COINMETRICS

# TRUSTED EXCHANGE FRAMEWORK 2.3

By Victor Ramirez, Jo Suddreth,
and the Coin Metrics Team

# INTRODUCTION

Centralized exchanges are a vital part of the crypto ecosystem, serving as the primary interface to the blockchain for a significant percentage of market participants. Yet, exchanges vary widely in quality across several use-cases, whether as custodial platforms, as data sources, or as entities for executing programmatic actions.

The Coin Metrics' Trusted Exchange Framework thus aims to quantitatively assess exchanges to promote transparency, innovation, and trust for the industry and its users. The common usage patterns for an exchange are translated to criteria that define the fundamental properties of exchange trustworthiness: transparency, resilience & security, data quality, regulatory compliance, and API quality. The criteria sources public information about an exchange such as incident history, financial statements, and license disclosure as well as market activity that can be derived from exchange data.

During the course of our research, we conducted a comprehensive literature review to identify prior techniques in evaluating exchanges, including the extensive research in detecting wash trading, fake volume, and fraud. Several of these techniques are included in our framework. We also made use of Coin Metrics' unique experience of maintaining our market data collection system for over 50 exchanges over the past six years, which involves extensive interaction with exchanges' APIs and regular evaluations of data quality issues and interruptions in service. Our framework contributes to the literature by utilizing an approach that is primarily quantitative to calculate exchange features (keeping subjective determinations to a minimum) and presenting the most complete collection of all facets of exchange trustworthiness to date.

Coin Metrics utilizes the output from our Trusted Exchange Framework to select high-quality constituent exchanges in our prices, indexes, and metrics.

**Interested in learning more about the Coin Metrics products mentioned in this report?**

Get in touch with us.

# RELEASE NOTES

## 2.3 (July 2025)

- Data Quality:
  - Added a test for the frequency of round-trip trades. See section “Round Trip Trades”.
  - Increased the weight of the trade permutations test.
    - **Why we made this change:** The trade permutations test continues to provide a strong signal for market manipulation.
- Regulatory:
  - Adjusted the score to be based on the number of active licenses an exchange has. The list of licenses can be found on Table 3.
    - **Why we made this change:** Over time, some of our criteria have stopped providing differentiating signals. Most exchanges now offer KYC, and by virtue of being licensed across the world, exchanges will offer some fiat trading. We wanted to simplify the previous regulatory score, provide more differentiation across regulated entities, and adapt it to the compliance landscape of the crypto industry today.
- Applied the Trusted Exchange Framework criteria to DeFi exchanges dYdX and Uniswap.
- You may notice that grades are relatively lower compared to previous versions. This is largely due to the Regulatory Compliance and the Data Quality criteria being stricter and more comprehensive. The new scale for the scores are not a reflection that exchanges are getting worse, but rather, that we raised the standards for quality in the framework. We believe that these enhancements to our criteria lead to a more nuanced differentiation of perceived quality in exchanges.

## 2.2 (April 2024)

- Data Quality: For Benford’s Law, Trade Sizes, and Trade Permutations, we’ve tightened the requirements to pass and changed how we are weighing volume for computing errors. For price discovery and price anomalies, we allow partial credit for exchanges that fall within the middle-tail of the distribution.
- Resilience & Security: Added Bug Bounties.
- Transparency: Recalibrated our scoring slightly to allow for a higher score if an exchange has multiple methods demonstrating proof of reserve/liability (e.g. Public Address balances and 3rd Party Audit). Increased the value of Publicly Disclosed Financials which clearly report digital asset holdings and liabilities to a government regulatory entity.
- Regulatory: Added VASP license and FATF graylist flags. Adjusted for how regulatory agency quality is factored based on the World Bank’s methodology[7] for regulatory quality by jurisdiction.
- API Quality: Recalibrated how API outages are factored to reduce false positives and follow a more logarithmic scale.

## 2.1 (October 2023)

In response to feedback to the first iteration of the Trusted Exchange Framework V2, several improvements were made to create a simpler and easier to interpret framework. Below is a quick summary of what's new.

- Numeric scores have been translated into a letter grading scale for convenience. See Grading Scale.
- The Regulatory score was revamped to be significantly more reflective of how crypto exchanges are regulated globally. Exchanges are no longer penalized for not operating under the U.S. if they serve their international customers in a manner that is deemed acceptable to regulators in reputable jurisdictions. The new score is more focused on the regulatory compliance framework rather than general compliance (such as security standards, which are now moved to Resilience & Security)
- Transparency and Resilience were broken out into their own separate scores to make for more mutually exhaustive groupings.
- Infrastructure now refers to API Quality. New criteria such as rate limit assessments were added.

Although the notion of a "Trusted Exchange" is a bit of a misnomer and highly dependent on the context, the framework serves as a rough approximation of how to holistically evaluate exchanges based on the most important qualities it must have. As noted in the past, we highly encourage users to think closely about what context they may evaluate exchanges on, and apply that context with the evaluation criteria within the Framework in mind. Treat this framework more like a model; "All models are wrong, but some are useful".

Please note our **FAQ** for common questions regarding our framework.

## 2.0 (March 2023)

The original framework was primarily focused on quantifying the amount of fake volume per exchange. New techniques have since been developed that directly measure the footprints of fake volume using a wider variety of market data from the exchange. Additionally, the new framework expands beyond quantifying fake volume and into providing a more holistic assessment of an exchange's trustworthiness.

# OVERALL RANKINGS

## CENTRALIZED EXCHANGES

| EXCHANGE | DATA QUALITY (SPOT) | TRANSPARENCY | RESILIENCE & SECURITY | REGULATORY COMPLIANCE | API QUALITY | GRADE |
|---|---|---|---|---|---|---|
| Bitstamp | A | B+ | B+ | A- | B+ | A- |
| Coinbase | A | B+ | C | A | A- | A- |
| CME | N/A | B+ | A | A | B- | A- |
| Kraken | A | A | C | A- | C- | A- |
| Gemini | A | C+ | C+ | A | B+ | A- |
| Binance | B+ | A | B | B | C- | B+ |
| Crypto.com | B- | B- | B+ | A | B+ | B+ |
| OKX | A | B+ | B- | B- | C- | B |
| Gate.io | A | A- | A | C- | C | B |
| HTX | B | B+ | A- | B- | C- | B |
| Bybit | B+ | A | C- | C | C+ | B |
| Deribit | B+ | B+ | A | C- | B- | B- |
| Itbit | A | D | A- | B | D | B- |
| Bullish | B+ | B+ | A | C | D | B- |
| Bithumb | A | B+ | C+ | C- | C | B- |
| BitMEX | N/A | B+ | B+ | C | C | C+ |
| Binance US | B- | D | B+ | B- | B+ | C+ |
| Bitbank | A | D | A- | C | C- | C+ |
| bitFlyer | C+ | D | C+ | A- | C- | C |
| LMAX | B- | D | C+ | B- | C- | C |
| Upbit | B- | B+ | B- | C- | D | C |
| KuCoin | A | A | B+ | D | D | C |
| MEXC | B- | B+ | B | D | D | C- |
| Bitfinex | A | D | C | D | C- | C- |
| HitBTC | B+ | D | C- | D | B- | C- |
| Poloniex | D | B+ | C | D | B- | D |
| LBank | D | D | C | D | D | D |

| EXCHANGE | DATA QUALITY (SPOT) | TRANSPARENCY | RESILIENCE & SECURITY | REGULATORY COMPLIANCE | API QUALITY | GRADE |
|---|---|---|---|---|---|---|
| dYdX | A | A | C+ | N/A | C | A- |
| Uniswap | B | A | C+ | N/A | A | B+ |

*Table 1. Overall Rankings. *Exchanges without sufficient data were not penalized for their data quality scores. They are however disqualified for being included in the trusted volume metric.*

## Grading Scale

We translated the numeric scores into a letter grading scale for convenience. The table below illustrates how to interpret the resulting scores.

