2023). This, in turn, leads to an increase in nonwash transaction fee revenues, as quantified, e.g., by Aloosh and Li (2024) for the historical Bitcoin data. They documented a notable immediate increase of approximately 1.5% in revenue following a positive shock from wash trading, based on a vector autoregression (VAR) impulse response function analysis, with the impacts fading out quickly over the following hour. Conversely, for a more recent sample spanning from 2013 to 2021, Amiram et al. (2025) concluded that while wash trading increases the popularity of exchanges in the short term, it has considerable negative long-term reputational trade-offs. Cryptocurrency exchanges may achieve these impacts either directly by fabricating trading volumes themselves, using naive round-trip transactions or sophisticated automated algorithms, or indirectly by establishing perverse incentives, such as direct payments, fee discounts, refunds, or preferential rights to encourage other participants to engage in wash trading (Cong et al., 2023).

Similarly, market makers may respond to these incentives or use such practices to attract more customers. By creating the illusion of liquidity, they make an asset appear more active and reduce the perceived risk of illiquidity. Next, individual cryptocurrency projects may engage in wash trading to inflate the perceived demand for their token, manipulating its price for individual financial gain and attracting more investors or venture capital. Newly launched projects, in particular, may need to meet specific trading volume criteria to secure listings on larger exchanges, providing further motivation for wash trading (Victor & Weintraud, 2021). Finally, individual cryptocurrency investors most likely engage in wash trading to exploit the benefits offered by exchanges, such as rewards for achieving certain trading thresholds or high trading volumes. Each of these types of participants might individually benefit from the distorted perception of market activity, although at the expense of market transparency and trust.

The main contributions of this paper are as follows: the existing literature has focused primarily on developing methodologies to detect wash trading using qualitative approaches or indirectly quantifying its occurrence on individual exchanges, largely to guide and support regulatory efforts (Fusaro & Hougan, 2019) in the cryptocurrency sector. In contrast, we adopt a broader economic approach and aim to position wash trading in the global economic context. This study, therefore, examines wash trading as a market-wide phenomenon in centralized exchanges and explores its dynamics empirically, utilizing a large, daily frequency dataset of publicly available variables from both the cryptocurrency segment and the world economy as a whole. Furthermore, it assesses the predictive potential of the variables selected by the Bayesian model averaging (BMA) methodology to guide

³ Regulated exchanges consist of those obtaining the BitLicense by the New York State Department of Financial Services, one of the pioneering agencies implementing regulations on cryptocurrencies. The details about the requirements and costs of the application process can be found at www.coindesk.com.

market participants and regulators in identifying future periods with a high likelihood of wash trading, which was identified as a critical next research step by [Aloosh and Li \(2024\)](#).

The structure of the paper is as follows: Section 2 introduces the two metrics used to quantify daily wash trading activity. It also previews other key detection metrics and tools established in the existing literature. Section 3 outlines the econometric methodology employed to identify the main empirical determinants of wash trading. Section 4 presents the empirical dataset used in the analysis. In Section 5, we identify the key factors driving wash trading, examine its dynamics, and assess the potential of these factors to predict future wash trading activity. Finally, Section 6 provides the conclusion of the paper.

2. Quantifying wash trading

This study employs two established detection methods to quantify the extent of wash trading on centralized cryptocurrency exchanges at a daily frequency. The first method, introduced by [Cong et al. \(2023\)](#), utilizes the behavioral tendency of human-generated trades to cluster at round sizes. Wash traders typically violate owing to the automated algorithmic nature of their manipulative activities or to escape detection. The second method, developed by [Amiram et al. \(2025\)](#), numerically evaluates the conformity of the transaction size distribution to Benford's law, an empirically observed statistical regularity that predicts the frequency of leading digits in real-life numerical datasets ([Benford, 1938](#)). The effectiveness of both methods in identifying wash trading activity was validated by [Aloosh and Li \(2024\)](#) using leaked, unique historical individual transaction data from the Mt. Gox exchange.

2.1. Trade volume roundness-based metric

Trade size clustering at round numbers is a widespread phenomenon observed in financial markets, including the largest stock and forex exchanges ([Grossman et al., 1997](#)). This well-documented behavioral pattern ([Rosch, 1975](#)) reflects a natural preference for simple, easy-to-remember figures in human decision-making, leading genuine traders to favor round execution prices and trade sizes such as multiples of 10. Such behavior is driven by psychological factors and cognitive biases, such as anchoring to a reference point ([Tversky & Kahneman, 1974](#)), as well as by market standards and conventions. For example, institutional investors often trade in standardized sizes and round lots to simplify trade execution, align with portfolio structures, and minimize transaction costs ([Grossman et al., 1997](#)).

In contrast, wash traders typically violate the clustering regularity of their trade sizes. From a technical point of view, this phenomenon is much less common in algorithmic trading conducted by automated systems and bots ([Mahmoodzadeh & Gençay, 2017](#)), which are frequently employed for wash trading. Moreover, wash traders may intentionally use random or irregular sizes to avoid the risk of detection, as simple volume faking strategies, suspicious repetitive activity, or conflicting rounded counterorders can be easily identified ([Chen et al., 2022](#)). Consequently, wash trading activities typically involve nonrounded transaction sizes.

Wash trading thus reduces or even eliminates the trade size clustering effect. [Cong et al. \(2023\)](#) thus suggested using the proportion of the market volume represented by unrounded trades to quantify the extent of wash trading activities. For the practical application of this methodology, it is essential to carefully specify the base unit for each cryptocurrency under study. Given that cryptocurrencies are often traded in small fractions, they recommend setting the base unit so that it represents the value of approximately one U.S. dollar. Considering cryptocurrency prices during the sample period covering roughly Q3:2019, this corresponds to 10^{-4} BTC, 0.001 ETH, 0.01 LTC, and 1 XRP in their study. It can subsequently be analyzed if trade size clustering occurs at multiples of 100 base units, representing a reasonable trading amount of approximately 100 U.S. dollars.

The challenge lies in the fact that nonrounded transaction sizes are also a common feature of other legitimate algorithmic trading activities, such as market making, high-frequency trading, and automated arbitrage. To differentiate between these legitimate sources of unrounded trading volume and wash trading, [Cong et al. \(2023\)](#) proposed benchmarking the legitimate volume of unrounded trades on the basis of activity observed on regulated exchanges, where their analysis identifies no systematic patterns of wash trading. The nonnegative discrepancy between all unrounded and legitimate unrounded trades on unregulated exchanges, the so-called 'abnormal' unrounded trading volume, then serves as an effective proxy for the volume attributed to wash trading.

For this purpose, we use Bitfinex, the only consensual exchange that [Cong et al. \(2023\)](#), [Hougan et al. \(2019\)](#), [Pennec et al. \(2021\)](#) suggested was not affected by wash trading manipulation. We follow the baseline procedure of [Cong et al. \(2023\)](#) and first estimate a simple linear regression model for the legitimate unrounded volume as follows:

$$\ln(V_{unrounded,t}) = \alpha + \beta \ln(V_{rounded,t}) + \varepsilon_t, \quad (1)$$

where $V_{unrounded,t}$ and $V_{rounded,t}$ are the unrounded and rounded trading volumes on Bitfinex on day t , respectively, whereas the rounded volumes are classified by assessing whether the modulo 100 base units of a certain trade size are zero or if the trade size is exactly one base unit. The estimated coefficients of the model are then applied to proxy the legitimate extent of unrounded trading volumes on the cryptocurrency exchanges under study on the basis of their rounded volumes. Finally, we subtract the legitimate unrounded volume from the overall unrounded volume to obtain the abnormal wash trading volume on day t and report this metric as a proportion of the overall daily trading volume.

2.2. Benford's law-based metric

While trade size roundness focuses on the last significant digit of transaction volumes, Benford's law ([Benford, 1938](#)) refers to the empirical distribution of the first digit. The law builds on the multiplicative nature of many real-world systems ([Li et al., 2004](#)), proposing that in natural datasets, smaller digits should appear more frequently in the leading position than larger ones, following a logarithmic distribution:

$$P_n = \log_{10}(n+1) - \log_{10}(n) = \log_{10}(1 + n^{-1}), \quad (2)$$

where $n \in 1, 2, \dots, 9$ and P_n denotes the probability of the first digit being n . In practice, this rule implies a decreasing occurrence of higher digits, with the digit '1' being the leading digit in approximately 30.1% of the cases, followed by the digit '2' in 17.6% of the cases. The digits '3' to '9' appear at the leading position in 12.5%, 9.7%, 7.9%, 6.7%, 5.8%, 5.1%, and 4.6% of the cases, respectively. Deviations from this distribution in empirical real-life datasets are often considered a signal of potential manipulation. [Cong et al. \(2023\)](#) argued that it is widely applied in various fields to detect fraudulent activity, particularly in macroeconomics ([Michalski & Stoltz, 2013](#)), accounting ([Amiram et al., 2015](#)), and engineering. Their simulations further illustrate its effectiveness in identifying wash trading activity by analyzing trade size distributions on cryptocurrency exchanges.

Building on this qualitative detection approach, [Amiram et al. \(2025\)](#) extended the application of Benford's law to a quantitative framework. They defined wash trading metrics using the mean absolute deviation (MAD), which quantifies the distance between the theoretical distribution of the first digit as prescribed by Benford's law and the empirical distribution observed in the data as follows:

$$MAD_{Benford} = \frac{1}{N} \sum_{n=1}^N |D_n|, \quad (3)$$

where $N = 9$ is the number of possible first digits, $n \in 1, 2, \dots, 9$, and D_n represents the difference between the theoretical probability of the

first digit being n from Benford's law and the empirical frequency observed across time. Essentially, this metric reflects the average absolute deviation between theoretically expected and observed frequencies. For example, a $MAD_{Benford} = 0.02$ corresponds to an average deviation of 2 percentage points up or down across all nine possible leading digits. This would be a sizable discrepancy, especially given that digits '5' through '9' should appear as the first digit in only 7.9% to 4.6% of cases, as previously discussed.