| GRADE | INTERPRETATION |
|---|---|
| A- to A | These exchanges excel in most or all of the factors assessed. Exchanges in this tier often have the highest quality data and efficient markets relative to its peers. |
| B- to B+ | These exchanges are generally of good quality across most of the factors assessed. Minor penalties in Data Quality, Regulatory Compliance, Transparency, and/or Resilience & Security keep exchanges in this tier from being in the A tier. |
| C- to C+ | These exchanges are generally middle-of-the-road across most of the factors assessed. Some exchanges in this tier may have organic trading activity, but suffer from incomplete compliance and/or past minor security incident(s), and vice versa. |
| D | These exchanges score poorly across most of the factors assessed. Exchanges in this tier tend to not be compliant in major jurisdictions, have suffered major security incidents, and/or contain significant amounts of inorganic market activity. |

*Table 2. Grading Scale*

# Weights

Each category is scored across numerous subcategories (see next section) and normalized to a max score of 1. These categories are then computed with a weighted average to create an overall score. The categories are weighted using the following values:

| EXCHANGE TYPE | DATA QUALITY (SPOT) | TRANSPARENCY | RESILIENCE & SECURITY | REGULATORY COMPLIANCE | API QUALITY | TOTAL |
|---|---|---|---|---|---|---|
| **Spot** | 30% | 15% | 15% | 30% | 10% | 100% |
| **Futures-Only*** | 0% | 21.4% | 21.4% | 42.9% | 14.3% | 100% |
| **DeFi** | 40% | 15% | 30% | 0% | 10% | 100% |

*Table 3. Weights.*

**Futures-Only weights were derived by omitting spot data quality and proportionally distributing weights across the remaining categories. Exchanges with an N/A for Data Quality (Spot) are applied with the Futures-Only weighting.*

Note that the "true" value of the weights are relative to the context of how an exchange is being used. A custody-focused use-case looking to avert risk may downweight Data Quality and API Quality in favor of Transparency, Resilience & Security, and Regulatory Compliance. In contrast, a data-provider that relies on exchange APIs but does not hold assets in custody in exchanges may want to weigh these categories in the opposite direction. We publish the scores for each category so that users with unique use cases can make their own assessment if needed.

As of version 2.3, we applied a separate weighing scale for DeFi exchanges. We did not apply the regulatory compliance to DeFi due to the uncertainty of how permissionless protocols are regulated. To counteract this, we scaled up the weighing for Data Quality and Resilience & Security. Resilience is of particular emphasis due to the prevalence of hacks and larger surface area for smart contract risks in DeFi. This scale is still experimental and is subject to change as we add more applicable criteria for DeFi and expand our DeFi exchange coverage universe.

Each category is explained in depth in the next section along with a description of each feature used to evaluate exchange trustworthiness.

# CATEGORIES

The Trusted Exchange Framework categorizes the major categories of exchange trustworthiness as: Data Quality, Transparency, Resilience & Security, Regulatory Compliance, and API Quality. These categories are broadly defined below.

## Data Quality (Spot)

The Data Quality score assesses the confidence that an exchange's reported data is accurate. Exchanges are the primary source for market data metrics such as price and volume, yet exchange-reported data has historically been fabricated. Crypto exchanges are known to create "fake" or non-economic volume to attract users to its platform.

Estimates vary on how much of the reported volume is fake. Bitwise estimated in 2019 that 95% of volume reported by exchanges are fake by observing anomalous trade patterns.[2] Forbes estimated in 2022 that about 51% of bitcoin trading volume is fake by weighting likelihood of fake volume given regulatory and web usage patterns.[3] However, fake volume boiled down to a percent value is misleading because a) it is trivial to manufacture artificial volume and b) it depends on which exchanges are included in the calculation, the reported volume of each exchange, and the proportionality of volume that is fake. For example, if a well-behaved crypto exchange accurately reports $100M of volume in one day and an ill-behaved crypto exchange reports $1.9B of mostly fake volume in one day, then ~95% of the volume between those two exchanges is fake.

Fake volume is not evenly distributed across crypto exchanges. The variance in our data quality test results is strong evidence of this. Thus, we want to only include trustworthy exchanges and exclude exchanges whose volumes are shown to be mostly inorganic when calculating metrics such as total trading volume.

This category utilizes several techniques that have been used to identify fake volume on an exchange: the distribution of leading digits and fitting against Benford's Law, quantifying cross-correlation of volume across markets, examining the distribution of buy/sell flag permutations on trade sequences, examining the distribution of trade sizes, an analysis of lead/lag of asset prices to determine where price discovery occurs, an analysis of pricing anomalies, and identifying round trip trades.

# Trusted Volume

A score of **B** or above on the Data Quality category qualifies an exchange to be a part of our "trusted volume" universe, an important designation that is used to select high-quality constituent exchanges for the calculation of Coin Metrics' prices, indexes, metrics, and other data products. An exchange with this designation is determined to have *most* of its trading volumes to be organic.

# Calculation Methodology

## Data Sampling

All data was sourced from Coin Metrics Market Data Feed. Due to large amounts of transaction data that can accrue over time, it's unfeasible to apply all of these techniques across all transactions and markets for a long period of time. These techniques were instead applied on a synthetic 24 hour dataset randomly sampled across the most liquid markets across several time windows bounded by a time period of interest (e.g. Q3 2023-Q1 2024) to minimize sampling bias.

## Scoring

Each feature is scored by how well an exchange's observed market data is distributed relative to a known ideal or expected value or distribution. Using a representative sample of market data for each exchange, a goodness-of-fit score is calculated against the expected distribution for a given feature.

For tests that have a "true" distribution to test against, such as the Leading Digit Test (Benford's Law), Trade Patterns (not a uniform distribution) and Trade Sizes (log normal), we calculate the mean-absolute error for each market, weight the error by volume for a given exchange, and compute the z-score across all exchanges. An exchange passes if its z-score is within a reasonable range for a test and fails when it's outside.

For tests where there exists no perfect score but imply some directional relationship between well-behaved exchanges (Volume Correlation), we group each market by the metric (correlation), weigh the metric by volume, and calculate the z-score. An exchange passes if its z-score is within a reasonable range for a test and fails when it's outside.

For tests where a perfect score exists, such as frequency of price anomalies (0%), frequency of round trip trades (0%), or price discovery lag (0s), exchange scores are calculated using a sigmoid function whose midpoint is close to some reasonable amount close to the median of the distribution. These scores are continuous and are better the more desirable the feature.

Note that this score represents an estimated "confidence level" (not in the traditional statistical sense) of how likely most of an exchange's volume is representative of organic and informed market activity. An exchange that fails one test signals a moderate amount of confidence about the exchange's data quality but does not rule out the possibility that most of the exchange's volume is organic due to the likelihood of false positives. An exchange that fails multiple tests signals a general lack of confidence in data quality. Similarly, an exchange

with a perfect score signals a general confidence for data quality but does not imply that their data is 100% accurate. Thus, the score is not meant to be interpreted as a strict probability or proportion of data that is legitimate.

| Data Quality Test | Computation | Threshold |
|---|---|---|
| Leading Digit Distribution | Z-score of the mean absolute error against Benford's Law weighted by volume. | Pass if z-score < 1. Partial credit if <= 1 z-score < 2. Fail if z-score > 2. |
| Volume Correlation | Z-score of the average correlation weighted by volume. | Pass if z-score > -1. Fail if z-score <= 1. |
| Trade Permutations | Z-score of the mean absolute error against an even distribution of trade side permutations. | Pass if z-score > -1. Partial credit if -1 >= z-score > -2 . Fail if z-score < -2. |
| Trade Size Distribution | Z-score of the mean absolute error of the residuals. | Pass if z-score < 1. Partial credit if <= 1 z-score < 2. Fail if z-score > 2. |
| Price Discovery | Sigmoid function. Midpoint = 0.6s (between 0.2s and 1s). K = 100. | Pass if < 0.2s. Continuous score from 0-1 at x more 0.2s to 1s. Fail as the price discovery approaches and exceeds 1s. |
| Price Anomalies | Sigmoid function. Midpoint = 5.5% (between 1% and 10%). K = 100. | Pass if <1%. Continuous score from 0-1 at x = 1-10%. Fail as the price anomaly frequency approaches and exceeds 10%. |
| Round Trip Trades | Sigmoid function. Midpoint = 10%. K = 1 | Pass if < 10%. Continuous score from 0-1. Approaches zero as x exceeds 10%. |

*Table 4. Data Quality test scoring methodology.*

## Subcategories

### Leading Digit Distributions (Benford's Law)

An assessment of how well an exchange's trade patterns follow a natural order of leading digits where leading digits tend to be small, also known as Benford's Law. Benford's Law has been used to detect fraud in financial (such as trade amounts in traditional markets) and non-financial applications (such as elections) where the distributions of quantities of leading digits do not follow Benford's Law. If an exchange's trade amounts in base units do not fit Benford's Law, it's an indicator of manipulated behavior.[4] See Figure 1.

## Volume Correlations

An assessment of how correlated an exchange's volume is relative to its peers. Well-behaved exchanges tend to behave similarly to each other and on average have a higher correlation of relative changes in volume across time. More precisely, we expect volume to increase and decrease at the same time across exchanges in response to material releases of new information. Inversely, less well-behaved exchanges are different in different ways, and thus exhibit lower correlation across the rest of the exchanges. Exchanges that manipulate their reported trading volume via artificial processes such as wash trading exhibit volume profiles that differ from their legitimate peers.[2]

## Trade Permutations

An assessment of the legitimacy of the distribution of trade buy/sell flags of an exchange. Exchanges that have historically fabricated volume are known to show an even distribution of buy/sell flags when examining trade sequences, likely due to wash trading, non-economic trading activity, or other trading activity generated from an artificial process. Legitimate market activity tends to heavily skew towards several consecutive buy or sell trades due to the presence of informed traders that are willing to cross the spread and take liquidity in response to material new information.[5] See Figure 2.

## Trade Sizes

A measure of how trade sizes are distributed on an exchange. Organic trading activity tends to result in a log-normal distribution of trade size, which is normally distributed when the log transformation is applied to both trade counts and trade sizes. This can be ascribed to the presence of retail and institutional market participants as well as the effect of liquidity constraints on order sizing. In contrast, inorganic trading activity generated from artificial processes can exhibit a different distribution with large numbers of trades executed with unusual trade sizes. The degree to which the distribution of trade sizes on an exchange follows or deviates from this distribution can be measured by calculating the error from the residuals between the log of observed trade sizes and the log of the log-normal probability distribution function. See Figure 3.

## Price Discovery

A measure of the lead/lag of asset prices on an exchange relative to a benchmark price. Exchanges that are centers of price discovery tend to lead price movement by an observable amount of time. This is measured using the Hayashi-Yoshida Estimator, allowing for a ranking of exchanges based on their lead/lag dynamics.[6] Exchanges found to lead other exchanges represent exchanges where price discovery occurs and thus score more favorably for this criteria.

In our implementation, we use a 24-hour sample of trades collected from each considered market and aggregate these trades onto a 10th-second time grid using volume-weighted-average price. The result is a series of prices, at 10th-second granularity, for each considered market. The Hayashi-Yoshida Estimator is calculated using these series and a reference market's time series. The time by which each market leads or lags the reference market is then identified, allowing exchanges to be ranked by their relative role in leading asset prices. See Figure 4.

### Pricing Anomalies

A measure of the frequency in price anomalies. A price anomaly is defined as having a price beyond 2 standard deviations across a common set of markets. The score is assigned based on the frequency of price anomalies for a given exchange's most common price pairs.

### Round Trip Trades

The estimated proportion of volume of "round-trip" trades. A pair of round trip trades is defined as two trades with nearly identical prices and amounts executed on the opposite side (buy or sell) in close proximity in time and sequence order. These trades inflate a market's trading volume with activity that involves zero market risk. Although not all trades using this definition are a direct result of market manipulation, a high frequency of these trades with nearly identical footprints strongly suggests trading activity that deviates from expected market activity.

## Transparency

The Transparency score is an assessment of the quality of publicly disclosed information from an exchange. This is valued using the following heuristic:

- Quality of Proof of Reserves (PoR) is assessed based on Nic Carter's PoR rubric. [1]
- Publicly disclosed financial information is graded relative to the PoR rubric.
- If an exchange has both PoR and publicly disclosed financial information, their Transparency score is the highest of the two.
- Exchanges may be exempt from PoR if it is not relevant to their operations, i.e. if they do not custody crypto and settle their contracts in cash.
- Wallet disclosures, "proof" of assets but no credible proof of liabilities, ownership or oversight give partial credit.

In contrast to the Regulatory Compliance category described in further detail below, the Transparency criteria here focuses on the *self-regulating* processes that an exchange offers. Criteria in this category include the quality of an exchange's PoR (where applicable), public disclosure of finances, and the public disclosure of an exchange's addresses.

## Subcategories

### Proof of Reserves Quality

This criteria evaluates the quality of an exchange's proof of reserves attestation. A selection of major exchanges have begun publishing proof of reserves attestations, yet a closer examination of these disclosures reveal that the attestations published are of varying quality.  Quality is assessed on the basis of cryptographic

verification of just assets or both assets and liabilities, the breadth of assets covered in the proof of reserves attestation, the frequency of the proof, user verification of liabilities, and the presence of a third party audit. These features are individually converted to a binary flag and then summed to create a score between 0 and 6, with 6 indicating a proof of reserves attestation with the strongest assurances. Our methodology is informed by Nic Carter's prior research on proof of reserves.[1]

## Publicly Disclosed Financials

An indicator for whether an exchange has disclosed their finances to the public. Exchanges that do not publish Proof of Reserves can get partial credit for disclosing their assets and liabilities.

## Wallet Disclosures

An indicator for whether an exchange has disclosed their wallets or if they can be traced on-chain. This criteria can be thought of as a small fraction of fulfilling a complete Proof of Reserves, as this does not include proof of liabilities, a third party audit, frequent updates, or cryptographic proof of ownership.

# Resilience & Security

**Resilience & Security** refers to how well an exchange protects its users. This criteria assesses exchanges for enacting proactive compliance measures such as complying to security standards (SOC2 Type II or ISO/IEC 27001) and penalizes exchanges for historical major market incidents (extensive pausing of withdrawals due to market conditions) and security incidents (hacks, data breaches). Incidents are weighted by recency and value lost.

## Subcategories

### SOC 2 Type 2 or ISO/IEC 27001

A binary flag for whether an exchange has demonstrated the ability to meet standardized security compliance procedures, such as SOC 2 or ISO/IEC 27001. Developed by the American Institute of CPAs (AICPA), SOC 2 defines criteria for managing customer data based on five "trust service principles"—security, availability, processing integrity, confidentiality and privacy. ISO/IEC 27001 is an international standard to manage information security.

### Offers Insurance

A binary flag for whether an exchange offers cash or crypto insurance on customer deposits.

### Offers a Bug Bounty

A binary flag for whether an exchange offers a reward for uncovering security vulnerabilities.

### Security Incident History

A score for whether an exchange has suffered a major security incident, defined as a breach in the exchange that leads to the exposure of private consumer data or loss of customer funds. Major security incidents were identified by searching major news publications focused on coverage of cryptocurrencies. A score is calculated that is a function of the recency of the incident and the amount of lost value in U.S. dollars, where exchanges that have experienced a more severe loss of funds are penalized more but where any penalty decays gradually over time.

### Market Incident History

A binary flag for "market incidents", defined as whether an exchange has paused withdrawals for reasons beyond regular site maintenance or known exogenous events (such as the Ethereum Merge). An exchange can pause withdrawals due to a loss in banking relationships or in response to a serious security incident that compromises the security of their wallets.