2.3. Other wash trading detection methodologies

While the literature on wash trading in cryptocurrency markets is sparse, several other methodologies have been developed to detect wash trading activities. However, most of these methods are not suitable for quantifying daily wash trading intensity on centralized exchanges. The reason is primarily their qualitative nature, which applies to entire datasets, or their reliance on blockchain data that are unavailable for centralized platforms.

For example, Cong et al. (2023) suggested that detecting wash trading volume can be enhanced by analyzing the fat-tailed nature of trade size distributions, characterized by power law dynamics similar to those observed in traditional financial markets. However, Aloosh and Li (2024) found this method to be inaccurate. An alternative approach was proposed by Amiram et al. (2025), who introduced an E-divisive with medians (EDM) algorithm to identify anomalous trading by detecting structural breaks in the data. Nevertheless, this method is not suitable for daily analysis, as it typically detects only a few structural breaks over a monthly period. Similarly, Aloosh and Li (2024) recommended exploiting the log-normal distribution of trading volumes and computing the deviation between theoretical and empirical distributions, akin to the $MAD_{Benford}$ approach. Nonetheless, they cautioned that the quantitative results may be sample dependent, making this method much better suited for only qualitative comparisons.

In addition, Victor and Weintraud (2021) developed a sophisticated volume-matching algorithm capable of identifying coordinated trading groups or accounts that execute seemingly unrelated transactions, which ultimately return to the same market positions, fulfilling the legal definition of wash trading. This method, however, is only applicable when individual trader or account data are identifiable, such as on decentralized exchanges using blockchain data or labeled transaction IDs. Chen et al. (2022) further contributed by defining three trading detection metrics on the basis of transaction numbers and amounts, but their approach similarly depends on combining on-chain and off-chain transaction data alongside exchange wallet data. Finally, Gu et al. (2022) identified key factors influencing transaction volumes and developed a deep learning prediction model to quantify the impact of these factors across different exchanges. The model can also be utilized to detect abnormal volumes; however, its dependence on extensive data presents practical challenges for daily wash trading detection.

3. Methodology

We process all trades in the available dataset for each exchange and asset, and resample them into daily aggregates on the basis of the UTC timestamps. During this process, we calculate two key metrics that estimate the extent of wash trading, as introduced in Section 2.1.

3.1. Wash trading metrics

First, we classify the 'roundness' of each trade by following the methodology outlined by Cong et al. (2023) to calculate the volume of rounded and unrounded trades. A trade is considered 'rounded' if its size is a multiple of 100 base units or if the size equals the base unit. The base unit is defined as the lot size of the contract, as explained in more detail in Section 2.1. However, for our dataset described in detail in Section 4, this base unit specification for individual cryptocurrencies

requires a slight adjustment owing to both the overall price growth in the segment since then and the lack of standardization across different exchanges. In line with (Cong et al., 2023), we strictly maintain the conceptual approach but define the base unit numerically as the minimum allowed transaction size for a given currency on each exchange.⁴ For example, on the current world's largest centralized cryptocurrency exchange, [Binance.com](https://binance.com), this translates to 10^{-5} BTC, 0.0001 ETH, 0.001 LTC, and 1 XRP.

For example, a trade with a volume of 1.040 LTC, where the base unit is 0.001, amounts to 1040 base units. Since this is not a multiple of 100, the trade is classified as unrounded. Conversely, trades of exactly 0.001 LTC or 0.1 LTC correspond to 1 base unit or 100 base units, respectively, and are thus classified as rounded. Importantly, we categorize trades on the basis of their size in units of cryptocurrency, not their value in USD or other denominations. This ensures consistency across different crypto assets, given their varying prices and trading behaviors.

Second, we determine how often each first significant digit appears in the trade sizes of both rounded and unrounded trades. For example, if the trade size is 0.00321 BTC, the first significant digit is 3. If the trade size is 123.45 XRP, the leading digit is 1. Whether the leading digit is in the integer or decimal part of the number is ignored, allowing all observations to be treated consistently.

3.2. Variable selection

Let us describe the procedure of selecting the relevant variables for the models, whereas a detailed description of all the variables and their origin follows in Section 4. We begin the selection of variables relevant to the models by categorizing them into two groups:

1. *General variables*: These are variables that are not specific to any single cryptocurrency, such as U.S. interest and inflation rates, commodity prices and financial market variables, among others.
2. *Asset-specific variables*: These variables pertain to individual cryptocurrencies and are labeled with suffixes corresponding to each asset, such as '_BTC' for Bitcoin or '_ETH' for Ethereum. If a variable has been logarithmically transformed, we add an additional '_l' suffix (e.g., '_BTC_l' or '_ETH_l').

When constructing the model for Bitcoin (BTC), we include all the general variables along with variables specific to Bitcoin and altcoins. Altcoins are shorthand for alternative coins to Bitcoin, which, in our case, encompasses ETH, LTC, and XRP. Examples of such variables are `Mcap_Alts_l`, `Momentum_Alts`, `Trades_Alts_l`, and `Volume_Alts_l`.

For models on other cryptocurrencies (i.e., Ethereum, Litecoin, and XRP), we include all the general variables and the asset-specific variables for that particular cryptocurrency. Additionally, we add all Bitcoin-specific variables. This approach recognizes Bitcoin's dominant influence on the cryptocurrency market and allows us to capture any spillover effects from Bitcoin to other assets.

We address multicollinearity in our dataset by employing a systematic variable elimination process based on the variance inflation factor (VIF). Specifically, we iteratively remove the variable with the highest VIF score and recalculate the VIFs for the remaining variables. This process is repeated until all variables have a VIF less than or equal to 10, a commonly accepted threshold indicating acceptable levels of multicollinearity (Dodge, 2008). We implement this procedure using the `car` package in R by Fox and Weisberg (2019).

⁴ Aloosh and Li (2024) emphasized the importance of carefully selecting base units, considering individual trade sizes, particularly when applied over an extended time horizon where prices can change significantly across several orders of magnitude. However, this issue is mitigated in our study, as prices remain within the same order of magnitude throughout the sample period.

After applying the VIF-based elimination, we are left with approximately 15 variables for each cryptocurrency. These variables are then subjected to a second selection step using BMA. This approach allows us to account for model uncertainty and identify the most robust explanatory variables for our final regression models.

3.3. Bayesian model averaging

BMA is a statistical technique designed to handle model uncertainty by averaging across a set of possible models, with each model weighted according to its posterior probability and complexity (Moral-Benito, 2015; Raftery et al., 1997). When dealing with a large number of potential explanatory variables, including all of them in a single model can lead to incorrect inferences or inefficient estimation because some variables may be superfluous or highly correlated. BMA provides a principled approach to account for this uncertainty by considering multiple models simultaneously.

More specifically, following the approach detailed by Moral-Benito (2015), suppose we have H potential models under consideration, which are indexed by $h = 1, \dots, H$. Each of these models yields a different estimate of the model coefficients, denoted as $\hat{\beta}_1, \dots, \hat{\beta}_H$. Rather than arbitrarily selecting one model from this potentially vast set, BMA accounts for all H candidate models. The model-averaged estimate, denoted by $\hat{\beta}_{MA}$, is computed as a weighted sum of the estimates from all considered models:

$$\hat{\beta}_{MA} = \sum_{h=1}^H \omega_h \hat{\beta}_h, \quad (4)$$

where the weight assigned to each model, ω_h , depends on a predetermined weighting scheme. This allows BMA to balance the contributions of all the models rather than selecting a single best-fitting model. To illustrate this process in more concrete terms, consider the following regression framework:

$$\begin{aligned} y &= X\beta + \epsilon, \\ \epsilon &\sim N(0, \sigma^2 I_T), \end{aligned} \quad (5)$$

where $y = (y_1, \dots, y_T)'$ is an $T \times 1$ vector representing the dependent variable with random disturbances ϵ . The matrix X is an $T \times q$ matrix of regressors, where each regressor is a candidate to be included in the final model. The vector β , of size $q \times 1$, contains the parameters to be estimated. Since there are q possible regressors, the total number of possible models to consider is 2^q , with each model denoted by M_j , where $j \in \{1, \dots, 2^q\}$.

Each model M_j is characterized by a unique set of parameters, β^j . Within the Bayesian framework, the posterior distribution of the parameters, given that model M_j is selected, is defined as follows using Bayesian principles:

$$g(\beta^j | y, M_j) = \frac{f(y | \beta^j, M_j) g(\beta^j | M_j)}{f(y | M_j)}. \quad (6)$$

In this equation, the posterior distribution $g(\beta^j | y, M_j)$ is a product of the likelihood $f(y | \beta^j, M_j)$ and the prior $g(\beta^j | M_j)$ for the model M_j . Given a prior probability assigned to each model, $P(M_j)$, the posterior probability of a model, conditional on the data y , can be computed using Bayes' theorem as follows:

$$P(M_j | y) = \frac{f(y | M_j) P(M_j)}{f(y)}. \quad (7)$$

Eqs. (6) and (7) highlight the necessity of specifying priors for both the model parameters and the probability of the model itself.