# Regulatory Compliance

The Regulatory Compliance score is an assessment on an exchange's ability to meet regulatory requirements across multiple jurisdictions via its existing licenses.

## Subcategories

### Regulatory Licenses

The number of regulatory licenses an exchange has secured from a whitelisted set of licenses. The more licenses an exchange has, the higher the score. These licenses are listed below.

| Code | License |
|---|---|
| FATF | FATF Travel Rule Compliance (AML/CFT Standards for VASPs) |
| US_MTL | U.S. State Money Transmitter License |
| US_BitLicense | New York State Department of Financial Services (NYDFS) BitLicense |
| US_FinCEN | FinCEN Money Services Business (MSB) Registration |
| US_SEC | U.S. SEC Broker-Dealer License |
| US_ATS | SEC-Registered Alternative Trading System (ATS) |
| US_CFTC | U.S. CFTC Oversight / Registration for Derivatives |
| EU_MiCA | EU MiCA CASP License (Crypto Asset Service Provider) |
| EU_EMI | EU Electronic Money Institution (EMI) License |
| UK_FCA_Registration | FCA Cryptoasset Business Registration (under AML regulations) |
| UK_FCA_Authorization | FCA Authorization for Regulated Financial Activities |
| SG_DPT | Singapore MAS Digital Payment Token (DPT) License |
| SG_PI | Singapore MAS Standard/Major Payment Institution License |
| HK_SFC | Hong Kong SFC Type 1 & Type 7 Licenses (Securities & Automated Trading) |
| AU_AUSTRAC | Australia AUSTRAC Digital Currency Exchange (DCE) Registration |
| AU_AFSL | Australian Financial Services License (AFSL) from ASIC |
| CH_AML | Switzerland FINMA Financial Intermediary / VQF (AML-SRO) Membership |
| CH_Bank | Switzerland FINMA Banking License |
| JP_FSA | Japan FSA Virtual Currency Exchange License |
| UAE_VARA | Dubai VARA Virtual Asset Service Provider (VASP) License |
| UAE_ADGM | Abu Dhabi ADGM FSRA Financial Services Permission (FSP) for VASPs |

*Table 5. List of licenses examined for the Regulatory Compliance score.*

# API Quality

The API Quality score is an assessment of an exchange's quality to be used as a programmable entity. The quality of technical infrastructure informs the ease-of-use of being able to execute actions programmatically, such as for reading data or executing trades, via an exchange's API. Quality is assessed based on quantitative and quantitative factors that the Coin Metrics team has identified from building data feeds from each exchange on this list.

Criteria in this category include the availability of an exchange's historical data and whether the exchange offers features that are critical for users who wish to collect market data or trade programmatically: a streaming API interface, a FIX API interface, a status page, trade buy/sell indicators, unique trade identifier, trade execution time, and sequential integer trade IDs. The selection of these features are informed by Coin Metrics' experience in developing and maintaining our market data collection system.

## Subcategories

### Historical Data

A binary flag for indicating whether an exchange allows users to query historical trades data. Exchanges differ in the amount of historical trades that are served via their API. Some exchanges only allow a user to query a fixed amount of trades, such as the past 1,000 trades that occurred on a market, or a fixed time window, such as the previous 24 hours of trades. The most transparent exchanges offer the full history of trades starting from the inception of the exchange. Exchanges that limit the ability to query historical data receive a 0 while exchanges that offer full history receive a 1.

### FIX API

A binary flag indicating whether an exchange offers a FIX API interface.

### Status Page

A binary flag indicating whether an exchange has a status page.

### Buy/Sell Indicator

A binary flag indicating whether an exchange serves trades data with a buy/sell flag.

### Unique Trade Identifier

A binary flag indicating whether an exchange's API provides a unique trade identifier.

### Sequential Integer Trade ID

A binary flag indicating whether an exchange's API returns a trade ID that is ordered sequentially by time. Exchanges that offer a sequential integer trade ID allow for market participants to independently verify that they have the complete set of trades from an exchange by looking for any gaps in trade IDs.

### Rate Limiting

An indicator for how well an exchange API's rate limits are commensurate to actual usage. Penalties are incurred if an exchange has no rate limits or if an exchange's rate limits are overly prohibitive. The proportion between the exchange API's rate limits and actual usage is based on empirical observations from extracting real-time data.

### Outages

A metric for the duration of API outages that the exchange has experienced. Exchanges are penalized based on how long outages occur.

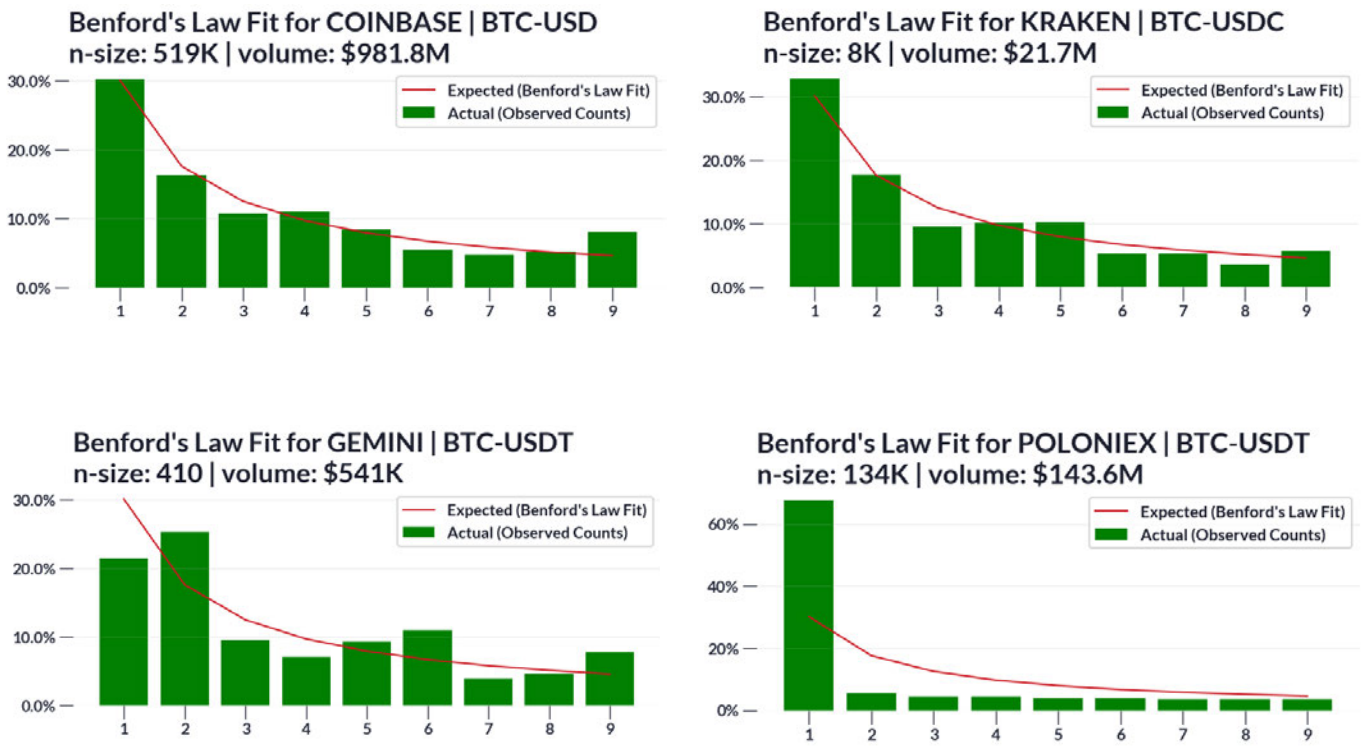
# CHANGELOG

| DATE | NOTES |
|---|---|
| July 31, 2025 | Trusted Exchange Framework V2.3 released. |
| April 30, 2024 | Trusted Exchange Framework V2.2 released. |
| March 6, 2024 | Added a new rule to remove failed exchanges ahead of the next official Trusted Exchange Framework release 3 months after trading ceases. |
| October 31, 2023 | Trusted Exchange Framework V2.1 released. Revamped Transparency, Resilience & Security, Regulatory scores; added grading scale, added rate limiting score. |
| March 1, 2023 | Trusted Exchange Framework V2.0 released. |

# APPENDIX

## Figure 1. Benford's Law Fits by Exchange

Organic market activity tends to follow specific properties from Benford's Law – leading digits tend to be low and most frequent, and the frequencies decrease as the leading digit increases. Markets that deviate from this behavior fail this test. The figures below illustrate the difference between well-behaved markets (Coinbase BTC-USD, Kraken BTC-USDC, Gemini BTC-USDT) and a market that would violate this law (Poloniex BTC-USDT). The x-axis represents the leading digits while the y-axis represents the frequency of the leading digits.