Next, to compute the marginal likelihood, the following integral is used:

$$f(y | M_j) = \int f(y | \beta^j, M_j) g(\beta^j | M_j) d\beta^j. \quad (8)$$

By taking the ratio of marginal likelihoods from Eq. (8) for two models, we can compute the Bayes factor. In contrast to the likelihood ratio test, which considers only point-specific estimates, the Bayes factor incorporates the entire parameter space, thereby reflecting uncertainty in the estimates, and includes a penalty for considering overly complicated models. This makes it a more robust tool against overfitting.

Following the framework proposed by Moral-Benito (2015) and Leamer (1978), we can view β as a function of β^j for each model, $j = 1, \dots, 2^q$. Thus, the posterior distribution of β , denoted $g(\beta | y)$, is computed as a weighted average of the posterior distributions for β^j across all models. The weights correspond to the posterior probabilities $P(M_j | y)$:

$$g(\beta | y) = \sum_{j=1}^{2^q} P(M_j | y) g(\beta | y, M_j). \quad (9)$$

Hence, the full posterior distribution for β represents a combination of the distributions across all possible models, weighted by the probability of each model. The process of BMA allows for inference and estimation directly from this posterior distribution, thereby accounting for model uncertainty in a principled manner. In the context of the BMA framework, it is also possible to compute the posterior inclusion probability (PIP) for any given variable. The PIP represents the probability that a specific variable is included in the model and is calculated by summing the posterior probabilities of all the models that incorporate that variable.

The PIP is the last criterion we consider when including a variable in our model. After the VIF procedure, we use variables whose PIP is greater than 90%, thus appearing in more than 90% of the tested models. Since we obtained approximately 15 variables from the VIF, it is possible to enumerate all model possibilities in a reasonable amount of time. These variables are reported as the models in Section 5. Finally, we use a uniform model prior ($P(M_j)$) and the 'Unit Information Prior' (UIP) for the coefficients, $g(\beta) = N$, where N is the number of observations. The UIP prior is well balanced; it adds some shrinkage while still permitting the data to convey significant information. Finally, we use BMA implementation in package BMS by Feldkircher and Zeugner (2015).

4. Data and setup

We perform the analysis of trades to obtain wash trade metrics with a trade-by-trade dataset from the data provider Tardis.dev, and we provide a short snapshot of the data in Table A.4. Our sample period runs from November 15, 2020, to January 31, 2022, covering 443 trading days with a variable number of realized trades across different exchanges and assets. We focus on four major assets: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), and XRP (formerly Ripple), which consistently represent a significant portion of the total cryptocurrency market capitalization. These assets are available on all selected exchanges: Binance Futures, Bitfinex Derivatives, FTX, OKEX Swap, and Huobi Futures. The specific data series we obtained are listed in Table 1, which also maps tickers to their respective assets.

4.1. Comprehensive market coverage

To capture a comprehensive view of market activity and potential wash trading across cryptocurrency markets, our dataset comprises perpetual swaps and futures contracts, which dominate cryptocurrency trading in terms of both volume and price discovery. Exchanges on platforms such as CoinMarketCap.com are ranked by these same metrics for both derivatives and spot markets, highlighting the importance of derivatives for overall market behavior.

Perpetual futures, commonly referred to as perpetual swaps, are the most traded derivatives in cryptocurrency markets, with over \$100 billion in daily trading volume (He et al., 2022). These contracts are

Table 1
Tickers on [Tardis.dev](https://tardis.dev).

Asset	Exchange				
	Binance futures	Bitfinex derivatives	FTX	Okex swap	Huobi futures
BTC	BTCUSDT	BTCF0-USTF0	BTC-PERP	BTC-USD-SWAP	BTC_CQ
ETH	ETHUSDT	ETHF0-USTF0	ETH-PERP	ETH-USDT-SWAP	ETH_CQ
LTC	LTCUSDT	LTCF0-USTF0	LTC-PERP	LTC-USD-SWAP	LTC_CQ
XRP	XRPUSDT	XRPF0-USTF0	XRP-PERP	XRP-USDT-SWAP	XRP_CQ

The table reports the exchange-specific tickers obtained from Tardis.dev, along with their corresponding mappings to the respective cryptocurrency assets. All tickers refer to futures or perpetual swap contracts and are available over the full sample period from November 15, 2020, to January 31, 2022.

also readily available to retail investors. Unlike traditional futures contracts, perpetual swaps do not have expiration dates, making them a favored instrument for leveraged trading without the need to hold the underlying asset. Owing to their speculative nature and large volumes, perpetual swaps account for a significant share of market activity and are crucial for identifying potential wash trading patterns. Including these instruments ensures that a substantial portion of speculative trades, which may be susceptible to manipulation, are represented in our analysis (Alexander & Heck, 2020). In the case of Ethereum, perpetual swaps also demonstrate dominant trading volumes compared with spot markets, particularly on exchanges such as BitMEX (Alexander et al., 2020). This dominance of perpetual swaps motivates our focus on these instruments, which are the key venue for institutional and speculative trading. Given their large volumes and influence, it is highly plausible that wash trading may occur at scale within these markets. The inclusion of futures and perpetual swaps in our dataset provides a more holistic view of trading activities, enabling us to better detect wash trading across various market segments. By focusing on the derivatives markets that dominate trading volume, our study captures the market segment most exposed to manipulative trading behaviors.

4.2. Explanatory variables

This analysis aims to explain the dynamics of wash trading metrics in the four major cryptocurrencies. We search for drivers within crypto markets as well as external factors that influence how exchanges engage in wash trading, acknowledging that cryptocurrencies have become increasingly connected to traditional financial markets and reflect macroeconomic and monetary indicators (Kočenda & Albrecht, 2026; Kukacka & Kristoufek, 2023; Sila et al., 2024).

To provide a clear understanding of the explanatory variables, we organize them into 11 categories, such as blockchain metrics, macroeconomic indicators, commodity prices and financial markets, trading activity, and others, summarized in Table B.5 in Appendix A. The respective descriptive statistics are presented in Table C.6 in Appendix A.

In summary, our dataset encompasses a rich array of variables that capture the multifaceted nature of cryptocurrency markets. By combining a wide range of explanatory variables, including blockchain metrics, financial indicators, macroeconomic factors, behavioral measures (as proxies for public sentiment), and trading activity, we establish a robust foundation for analyzing wash trading dynamics. All variables are carefully processed and, where appropriate, logarithmically transformed to enhance numerical stability, as indicated by the ‘*l*’ suffix in the descriptive statistics in Table C.6. The breadth and depth of our data enable a thorough investigation into the determinants of wash trading in major cryptocurrencies.

5. Explaining wash trading

The estimated wash trading metrics depicted in Fig. 1 reveal significant differences in wash trading behavior across Bitcoin, Ethereum, Litecoin, and XRP from late 2020 to early 2022. Overall, the percentage of wash trading volume estimated using the volume roundness-based

metric (left panel) mostly fluctuates between 55% and 85%, strongly supporting the main finding of Cong et al. (2023) that over 70% of the volume reported by unregulated exchanges is wash trading. The Benford-law-based $MAD_{Benford}$ metric (right panel) generally moves between 0.5 and 3 percentage points, representing a sizable average absolute deviation for each of the nine digits, as discussed in Section 2.2.

According to the wash trading volume roundness-based metric (left panel), Ethereum presents the highest levels of wash trading, followed by XRP and Bitcoin, with Litecoin showing the lowest levels. Ethereum’s peaks in wash trading align with major price surges, indicating attempts to inflate wash trading volume and enhance perceived liquidity, a pattern also observed by Cong et al. (2021) in their study on decentralized exchanges during price increases. XRP shows significant wash trading, particularly during its legal battle with the U.S. SEC, which is likely aimed at countering negative market sentiment. This finding is consistent with that of Pennec et al. (2021), who observed increased wash trading during periods of regulatory scrutiny. Bitcoin, while still showing notable wash trading, ranks lower than both Ethereum and XRP in terms of wash trading volume, aligning with the findings of Foley et al. (2019), who documented substantial but lower levels of wash trading in Bitcoin than in altcoins. Litecoin, as expected, experiences the least amount of wash trading, reflecting its reputation as a more stable asset with less speculative activity, as also observed by Makarov and Schoar (2020). This stability likely results in fewer attempts to artificially boost trading volumes.

When these findings are compared with the deviations from Benford’s law (right panel), the patterns of manipulation differ. Bitcoin shows the greatest deviations, suggesting persistent statistical anomalies and market manipulation that subtly distort trading patterns, a finding that is consistent with that of Osterrieder and Lorenz (2017), who documented similar deviations in Bitcoin’s trading behavior. Ethereum and XRP show similar levels of deviations, indicating significant manipulation, but less pronounced deviation compared with Bitcoin. Litecoin, once again, has the lowest deviation from Benford’s law, indicating the least amount of disruption to its natural trading patterns. This finding aligns with that of Vergne (2020), who reported that Benford’s law is less effective at detecting manipulation in smaller, more stable cryptocurrencies.

Overall, the wash trading volume and Benford’s law-based metrics reveal distinct facets of wash trading. Ethereum and XRP exhibit the highest wash trading volume, with Bitcoin showing moderate levels, whereas deviations from Benford’s law highlight Bitcoin as the asset with the most detectable manipulation, followed by Ethereum and XRP. Litecoin remains the least manipulated asset by both measures. We now explore the drivers and indicators of the dynamics of each metric.