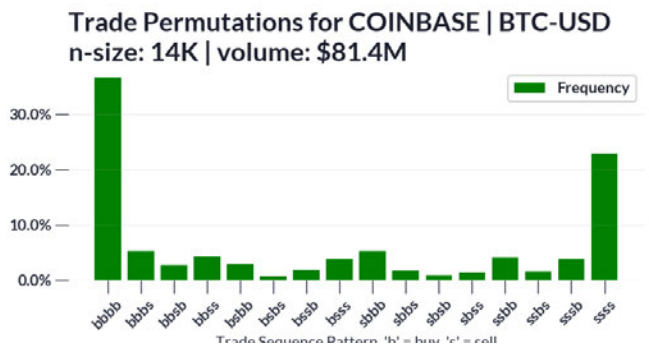


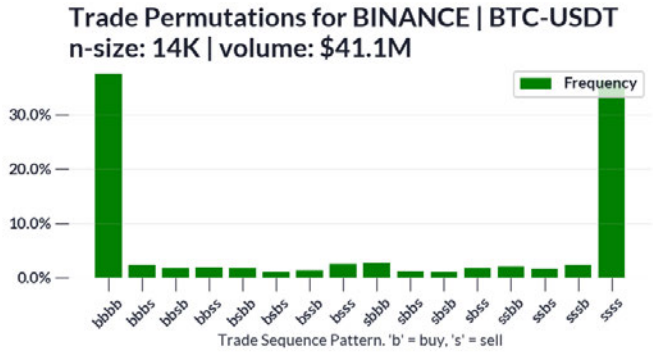
It's worth noting that although most well-behaved markets fit Benford's Law (with some tolerance to account for idiosyncratic exchange activity), there are cases where leading digits perfectly fit Benford's Law. This is the case for MEXC. It's inconclusive whether these are due to natural trading activity or manipulated to fit Benford's Law, but we will continue to tune this test in future revisions

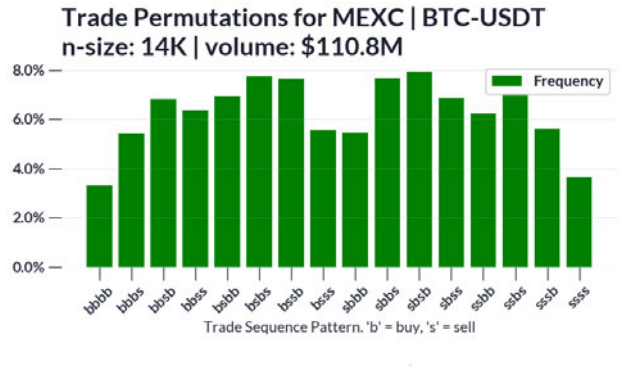


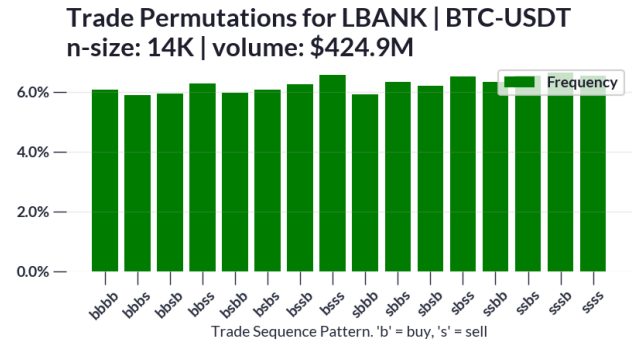

# Figure 2. Trade Permutations

The majority of well-behaved  crypto markets follow trade permutation patterns with several consecutive buy or sell orders. Conversely, markets that often exhibit heavy wash trading will have a uniform distribution for their trade permutations. Below we illustrate the difference between a well-behaved market (coinbase-btc-usdt) and a market which has trade permutation patterns that are usually indicative of heavy wash trading (mexc-btc-usdc).



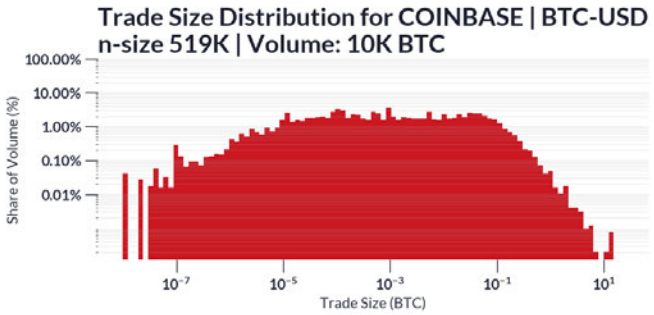



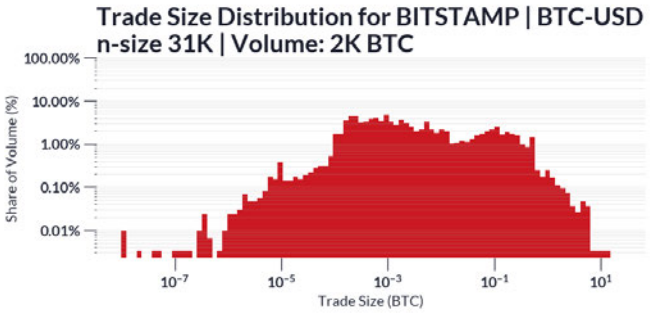



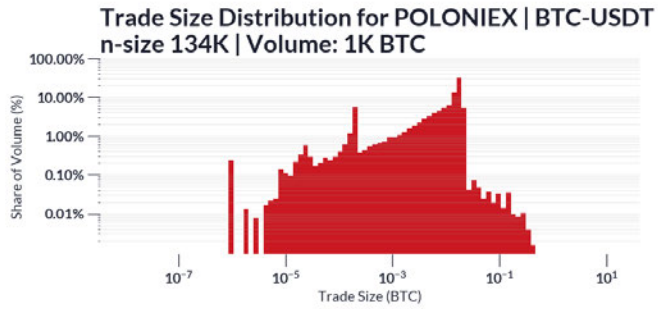
# Figure 3. Trade Size Distributions

Organic trading activity tends to cause the distribution of trade sizes to follow a log-normal distribution. Markets with significant levels of inorganic or spurious trades deviate from this distribution significantly, as measured by the mean absolute error of the residuals between the log of the observed trade sizes and a log of the log-normal probability distribution. Well-behaved markets tend to have residuals fairly flat, while anomalous markets have a high magnitude of residuals, sometimes linearly trending at higher trade sizes. For illustrative purposes, we plotted the raw trade size histogram between some well-behaved market (coinbase-btc-usd-spot, bitstamp-btc-usd-spot) and a market that does not fit this distribution (poloniex-btc-usdt-spot).

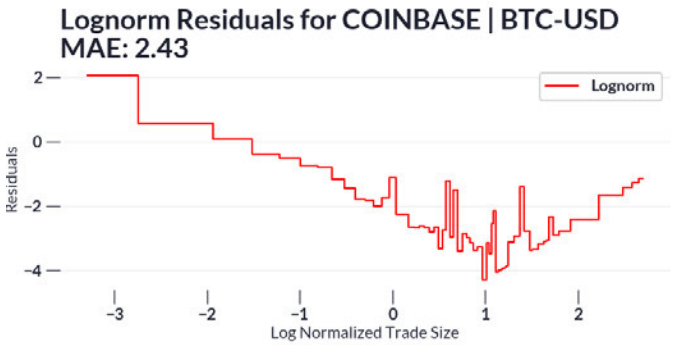
### Trade Sizes

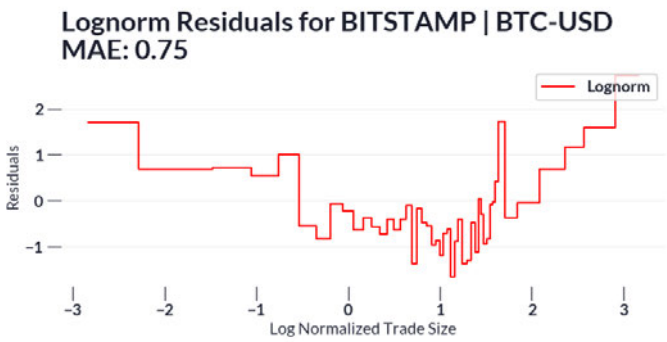


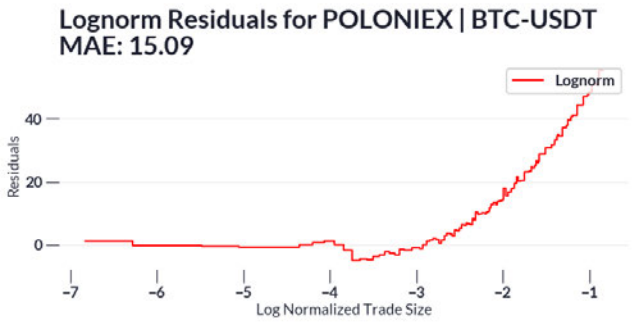



## Residuals


Lognorm Residuals for COINBASE | BTC-USD
MAE: 2.43
Lognorm
Residuals
Log Normalized Trade Size

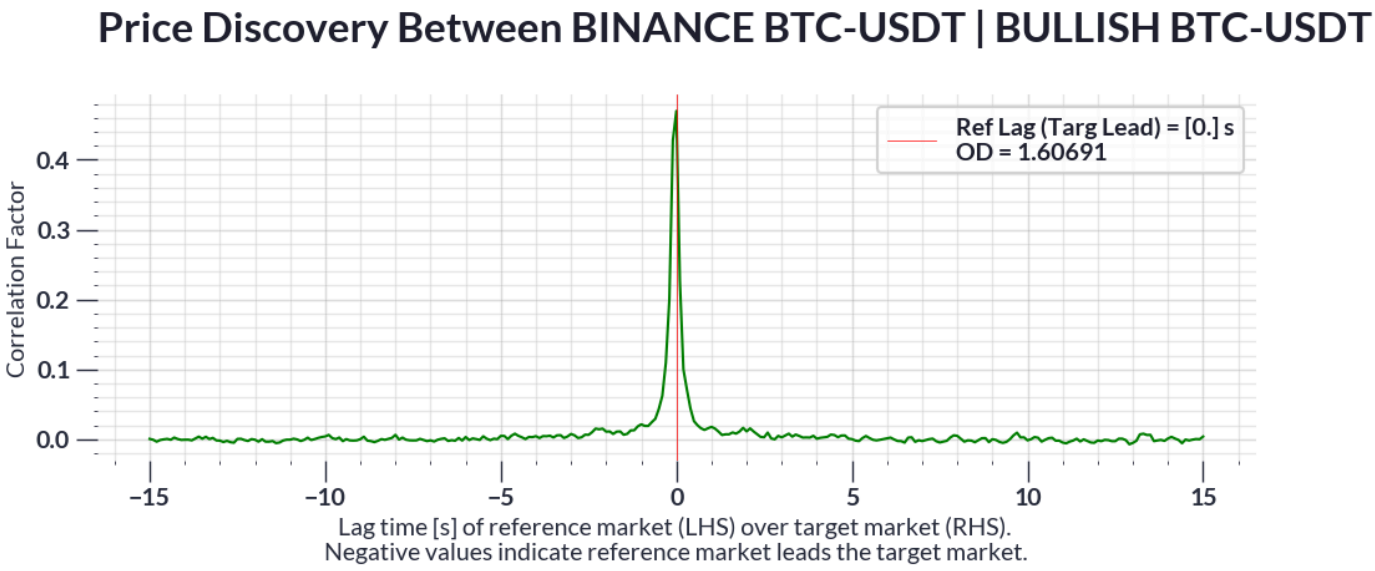

Lognorm Residuals for BITSTAMP | BTC-USD
MAE: 0.75
Lognorm
Residuals
Log Normalized Trade Size

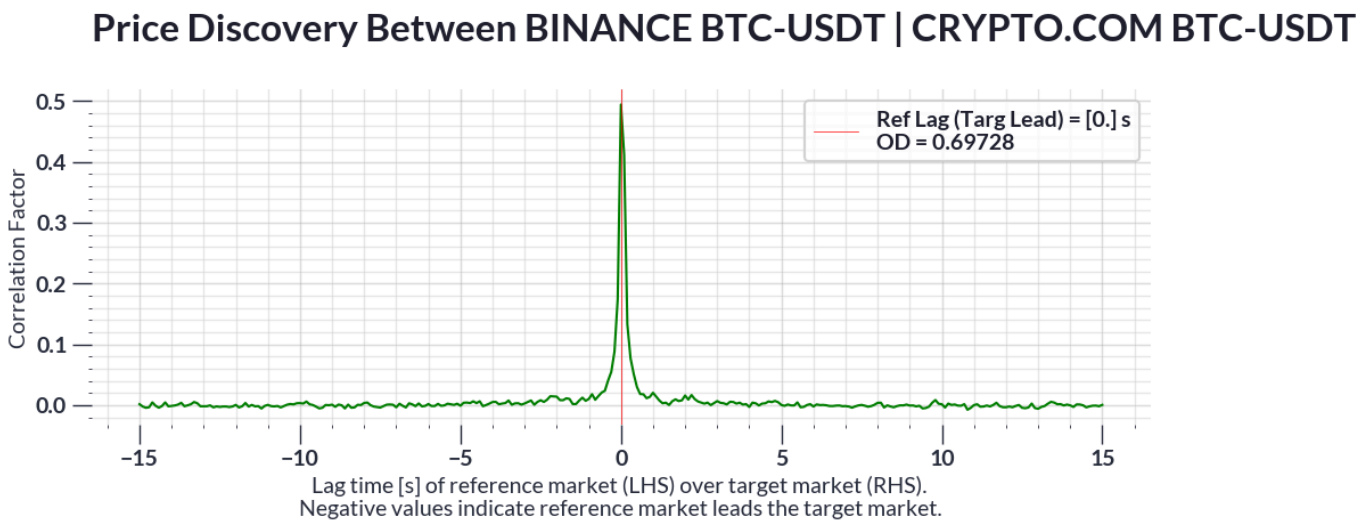

Lognorm Residuals for POLONIEX | BTC-USDT
MAE: 15.09
Lognorm
Residuals
Log Normalized Trade Size

# Figure 4. Price Discovery

Lag curves are a crucial component of the Hayashi-Yoshida methodology used to quantify Price Discovery. In this method, returns of an asset on one market are shifted forwards and backwards in time relative to the returns of the same asset on a reference market. Estimating the correlation between these returns, as a function of time displacement, allows for analysts to observe how much a given market should be lagged for its prices to most strongly correlate with the adjacent reference market.