5.1. Trade volume roundness-based metric

The results for the trade volume roundness-based metric reported in Table 2 reveal significant associations between wash trading and various factors, both crypto-related and external. We emphasize that the analysis is associational and predictive rather than causal. Daily observational data and BMA-based variable selection identify robust

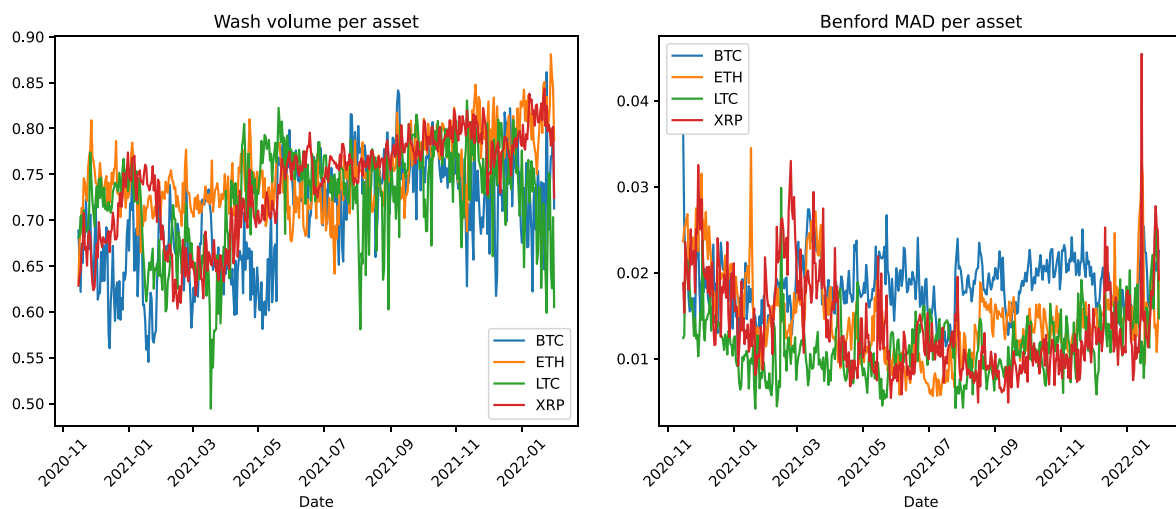


Fig. 1. Wash trading metrics. The left panel depicts the wash trading volume as a percentage of the overall daily trading volume on the basis of the volume roundness-based metric for the four crypto assets under study. The right panel depicts the daily Benford's law-based $MAD_{Benford}$ metric.

statistical relationships and indicators of future wash trading, but they do not isolate exogenous variation and therefore do not establish causal effects. Terms such as 'driver,' 'effect,' or 'leads to,' where they appear, should be read in this associational and predictive sense. In the table, we summarize the final models of the BMA procedure, with adjusted R^2 values as an indicative measure of the model's quality. The first column of numerical results shows the estimation for the immediate effect. The next three columns display the estimates for models with the explanatory variables lagged by 1, 7, and 28 days, respectively, indicating whether the explanatory variables serve as indicators of future wash trading, either positive or negative. The specification deliberately omits an autoregressive term because our objective is to identify the factors associated with wash trading and their value as predictive indicators, not to model the serial dependence of the dependent variable itself. As is common for daily financial data, the Breusch–Godfrey test rejects the null of no serial correlation in all specifications, indicating remaining residual autocorrelation, and the Breusch–Pagan test indicates heteroskedasticity in several models. Neither biases the OLS point estimators, but both invalidate conventional standard errors. We therefore base all inference on heteroskedasticity- and autocorrelation-robust (HAC) standard errors (Stock & Watson, 2003), which remain valid under both. The augmented Dickey–Fuller (ADF) and KPSS tests provide strong statistical evidence for stationary residuals, indicating that the reported relationships are not spurious.

5.1.1. Bitcoin

For BTC, the transaction transfer count ($Transaction\#$) has a significant negative relationship with wash trading volume in both the immediate effect and at lag 1. This suggests that higher on-chain transaction activity corresponds to reduced wash trading, possibly because increased legitimate market participation dilutes these manipulative practices. This negative relationship between transaction activity and wash trading aligns with earlier findings by Foley et al. (2019), who argued that legitimate participation and liquidity can reduce the necessity for wash trading in more liquid markets such as Bitcoin. However, further research is needed to explore the specific mechanisms through which legitimate activity crowds out wash trading. Order book analysis or real-time monitoring of transaction flows could provide deeper insights into how legitimate market engagement diminishes wash trading. The volatility of BTC (σ_{BTC}) is positively associated with wash trading volume in the immediate term and at lag 1, indicating that higher price volatility coincides with more wash trading. Exchanges may exploit volatile markets to engage in wash trading, creating an

illusion of liquidity or influencing prices. At lag 1, the momentum in BTC prices ($Momentum_{BTC}$) becomes significant and positive, implying that recent positive price trends may predict increased wash trading activity, but only in a short-lived manner, as the momentum is not significant for higher lags. Economic policy uncertainty (EPU) consistently has a negative relationship across all the models, indicating that greater economic uncertainty may deter wash trading in BTC, which is likely connected to the risk aversion of investors, who become less active in times of economic uncertainty, and that exchanges are less motivated to wash trade.

5.1.2. Ethereum

For ETH, the velocity ($Velocity_{ETH}$) consistently has a positive and significant effect on wash trading volume across all lag structures. This finding indicates that greater circulation of ETH is associated with increased wash trading, suggesting that as ETH is transacted more frequently, opportunities or incentives for wash trading rise. These results are consistent with the work of Cong et al. (2021), who identified a positive relationship between transaction volumes and wash trading on decentralized exchanges. This connection thus seems to hold for both centralized and decentralized exchanges. Immediate volatility (σ_{ETH}) positively affects wash trading volume, reinforcing the notion that price volatility encourages wash trading. Notably, Google Trends data for BTC ($GoogleTrends_{BTC}$) negatively impact ETH wash trading volume at lag 7, implying that increased public sentiment toward BTC may divert attention from ETH, leading to reduced wash trading in ETH. The effect of the VIX is statistically significant across different lags, yet its economic scale is small.

5.1.3. Litecoin

In the case of LTC, the VIX index (VIX) consistently shows a significant negative relationship with wash trading volume across all lag structures. This finding suggests that greater uncertainty, as measured by the VIX, is associated with reduced wash trading in LTC. This negative relationship between uncertainty and wash trading is unique to Litecoin in our analysis, contrasting with the findings for other cryptocurrencies. However, this finding is consistent with that of Makarov and Schoar (2020), who noted Litecoin's reduced speculative activity. Social interest indicators such as Google Trends for LTC and BTC negatively affect wash trading volume, suggesting that increased public sentiment corresponds to reduced wash trading. This may be attributed to heightened scrutiny and more informed market participants deterring wash trading, but is more likely due to more attention being

Table 2
Estimated models for the wash trading volume roundness-based metric.

	BTC			
		lag 1	lag 7	lag 28
Transaction#_1	-0.146***	-0.175***		
σ_{BTC}	0.436***	0.332***		
Momentum_BTC		0.137***		
VIX_1				-0.091***
EPU_1	-0.027***	-0.031***	-0.032***	-0.029***
Volume_Alts_1			-0.021***	-0.016***
Adjusted R ²	0.230	0.288	0.160	0.244
BP test	5.59	9.75**	8.87**	7.36*
BG test	177.8***	150.0***	189.1***	182.1***
ADF test	-4.93**	-4.99**	-4.31**	-4.06**
KPSS test	2.71	2.82	2.25	1.42
ETH				
Velocity_ETH_1	0.042***	0.032***	0.031***	0.039***
σ_{ETH}	0.322***			
GoogleTrends_BTC_1			-0.021***	
VIX_1	0.035***	0.041***	0.036***	
Momentum_BTC				-0.106***
Adjusted R ²	0.215	0.177	0.282	0.290
LTC				
VIX_1	-0.120***	-0.122***	-0.110***	-0.113***
GoogleTrends_LTC_1	-0.053***	-0.045***	-0.037***	-0.038***
Wiki_BTC_1				-0.026***
Wiki_LTC_1	0.029***	0.031***	0.025***	0.020***
σ_{LTC}	0.494***	0.418***	0.353***	
GoogleTrends_BTC_1	-0.031***	-0.040***	-0.051***	
Connectedness	0.001***	0.001***	0.001***	0.001***
Gold_1	0.269***	0.248***		
Adjusted R ²	0.440	0.429	0.367	0.288
XRP				
GoogleTrends_Crypto_1	0.041***	0.040***	0.028***	0.034***
Gold_1	0.705***	0.687***	0.478***	
VIX_1			-0.065***	-0.061***
σ_{XRP}	-0.204***	-0.224***	-0.148***	
GoogleTrends_BTC_1	-0.043***	-0.043***	-0.072***	-0.079***
Wiki_BTC_1	-0.037***	-0.037***		-0.036***
EPU_1	-0.021***	-0.023***	-0.017***	-0.014***
σ_{BTC}	0.563***	0.595***	0.519***	0.451***
Connectedness	-0.001***	-0.001***	-0.0005***	
Adjusted R ²	0.438	0.449	0.454	0.604

Based on a daily dataset of over 40 explanatory variables covering the period from November 15, 2020, to January 31, 2022, i.e., 443 observations. First, explanatory variables with the overfitting risk ($VIF > 10$) are eliminated step-by-step. The final model results from the subsequent BMA model selection under $PIP > 0.9$. The HAC standard errors, according to Stock and Watson (2003), are considered. For the tests, ***, **, and * mark statistical significance at the 99%, 95%, 90% confidence level, respectively. The tests for the ETH, LTC, and XRP model residuals produced qualitatively similar results to those for BTC and are therefore omitted for brevity.

directed toward the Bitcoin markets, where the increased wash trading activity siphons it from the other markets. Consistent with BTC and ETH, LTC volatility (σ_{LTC}) is positively associated with, and predictive of, increased wash trading in LTC. The total connectedness of the market (*Connectedness*) is significant across the forecasting horizons, but the effect is practically tiny.