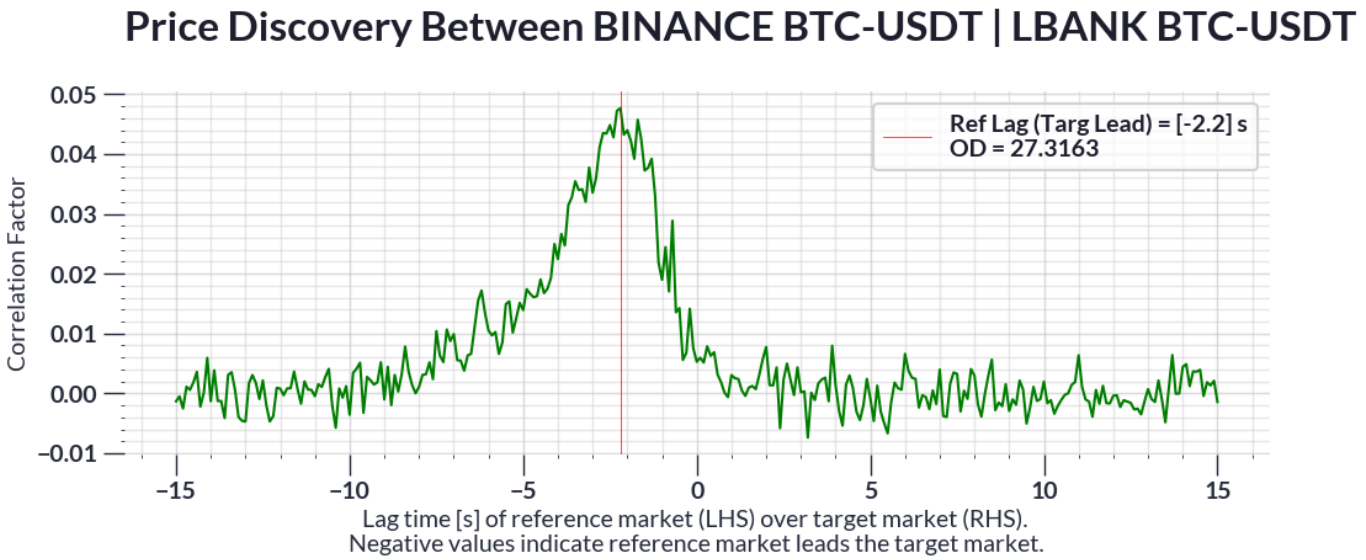
Consider the lag curves on a random sample during 2025 Q2 for the example BTC-USDT markets below, taking Binance as a reference exchange. Bitcoin prices between Binance and Bullish as well as Binance and Crypto.com have a max correlation at t = 0s, meaning the prices are in sync. These markets are generally "well-behaved".



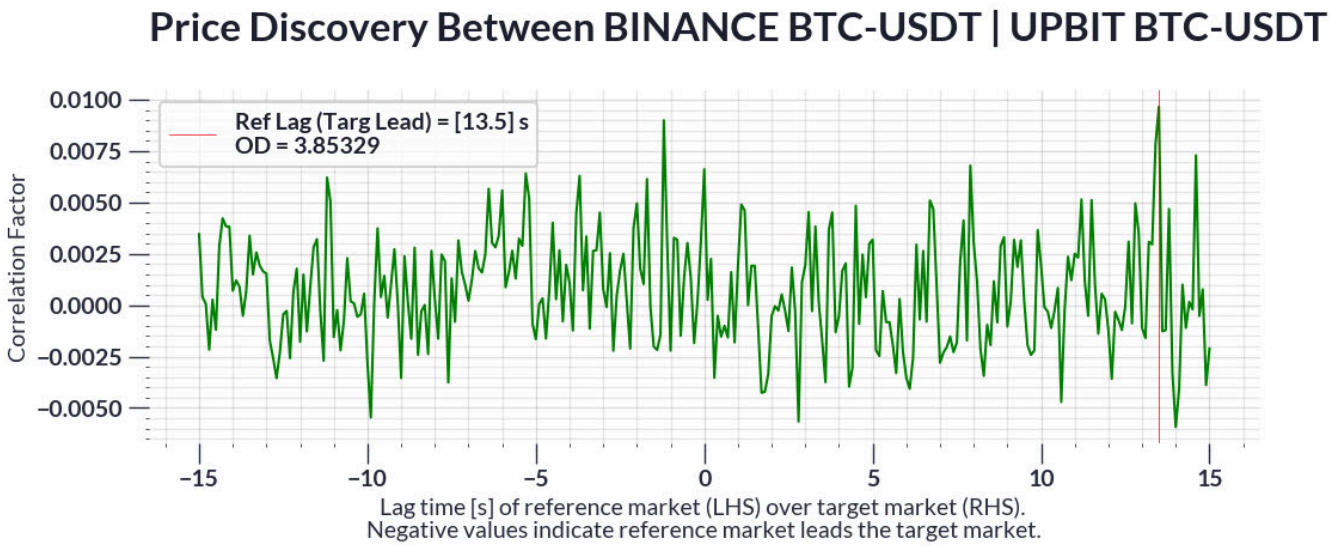



There are two common cases for markets to be ill-behaved: one where the target market clearly lags the reference market, and the other where there is no clear relationship between the prices of the target and reference markets.

An example of the first case is LBank. LBank's prices have its highest correlation 2.2s after Binance's, meaning LBank's prices lag Binance's by 2.2s on average. This is a relatively high lag time across crypto exchanges, as the median lag time was found to be less than 0.1s.




An example of the second case is Upbit. Upbit's correlation has a maximum at 13.5s, but note that the presence of several other peaks make this an imprecise estimate. Consequently, this leads to common price dislocations between the target market and the rest of the market.

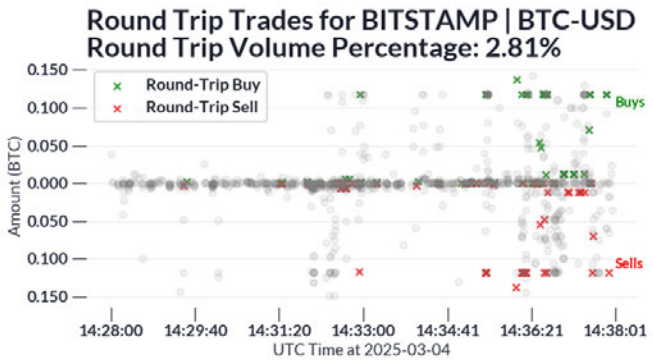


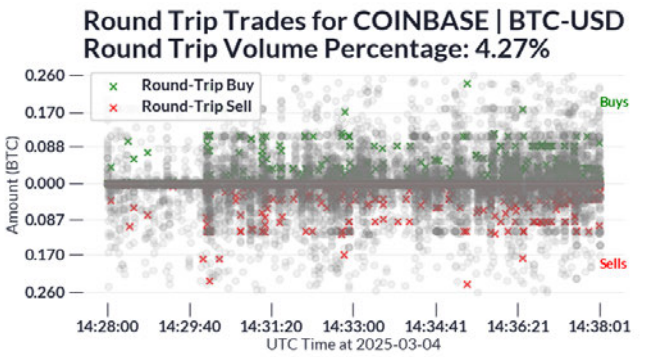
# Figure 5. Round Trip Trades

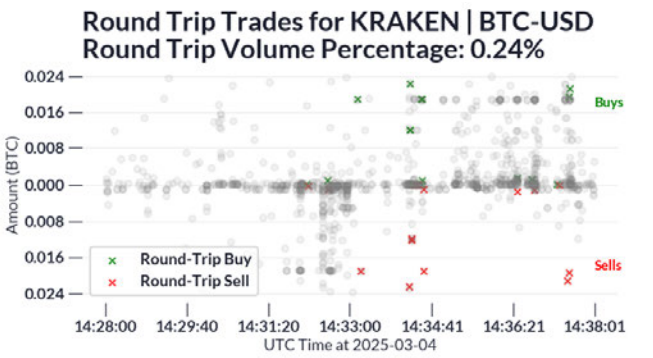
Below are a set of executed trades by market in a randomly sampled time period. Organic trade activity tends to be randomly dispersed across sides and various amounts but follow the market trend, resulting in few round trip trades. If the trades executed follow a pattern, such as the same amounts being consistently bought and sold back and forth, then it suggests a strong presence of non-economic activity executed by automated trading bots.