One of the more notable and consistent findings in our results is the negative relationship between the VIX index, a proxy of global financial uncertainty, and wash trading activity in Litecoin. This contrasts with the typical response in BTC and ETH, where volatility often coincides with increased wash trading. We interpret this pattern as a potential instance of endogenous market discipline in a lower-liquidity environment. Specifically, during times of elevated global risk aversion, retail and speculative demand in mid-cap cryptocurrencies such as LTC tends to decline, reducing the marginal benefits of inflating volume for exchanges or market makers. Moreover, the cost of manipulation, measured in terms of capital, slippage, or regulatory risk, may become

less justifiable when market attention and turnover fall. This could naturally suppress wash trading without the need for external oversight. Such a dynamic suggests that not all crypto assets are equally susceptible to wash trading at all times and that exchange-level wash trading strategies may be asset-specific and conditioned on macro-level volatility. For researchers and regulators alike, this finding implies that broad-brush enforcement may be inefficient. Some assets or platforms may already exhibit a degree of self-stabilization, especially during global risk-off episodes. Future work could explore whether similar behavior occurs in other mid-cap or legacy cryptocurrencies with long-standing infrastructure but low speculative turnover and whether the fee structure or liquidity fragmentation mediates this effect.

5.1.4. XPR

For XRP, the price of gold (*Gold*) has a positive influence on wash trading volume in the immediate term and at short-term lags, with a significant effect. This suggests that when gold prices rise, wash trading in XRP increases, possibly reflecting a shift in investment strategies or hedging behaviors that create opportunities for wash trading. The VIX index negatively affects wash trading volume at longer lags (lags 7 and 28), indicating that higher market uncertainty reduces wash trading in XRP over time. Additionally, increased public sentiment, as proxied by Google Trends for BTC (*GoogleTrends_BTC*) and Wikipedia searches for BTC (*Wiki_BTC*), negatively impacts XRP wash trading volume, highlighting cross-asset influences and the role of information dissemination in deterring wash trading as well as the dominance of BTC in the crypto markets.

5.1.5. Common patterns

Across the different cryptocurrencies analyzed, several common patterns emerge in the factors influencing wash trading volume. Market volatility appears to be a robust correlate and predictor of wash trading, with higher volatility associated with higher wash trading volume in BTC, ETH, and LTC. This suggests that exchanges may exploit volatile conditions to engage in wash trading across multiple assets. The only exception is XRP, in which wash trading activity is boosted by BTC volatility. As BTC and XRP volatilities are correlated, the negative estimated effect of its own volatility in XRP does not imply that it actually decreases wash trading.

Another common pattern is the negative impact of public sentiment on wash trading outside of BTC, suggesting that the public sentiment effect is likely dominated by interest in BTC, which indirectly decreases trading in other assets as wash trading activity gravitates toward the Bitcoin market.

From the aggregate perspective and on the basis of the adjusted R^2 values, we find that wash trading dynamics can be better explained, tracked, and anticipated for smaller cryptocurrencies, which are represented by LTC and XRP in our sample. We also observe that the quality of the models does not decrease much for further foresight. We find that the adjusted R^2 values decrease with lags only for LTC. The information and anticipation move across explanatory variables, directing toward the importance of a large set of explanatory variables to understand the complex interactions properly.

These empirical regularities point to the system-wide influence of Bitcoin on altcoin behavior, which invites further modeling of its transmission effects and wash trading spillovers. Our modeling framework explicitly incorporates Bitcoin-specific variables (volatility, momentum, public sentiment metrics, and network usage) as covariates in the altcoin models. The statistical significance and explanatory power of these variables confirm the structural dominance of Bitcoin in shaping market-wide trading behavior and wash trading incentives. However, this influence is not merely contemporaneous; it may also propagate through dynamic spillover channels. For example, during periods of strong upward momentum in BTC markets, altcoin exchanges may seek to capitalize on investor enthusiasm by artificially inflating volume in assets with perceived Bitcoin correlations. These wash

Table 3
Estimated models for Benford's law-based $MAD_{Benford}$ metric.

	BTC			
		lag 1	lag 7	lag 28
Connectedness				−0.00004***
Wiki_BTC_1				0.003***
Gold_1		−0.016***	−0.019***	−0.023***
GoogleTrends_BTC_1				−0.002***
Adjusted R ²	–	0.016	0.024	0.106
BP test		0.63	0.08	5.11
BG test		247.9***	245.5***	213.9***
ADF test		−4.42**	−4.40**	−5.23**
KPSS test		0.17	0.18	0.12
ETH				
VIX_1	0.007***	0.008***	0.012***	0.007***
EPU_1	0.002***	0.002***		
Addresses_ETH_1	−0.008***	−0.007***		−0.009***
Momentum_BTC			0.016***	
Adjusted R ²	0.222	0.204	0.215	0.166
LTC				
GoogleTrends_BTC_1	−0.004***			
InterestRate			0.001***	
Adjusted R ²	0.140	–	0.228	–
XPR				
VIX_1	0.012***	0.012***	0.016***	0.012***
EPU_1	0.003***			
GoogleTrends_Crypto_1		−0.002***		
σ_{BTC}			−0.032***	
Gold_1				0.040***
Adjusted R ²	0.235	0.216	0.227	0.205

Based on a daily dataset of over 40 explanatory variables covering the period from November 15, 2020, to January 31, 2022, i.e., 443 observations. First, explanatory variables with the overfitting risk ($VIF > 10$) are eliminated step-by-step. The final model results from the subsequent BMA model selection under $PIP > 0.9$. The HAC standard errors, according to [Stock and Watson \(2003\)](#), are considered. For the tests, ***, **, * mark statistical significance at the 99%, 95%, 90% confidence level, respectively. Tests for ETH, LTC, and XPR model residuals produced qualitatively similar results to those for BTC and are therefore omitted for brevity.

trading cascades are likely amplified in less-regulated trading environments, where cross-asset trading pairs and common infrastructure (e.g., stablecoin-based liquidity pools and shared wallets) further blur asset boundaries.

5.2. Benford's law-based metric

Moving to the deviations from Benford's law reported in [Table 3](#), the results are less clear. Note that the deviations from the law do not capture the volume of wash trading but rather whether wash trading is taking place in the analyzed market.

For BTC and LTC, we have difficulty detecting any reasonable drivers or indicators for the metrics. Although we find four statistically significant results for BTC at lag 28, the Wikipedia views and Google Trends mostly balance each other, and the other two are hardly economically relevant in magnitude. While we find limited evidence of drivers for deviations from Benford's law in BTC and LTC, the persistent statistical anomalies detected in Bitcoin align with the findings of [Osterrieder and Lorenz \(2017\)](#), who observed similar deviations indicative of market manipulation. For ETH and XRP, we find that VIX and EPU increase the likelihood of wash trading on the exchanges, which aligns somewhat with the results for wash trading volume in the previous section. The relationship between market volatility (VIX) and deviations from Benford's law in ETH and XRP aligns with the argument of [Vergne \(2020\)](#) that Benford's law may be more effective for detecting manipulative practices in highly speculative or volatile markets. Nevertheless, the Benford's law-based metric is rather narrow and challenging to model or predict.

5.3. Discussion

The analysis reveals that both the trade volume roundness-based metric and, to a lesser extent, the Benford's law-based metric respond to market volatility and uncertainty but capture different aspects of manipulative practices. The volume roundness-based metric is more sensitive to immediate market conditions and behavioral factors such as price volatility, momentum, and public sentiment. It highlights how exchanges and possibly traders might use wash trades to influence perceived liquidity and price movements in the short term. For example, in BTC and ETH, higher volatility is associated with higher wash trading volume, consistent with exchanges exploiting volatile conditions.

The Benford's law-based metric, on the other hand, captures systemic irregularities in trade data over time, revealing wash trading that deviates from expected numerical patterns. The consistent positive relationship between the VIX index and the metric for ETH and XRP suggests that it is more likely that exchanges engage in more wash trading during periods of market stress, potentially because of decreased oversight or attempts to capitalize on market volatility.

The two metrics are interconnected in that they both respond to market volatility and uncertainty but operate on different temporal and behavioral scales. While increased volatility may lead to immediate wash trading, as captured by the trade volume roundness-based metric, sustained periods of uncertainty might result in more systemic wash trading detectable by deviations from Benford's law. The metrics are designed to capture different temporal and structural aspects of wash trading. The volume roundness-based metric identifies artificial volume inflation by exploiting behavioral features of wash trading: round-number trading, mirror trades, and excessive intraday frequency in specific volume sizes. It is therefore highly sensitive to short-term wash trading bursts that typically align with market-moving news, pump-and-dump activity, or fee manipulation schemes. In contrast, the Benford's law-based approach reflects statistical anomalies in the distribution of leading digits of reported trade volumes, which accumulate over longer periods and are less prone to noise from temporary market fluctuations. Additionally, it is not volume-sensitive. While both metrics serve as measures to detect wash trading, they are not expected to be tightly correlated. Their divergence in explanatory power, particularly the weaker and more inconsistent signal from Benford's law, reflects both its strength as a cumulative indicator and its weakness in capturing tactical or microstructural wash trading. Moreover, Benford distributions can be affected by legitimate rounding behavior, token unit characteristics, or exchange-specific reporting formats. Taken together, the dual-metric strategy provides a complementary lens: the volume roundness-based metric is ideal for detecting event-driven or high-frequency wash trading episodes, whereas Benford's law serves as a lower-resolution tool for detecting long-run irregularities and is potentially used as a systemic check of persistent irregularities in exchange-reported data.

Cross-asset effects are evident in the analysis, highlighting the interconnected nature of the cryptocurrency market, a phenomenon similarly noted by [Foley et al. \(2019\)](#), who found that manipulation of major assets such as Bitcoin can influence the behavior of other cryptocurrencies. Variables related to BTC, such as Google Trends and Wikipedia searches, significantly impact the wash trading and manipulation metrics of the other studied cryptocurrencies. Events or trends affecting the major cryptocurrency can thus influence trading behaviors and wash trading across the market.

5.3.1. Regulatory developments

Although our empirical focus is on the structural and behavioral determinants of wash trading, the institutional context in which these dynamics occur deserves closer attention, especially in light of recent regulatory developments. Our analysis covers a period (November 2020 to January 2022) that precedes the formal implementation of major regulatory frameworks, most notably the EU Markets in Crypto-Assets

Regulation (MiCAR), the SEC's intensified enforcement efforts after 2022, and global coordination efforts led by the International Organization of Securities Commissions (IOSCO). While public discussions around the need for stronger regulation of centralized exchanges were already underway, triggered by repeated reports of suspicious activity and several historical exchange collapses, regulatory oversight during our sample was largely absent, fragmented, or unenforced. The persistence of statistically and economically significant wash trading across large and reputable exchanges in our results illustrates the limited deterrent effect of regulatory discourse alone, especially in jurisdictions lacking binding enforcement mechanisms or supervisory capacity. This raises important implications for the design and implementation of regulatory regimes.

First, regulation must be accompanied by robust auditing, surveillance technology, and cross-border information sharing, especially given the pseudonymous and globally mobile nature of crypto asset exchanges. Second, the incentives for exchanges to self-regulate in the absence of strict rules remain weak, particularly where high reported trading volumes translate into higher rankings on aggregator platforms, better token listings, and increased fee revenue. The post-2022 enforcement environment, including actions taken against Binance, Coinbase, and others, offers a natural follow-up period to assess whether the formal regulatory presence changes these behavioral patterns or simply shifts wash trading to less visible markets. Our results thus serve as an empirical benchmark against which future assessments of regulatory efficacy can be conducted.

5.4. Avenues for future research

Understanding and possibly foreseeing wash trading activity in crypto markets is essential for greater acceptance and further market penetration of emerging cryptocurrencies. To do so, we need reliable metrics and methodologies to detect and indicate such suspicious behavior. The development and incorporation of additional metrics for wash trading detection could validate and complement the current findings. Alternative approaches, such as order book analysis, trade clustering methods, or machine learning-based anomaly detection, could uncover wash trading that is not captured by volume roundness-based or, more importantly, Benford's law-based metrics, as the latter have been shown to be of limited informative value.

However, it is not only the metrics that can be explored. Examining the impact of regulatory changes or announcements on wash trading activities would provide valuable insights into the effectiveness of policy interventions in the cryptocurrency market. Analyzing how exchanges and traders adjust their behaviors in response to regulatory scrutiny could inform future regulatory strategies aimed at reducing manipulative practices. A natural next step is to extend the dataset temporally to cover the regulatory environment after 2022. The period from 2022 onward represents a fundamental turning point in the institutional treatment of cryptocurrency markets. Regulatory milestones include the EU's final adoption of MiCAR, which mandates licensing and transparency standards for crypto asset service providers (CASPs), the SEC's legal pursuit of major exchanges for unregistered securities trading, the Commodity Futures Trading Commission's expanded jurisdictional claims, and coordinated guidance by IOSCO proposing global market integrity standards. These measures are no longer just declaratory but are beginning to be enforced, with tangible legal consequences. Extending the analysis to this postregulatory period would enable a rigorous evaluation of whether these interventions have reduced wash trading behavior or merely displaced it to less visible or less regulated venues. In addition, asset-level heterogeneity, such as differences in DeFi integration, tokenomics, or investor base, may moderate the regulatory response, providing further empirical variation.

In addition to model extensions and dataset expansions, future research may benefit significantly from the application of machine learning-based anomaly detection techniques and clustering methods.

Unsupervised learning algorithms such as isolation forests, *k*-means, or DBSCAN (density-based spatial clustering of applications with noise) can be employed to detect anomalous trading patterns that deviate from historically typical behavior without the need for labeled data or manually defined rules. Furthermore, feature engineering based on wallet addresses, asset pairs, fee structures, and exchange-specific trade timestamps could enhance these models' precision and interpretability. Integrating these approaches with traditional metrics, such as those proposed in our current study, may also yield hybrid detection frameworks capable of capturing both real-time and structural wash trading behaviors. Other models, such as autoencoders or long short-term memory (LSTM) networks, can learn from historical data with a strong emphasis on diurnal patterns to identify subtle and nonlinear relationships that may evolve over time or vary by asset. In particular, these models hold potential for uncovering coordinated wash trading schemes at certain time periods, round-trip volume bursts, mirrored buy-sell sequences, and generally coordinated patterns across assets and/or exchanges. Their ability to generalize across different market regimes and adapt to shifting trading norms makes them attractive tools for scalable, real-time surveillance in cryptocurrency markets.

In addition to methodological innovation, comparative analysis between cryptocurrency and traditional financial markets represents another promising research direction. Wash trading, while widely associated with crypto assets, is not unique to them. Similar behaviors have been documented in emerging market equities, over-the-counter FX markets, and even regulated venues under relaxed supervision, such as small-cap stocks in the U.S. prior to regulatory crackdowns. Conducting empirical comparisons across asset classes and regulatory environments may help distinguish crypto-specific wash trading dynamics from more general market failures. Moreover, such cross-market studies could inform best practices in surveillance system design, help calibrate risk thresholds for different asset types, and guide policy harmonization across jurisdictions. Understanding whether wash trading in crypto follows a different logic or merely reflects underenforced market norms remains a fundamental question for market integrity research.

6. Conclusion

This study investigates wash trading behavior in major cryptocurrencies, Bitcoin, Ethereum, Litecoin, and XRP, using two distinct metrics: the trade volume roundness-based metric and the Benford's law-based metric. The analysis reveals significant differences in wash trading practices across these assets and underscores the complexity in detecting and understanding wash trading in cryptocurrency markets.

The trade volume roundness-based metric indicates that, owing to their market dominance, Bitcoin and Ethereum experience consistently high levels of wash trading, especially during periods of rapid price growth. This suggests that exchanges and possibly traders engage in wash trading to inflate perceived liquidity and drive market optimism during bullish phases. Litecoin, characterized by a steadier market environment, exhibits lower levels of wash trading, reflecting less speculative interest and less effort to artificially amplify trading volumes. XRP displays relatively high levels of wash trading, particularly during its legal challenges with the SEC, likely aiming to counteract negative market sentiment and create the illusion of liquidity amidst regulatory uncertainty.

The deviations from Benford's law provide a different perspective, as they capture systemic irregularities in trade data over time. While this metric is less sensitive to immediate manipulative practices, it highlights periods of sustained wash trading detectable through statistical anomalies in numerical patterns. The results suggest that while wash trading is prevalent across all crypto assets, the degree to which it distorts natural trading patterns varies, with Bitcoin and Ethereum exhibiting more sustained and widespread wash trading detectable by both metrics.

The analysis reveals that both metrics respond to market volatility and uncertainty but capture different aspects of wash trading. Market volatility emerges as a robust predictor of wash trading, with higher volatility associated with higher wash trading volume in Bitcoin, Ethereum, and Litecoin. Exchanges may exploit volatile conditions to engage in wash trading across multiple assets. Public sentiment negatively impacts wash trading, particularly in assets other than Bitcoin, stressing its dominance in the crypto markets as it gravitates the wash trading activities during bull runs. The findings highlight the interconnected nature of cryptocurrency markets, where variables related to Bitcoin significantly impact wash trading and manipulation metrics of other cryptocurrencies. This cross-asset influence suggests that wash trading may propagate systemically through spillover channels, especially in times of high market activity.

An additional finding of interest is the inverse relationship between global market volatility, as captured by the VIX index, and wash trading in Litecoin. This may indicate that certain assets exhibit endogenous stabilizing dynamics, where reduced retail activity during global risk-off episodes naturally curbs wash trading, in contrast to more speculative environments. Such asset-specific patterns imply that wash trading incentives may vary not only across exchanges and time but also by structural characteristics of the underlying asset.

This study emphasizes the need for reliable metrics and methodologies to detect and indicate suspicious trading behavior. While the trade volume roundness-based metric effectively captures immediate wash trading, the Benford's law-based metric has limited informative value because of its indirect nature and sensitivity to significant disruptions in natural trading patterns. The development of additional metrics and the use of advanced analytical techniques, such as machine learning-based anomaly detection, could enhance the detection of wash trading not captured by current methods.

Importantly, the study captures a largely preregulatory period, as major enforcement regimes such as the EU's MiCAR and increased SEC scrutiny were introduced only after our sample ended in early 2022. This makes our results a valuable empirical baseline for future assessments of regulatory effectiveness.

In conclusion, understanding and addressing wash trading in cryptocurrency markets is crucial for promoting market integrity and fostering broader acceptance of digital assets. Regulators, exchanges, and market participants must collaborate to develop effective surveillance mechanisms and regulatory frameworks that can adapt to the evolving landscape of cryptocurrency trading. Our findings reinforce the need for targeted enforcement, dynamic market monitoring, and an improved understanding of cross-asset wash trading dynamics. They also underscore the need to consider wash trading not only as a technical violation but also as a deeper market distortion that manipulates perceived liquidity and exploits asymmetric information. Future research should continue to explore alternative detection methods, examine the impact of regulatory interventions, and compare wash trading across different markets to enhance our understanding of wash trading and its implications for the financial ecosystem.