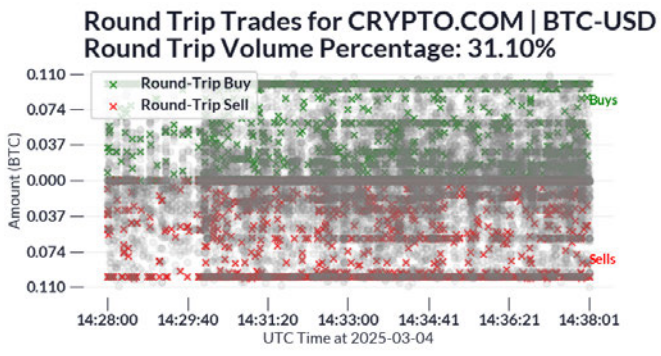
On average, we observed most exchanges have 1-8% of their trading volume come from trades labeled as a "round trip". The exchanges that had the highest proportion of volume come from round trips hovered around 15-30%.

### BTC-USD

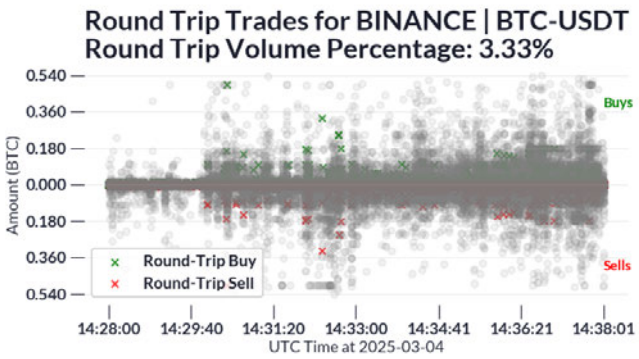


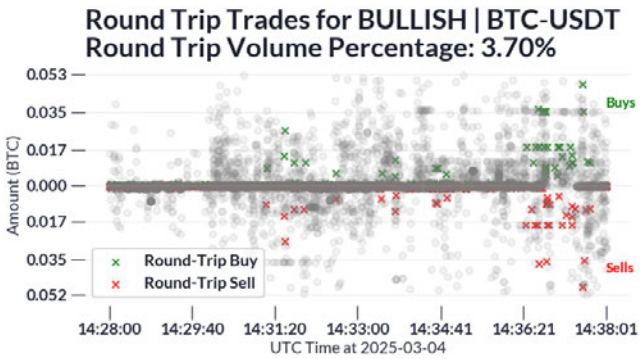



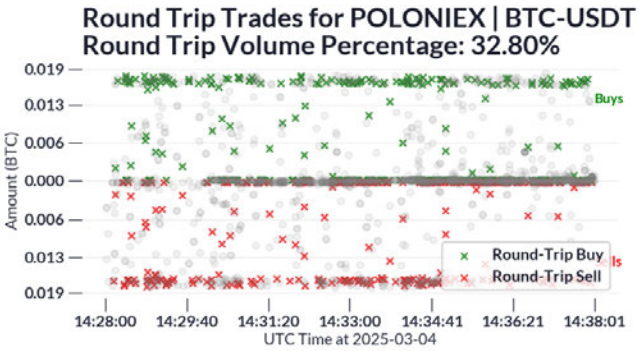



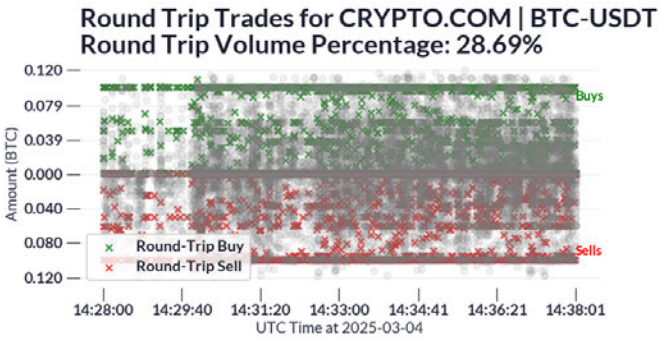



## BTC-USDT


Round Trip Trades for BINANCE | BTC-USDT
Round Trip Volume Percentage: 3.33%
0.540
0.360
0.180
0.000
0.180
0.360
0.540
Amount (BTC)
Buys
Sells
Round-Trip Buy
Round-Trip Sell
14:28:00
14:29:40
14:31:20
14:33:00
14:34:41
14:36:21
14:38:01
UTC Time at 2025-03-04


Round Trip Trades for BULLISH | BTC-USDT
Round Trip Volume Percentage: 3.70%
0.053
0.035
0.017
0.000
0.017
0.035
0.052
Amount (BTC)
Buys
Sells
Round-Trip Buy
Round-Trip Sell
14:28:00
14:29:40
14:31:20
14:33:00
14:34:41
14:36:21
14:38:01
UTC Time at 2025-03-04


Round Trip Trades for POLONIEX | BTC-USDT
Round Trip Volume Percentage: 32.80%
0.019
0.013
0.006
0.000
0.006
0.013
0.019
Amount (BTC)
Buys
Round-Trip Buy
Round-Trip Sell
14:28:00
14:29:40
14:31:20
14:33:00
14:34:41
14:36:21
14:38:01
UTC Time at 2025-03-04


Round Trip Trades for CRYPTO.COM | BTC-USDT
Round Trip Volume Percentage: 28.69%
0.120
0.079
0.039
0.000
0.040
0.080
0.120
Amount (BTC)
Buys
Sells
Round-Trip Buy
Round-Trip Sell
14:28:00
14:29:40
14:31:20
14:33:00
14:34:41
14:36:21
14:38:01
UTC Time at 2025-03-04

# WORKS CITED


1. Carter, Nic. “The Status of Proof of Reserve as of Year End 2022.” Medium, 29 December 2022, https://medium.com/@nic__carter/the-status-of-proof-of-reserve-as-of-year-end-2022-48120159377c. Accessed 30 January 2023.

2. Bitwise Asset Management. “Economic and Non-Economic Trading In Bitcoin: Exploring the Real Spot Market For The World's First Digital Commodity.” SEC.gov, 24 May 2019, https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5574233-185408.pdf. Accessed 27 January 2023.

3. “More Than Half of All Bitcoin Trades Are Fake.” Forbes, 10 August 2022, https://www.forbes.com/sites/javierpaz/2022/08/26/more-than-half-of-all-bitcoin-trades-are-fake/?sh=28d5007d6681. Accessed 30 January 2023.

4. Cong, Lin William, et al. “Crypto Wash Trading.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3530220.

5. Digital Asset Research. “Clean Prices: A Required Step for Crypto to Mature.” Medium, 25 October 2019, https://medium.com/digitalassetresearch/clean-prices-a-required-step-for-crypto-to-mature-bbb4156b76d8. Accessed 30 January 2023.

6. Robertson, Kevin, and Jiani Zhang. “Suitable Price Discovery Measurement of Bitcoin Spot and Futures Markets.” 12 January 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4012165.

7. Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi ( 2010 ). "The Worldwide Governance Indicators: Methodology and Analytical Issues". World Bank Policy Research Working Paper No. 5430 ( papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130 ).

# TRUSTED EXCHANGE FRAMEWORK 2.3

By **Victor Ramirez**, **Jo Suddreth**, and the **Coin Metrics Team**

COINMETRICS

To view more from Coin Metrics Research go to
**coinmetrics.io/pubs** or subscribe to **State of the Network**