CRediT authorship contribution statement

Jan Sila: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Evzen Kocenda:** Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. **Ladislav Kristoufek:** Conceptualization, Formal Analysis, Funding acquisition, Investigation, Resources, Writing – original draft. **Jiri Kukacka:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used the AI Research Assistant by the [Web of Science](#) to enrich the bibliography base of the paper and the generative AI system [ChatGPT-4o](#) to structure the content of the paper and to generally improve the readability and language. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors have no competing interests to declare.

Acknowledgments

We have greatly benefited from the comments received from Hideki Sato, Bo Yuan, and the presentation participants at Ghent (2023 Fin-Tech workshop), Giessen (2024 COMPSTAT conference), Kanazawa University, Paris (2025 FMND workshop), and Santiago (2025 CEF conference). We are grateful to the two anonymous referees for their inspiring comments and recommendations. We acknowledge the financial support from the Czech Science Foundation under the project 'Deep dive into decentralized finance: Market microstructure, and behavioral and psychological patterns' [grant number 23-06606S]. This work was supported by the Charles University Research Centre program No. 24/SSH/020 and the Cooperatio Program at Charles University, research area Economics. Jan Sila acknowledges financial support from the Charles University Specific University Research scheme [grant number SVV 260 729]. The usual disclaimer applies.

All the authors have read and approved the final manuscript.

Appendix A. Raw trade data example

See [Table A.4](#).

Appendix B. Categorized explanatory variables

See [Table B.5](#).

Appendix C. Descriptive statistics

See [Table C.6](#).

Data availability

The trades were purchased through the data provider Tardis.dev: <http://www.tardis.dev>. The empirical dataset collected and analyzed during the current study was downloaded from CoinMetrics: <http://www.coinmetrics.io>; from the St. Louis Federal Reserve database: <http://www.fred.stlouisfed.org>; from <http://www.investing.com>; from <http://www.spglobal.com>; from <http://trends.google.com>; from <http://pageviews.wmcloud.org>; and from <http://www.policyuncertainty.com>, as specified in Section 4. All the data are publicly available.

Table A.4

Snapshot of the raw trade data for Bitcoin on [Binance.com](https://www.binance.com) Futures exchange from Nov 15, 2020.

timestamp	local_timestamp	id	side	price	amount
1605398421366000	160 539.421376810	267 330.06	buy	16 065.00	0.135
1605398421501000	160 539.421510626	267 330.07	buy	16 065.00	0.031
1605398421833000	160 539.421845120	267 330.08	sell	16 064.99	0.002
1605398422088000	160 539.422098570	267 330.09	sell	16 064.99	0.007
1605398422248000	160 539.422257947	267 330.10	sell	16 064.99	0.002

The first column shows the timestamps in UTC provided by the exchange, and the second is the timestamp of the arrival of the data to the data provider. The column 'id' identifies a trade, not a trading entity, and is not always provided by the exchange. 'side', 'price', and 'amount' show the direction, price, and size of the trade, respectively.

Table B.5

Categorized explanatory variables.

Variable	Label(s)	Description
1. Blockchain metrics		
Active Addresses	Addresses_BTC \ETH	Number of active addresses on the Bitcoin and Ethereum blockchains, indicating user activity levels.
Hash Rate	Hashrate_BTC \ETH	Measures of the computational power securing the Bitcoin and Ethereum networks, reflecting network security and mining activity.
Blockchain Fees	Fees_BTC \ETH	Total transaction fees paid on the Bitcoin and Ethereum blockchains, serving as indicators of network congestion and transaction demand.
Transaction Counts and Values	Transaction#, TransactionVal_USD \Ntv	Number and adjusted value of transfers between ledger entities, both in USD and native cryptocurrency units.
Difficulty Mean	DiffMean	Average mining difficulty for the Bitcoin network.
2. Crypto exchange flows		
Exchange Flow Ratio	ExFlowRatio	Ratio of exchange inflows (ExInflow) to outflows (ExOutflow) for BTC and ETH, indicating the net movement of these cryptocurrencies onto or off the exchanges.
Asset-Specific Exchange Flows	FlowInEx_BTC \ETH, FlowOutEx_BTC \ETH	Detailed inflow and outflow data for each asset, reflecting trading activity and investor behavior.
3. Market index		
Market Return	BDMI_return	Market return based on the S&P Cryptocurrency Broad Digital Market Index . This index is chosen as an appropriate market representation on the basis of Sila et al. (2025) .
4. Financial variables		
Break-Even Inflation	BreakEvenInflation	Expected inflation derived from the difference between the yields of 10-year U.S. Treasury securities and 10-year U.S. Treasury Inflation-Protected Securities (TIPS).
Interest Rate	InterestRate	Market yield on U.S. Treasury securities at a 1-year constant maturity (ticker: DGS1), representing short-term interest rate expectations.
5. Macroeconomic indicators		
Economic Policy Uncertainty	EPU	An index quantifying the uncertainty surrounding economic policies, which can influence investor sentiment and market volatility.
6. Commodity prices and financial markets		
Gold Price	Gold	Gold Bullion index, indicating investor sentiment toward safe-haven assets.
Commodity Index	GSCI	S&P GSCI Commodity Index provides a broad measure of commodity market performance.
CBOE Volatility Index	VIX	Measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices.
S&P 500 Index	SP500	Proxy for overall U.S. equity market performance.
Exchange Rates	USD_EUR, USD_CNY	Exchange rates of USD to EUR and CNY, connecting crypto markets to international currency movements.

(continued on next page)

Table B.5 (continued).

7. Behavioral and sentiment indicators		
Google Trends Data	GoogleTrends_BTC \ETH \LTC \XRP, GoogleTrends_Crypto	Search interest indices for cryptocurrencies and for cryptocurrency in general, reflecting public attention and sentiment.
Wikipedia Page Views	Wiki_BTC \ETH \LTC \XRP	Daily views of cryptocurrency-related Wikipedia pages, serving as another measure for public attention and sentiment.
8. Supply metrics		
Issuance Continuity Percentage	IssuanceCont	Annualized inflation rate over Bitcoin supplies, indicating the growth in the number of coins.
Total Issuance	IssuanceTotal	Total supply of Bitcoin denominated in USD, reflecting the market value of the circulating supply.
9. Price features		
BTC Price	PriceUSD	Price of Bitcoin over time in USD.
Market Capitalization	Mcap_BTC \ETH \LTC \XRP \Alts	Market capitalization for individual assets and the sum of all altcoins.
Momentum Measures	Momentum_BTC \ETH \Alts	Logarithmic deviation of the current market capitalization from the previous seven days, capturing short-term market trends.
10. Volatility measures		
Asset Volatility	σ _BTC \ETH \LTC \XRP	Calculated using the Garman–Klass estimator (Garman & Klass, 1980) for each asset, providing a measure of intraday price volatility.
Aggregate Volatility Factor	Connectedness	Aggregate volatility measure calculated using the dynamic network methodology of Baruník and Ellington (2024) , which is discussed in detail in Sila et al. (2024) .
11. Trading activity		
Number of Trades	Trades_BTC \ETH \LTC \XRP \Alts	Daily number of trades for each asset and collectively for altcoins, indicating market activity levels.
Trading Volume	Volume_BTC \ETH \LTC \XRP \Alts	Daily trading volumes, serving as a proxy for liquidity and market interest.

The table contains variables' names, technical labels reported in regression outputs and their interpretation, and their description. The variables are divided into 11 categories.

Table C.6
Descriptive statistics of the explanatory variables.

Variable	Mean	St. Dev.	Min	Max
Addresses_BTC_l	13.79	0.16	13.16	14.13
Addresses_ETH_l	13.30	0.13	13.05	13.72
BDMI_return	0.01	0.06	−0.32	0.45
BreakEvenInflation_l	0.83	0.10	0.52	1.02
Connectedness	63.40	17.01	25.76	87.24
DiffMean_l	30.64	0.16	30.25	30.91
EPU_l	4.90	0.47	3.03	6.55
ExFlowRatio	−0.03	0.20	−0.66	1.03
ExInflow_l	21.34	0.51	19.37	22.90
ExOutflow_l	21.39	0.50	19.83	22.91
Fees_BTC_l	1.74	0.93	0.24	4.11
Fees_ETH_l	2.64	0.96	0.09	4.22
FlowInEx_BTC_l	20.73	0.51	18.83	22.19
FlowInEx_ETH_l	20.49	0.66	18.49	22.22
FlowOutEx_BTC_l	20.77	0.52	19.39	22.18
FlowOutEx_ETH_l	20.53	0.64	18.34	22.26
Gold_l	7.50	0.03	7.43	7.58
GoogleTrends_BTC_l	2.30	0.36	1.51	3.27
GoogleTrends_Crypto_l	5.23	0.56	3.51	6.53
GoogleTrends_ETH_l	4.34	0.49	2.65	5.54
GoogleTrends_LTC_l	2.75	0.40	1.90	3.81
GoogleTrends_XRP_l	2.60	0.43	1.25	4.17
GSCI_l	6.22	0.13	5.91	6.44
Hashrate_BTC_l	18.79	0.19	17.87	19.19
Hashrate_ETH_l	6.27	0.38	5.53	6.86
InterestRate	3.36	3.31	1.00	19.50
IssuanceCont_l	0.56	0.14	−0.35	0.87
IssuanceTotal_l	17.45	0.36	16.34	18.09
Mcap_Alts_l	25.98	0.61	24.49	26.79
Mcap_BTC_l	27.40	0.33	26.42	27.87

(continued on next page)

Table C.6 (continued).

Mcap_ETH_I	26.28	0.60	24.65	27.06
Mcap_LTC_I	23.13	0.32	22.14	23.98
Mcap_XRP_I	25.02	0.47	23.77	25.93
Momentum_Alts	0.01	0.11	-0.45	0.47
Momentum_BTC	0.01	0.07	-0.22	0.22
Momentum_ETH	0.01	0.09	-0.37	0.32
PriceUSD_I	10.65	0.32	9.68	11.12
σ _BTC	0.04	0.02	0.01	0.23
σ _ETH	0.05	0.03	0.01	0.37
σ _XRP	0.07	0.06	0.01	0.49
σ _LTC	0.06	0.04	0.01	0.40
SP500_I	8.35	0.08	8.18	8.48
Trades_Alts_I	14.97	0.65	13.06	17.01
Trades_BTC_I	14.31	0.37	13.46	15.66
Trades_ETH_I	13.90	0.48	12.52	15.53
Trades_LTC_I	12.63	0.67	11.16	14.42
Trades_XRP_I	13.27	0.70	11.76	15.25
Transaction#_I	13.59	0.16	12.98	13.95
TransactionVal_Ntv_I	12.44	0.40	11.22	13.50
TransactionVal_USD_I	23.09	0.51	21.61	24.32
USD_CNY_I	1.86	0.01	1.84	1.89
USD_EUR_I	-0.17	0.03	-0.21	-0.11
Velocity_BTC_I	4.31	0.40	3.26	5.53
Velocity_ETH_I	3.52	0.50	1.42	4.83
VIX_I	2.98	0.18	2.71	3.62
Volume_Alts_I	22.41	0.79	20.93	25.56
Volume_BTC_I	11.03	0.48	9.82	12.78
Volume_ETH_I	13.50	0.59	11.95	15.28
Volume_LTC_I	13.83	0.66	12.32	15.72
Volume_XRP_I	20.25	0.74	18.64	22.88
VolumeExch_I	24.44	0.40	23.41	26.58
Wiki_BTC_I	9.71	0.39	8.87	11.05
Wiki_ETH_I	8.80	0.41	7.52	10.61
Wiki_LTC_I	7.17	0.54	6.20	8.52
Wiki_XRP_I	7.24	0.50	6.39	8.90

The values are in daily granularity, covering the period from November 15, 2020, to January 31, 2022, i.e., 443 observations. The variables are sorted in alphabetical order.

References

- Alexander, C., Choi, J., Massie, H. R., & Sohn, S. (2020). Price discovery and microstructure in ether spot and derivative markets. *International Review of Financial Analysis*, 71, Article 101506.
- Alexander, C., & Heck, D. F. (2020). Price discovery in bitcoin: The impact of unregulated markets. *Journal of Financial Stability*, 50, Article 100776.
- Aloosh, A., & Li, J. (2024). Direct evidence of bitcoin wash trading. *Management Science*, 70(12), 8875–8921.
- Alvarez, F., Argente, D., & Patten, D. V. (2023). Are cryptocurrencies currencies? bitcoin as legal tender in El Salvador. *Science*, 382(6677), eadd2844.
- Amiram, D., Bozanic, Z., & Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20, 1540–1593.
- Amiram, D., Lyandres, E., & Rabetti, D. (2025). Trading volume manipulation and competition among centralized crypto exchanges. *Management Science*, 71(10), 8604–8622.
- Andrada-Félix, J., Fernandez-Perez, A., & Sosvilla-Rivero, S. (2020). Distant or close cousins: Connectedness between cryptocurrencies and traditional currencies volatilities. *Journal of International Financial Markets, Institutions & Money*, 67, Article 101219.
- Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, 64, 74–81.
- Baruník, J., & Ellington, M. (2024). Persistence in financial connectedness and systemic risk. *European Journal of Operational Research*, 314(1), 393–407.
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Bouri, E., & Jalkh, N. (2023). Spillovers of joint volatility-skewness-kurtosis of major cryptocurrencies and their determinants. *International Review of Financial Analysis*, 90, Article 102915.
- Charfeddine, L., Benlagha, N., & Khediri, K. B. (2022). An intra-cryptocurrency analysis of volatility connectedness and its determinants: Evidence from mining coins, non-mining coins and tokens. *Research in International Business and Finance*, 62, Article 101699.
- Chen, J., Lin, D., & 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, Article 126405.
- Cong, L. W., Li, X., Tang, K., & Yang, Y. (2023). Crypto wash trading. *Management Science*, 69(11), 6427–6454.
- Cong, L. W., Li, Y., & Wang, N. (2021). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3), 1105–1155.
- Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62, 182–199.
- Dodge, Y. (2008). *The Concise Encyclopedia of Statistics*. Springer New York.
- Feldkircher, M., & Zeugner, S. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of Statistical Software*, 68(4), 1–37.
- Foley, S., Karlsen, J. R., & Putninš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?. *The Review of Financial Studies*, 32(5), 1798–1853.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage.
- Fusaro, T., & Hougan, M. (2019). Bitwise asset management-presentation to the us securities and exchange commission [www document]. <https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf>, [Online; www document; (Accessed 13 August 2024)].
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1), 67–78.
- Grinblatt, M., & Keloharju, M. (2004). Tax-loss trading and wash sales. *Journal of Financial Economics*, 71(1), 51–76.
- Grossman, S. J., Miller, M. H., Cone, K. R., Fischel, D. R., & Ross, D. J. (1997). Clustering and competition in asset markets. *The Journal of Law and Economics*, 40(1), 23–60.
- Gu, Z., Lin, D., & Wu, J. (2022). On-chain analysis-based detection of abnormal transaction amount on cryptocurrency exchanges. *Physica A. Statistical Mechanics and its Applications*, 604, Article 127799.
- He, S., Manela, A., Ross, O., & von Wachter, V. (2022). Fundamentals of perpetual futures. *SSRN Electronic Journal*.
- Hougan, M., Kim, H., & Lerner, M. (2019). Economic and non-economic trading in bitcoin: exploring the real spot market for the world's first digital commodity. <https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5574233-185408.pdf>, [Online; www document; (Accessed 13 August 2024)].
- Ji, Q., Bouri, E., Lau, C. K. M., & Roubaud, D. (2019). Dynamic connectedness and integration in cryptocurrency markets. *International Review of Financial Analysis*, 63, 257–272.
- Kočenda, E., & Albrecht, P. (2026). Event-driven changes in return connectedness among cryptocurrencies. *Financial Innovation*, (forthcoming).
- Kukacka, J., & Kristoufek, L. (2023). Fundamental and speculative components of the cryptocurrency pricing dynamics. *Financial Innovation*, 9, 61.

- Leamer, E. E. (1978). Vol. 53, *Specification searches: Ad hoc inference with nonexperimental data*. New York: Wiley.
- Li, Z., Cong, L., & Wang, H. (2004). Discussion on Benford's Law and its application. arXiv preprint math/0408057.
- Li, X., & Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin. *Decision Support Systems*, 95, 49–60.
- Mahmoodzadeh, S., & Gençay, R. (2017). Human vs. high-frequency traders, penny jumping, and tick size.. *Journal of Banking & Finance*, 85, 69–82.
- Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2), 293–319.
- Mark, M., Sila, J., & Weber, T. A. (2022). Quantifying endogeneity of cryptocurrency markets. *The European Journal of Finance*, 28(7), 784–799.
- Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. *The Review of Economics and Statistics*, 95(2), 591–616.
- Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29(1), 46–75.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Technical report*.
- Osterrieder, J., & Lorenz, J. (2017). A statistical risk assessment of bitcoin and its extreme tail behavior. *Annals of Financial Economics*, 12(01), Article 1750003.
- Pennec, G. L., Fiedler, I., & Ante, L. (2021). Wash trading at cryptocurrency exchanges. *Finance Research Letters*, 43, Article 101982.
- Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437), 179–191.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7(4), 532–547.
- Sifat, I., Tariq, S. A., & van Donselaar, D. (2024). Suspicious trading in nonfungible tokens (NFTs). *Information & Management*, 61(1), Article 103898.
- Sila, J., Kocenda, E., Kristoufek, L., & Kukacka, J. (2024). Good vs bad volatility in major cryptocurrencies: The dichotomy and drivers of connectedness. *Journal of International Financial Markets, Institutions and Money*, 96, Article 102062.
- Sila, J., Mark, M., Kristoufek, L., & Weber, T. A. (2025). Crypto market betas: the limits of predictability and hedging. *Financial Innovation*, 11(1), 1–28.
- Stock, J. H., & Watson, M. W. (2003). *Introduction to econometrics*. Boston: Addison-Wesley.
- Tošić, A., Vičić, J., & Hrovatin, N. (2025). Beyond the surface: advanced wash-trading detection in decentralized NFT markets. *Financial Innovation*, 11(1), 86.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Vergne, J. (2020). Decentralized vs. distributed organization: blockchain, machine learning and the future of the digital platform. *Organization Theory*, 1(4), Article 2631787720977052.
- Victor, F., & Weintraud, A. M. (2021). Detecting and quantifying wash trading on decentralized cryptocurrency exchanges.. In *Proceedings of the web conference 2021* (pp. 23–32).
- Wang, S., Cheng, N., & Zhang, T. (2025). Wash trading and insider sales in NFT markets. *Journal of Banking & Finance*, 180, Article 107529.
- Wang, J., Ma, F., Bouri, E., & Guo, Y. (2023). Which factors drive bitcoin volatility: Macroeconomic, technical, or both?. *Journal of Forecasting*, 42(4), 970–988.
- Zhang, Z., & Zhao, R. (2023). Good volatility, bad volatility, and the cross section of cryptocurrency returns. *International Review of Financial Analysis*, 89, Article 102712.