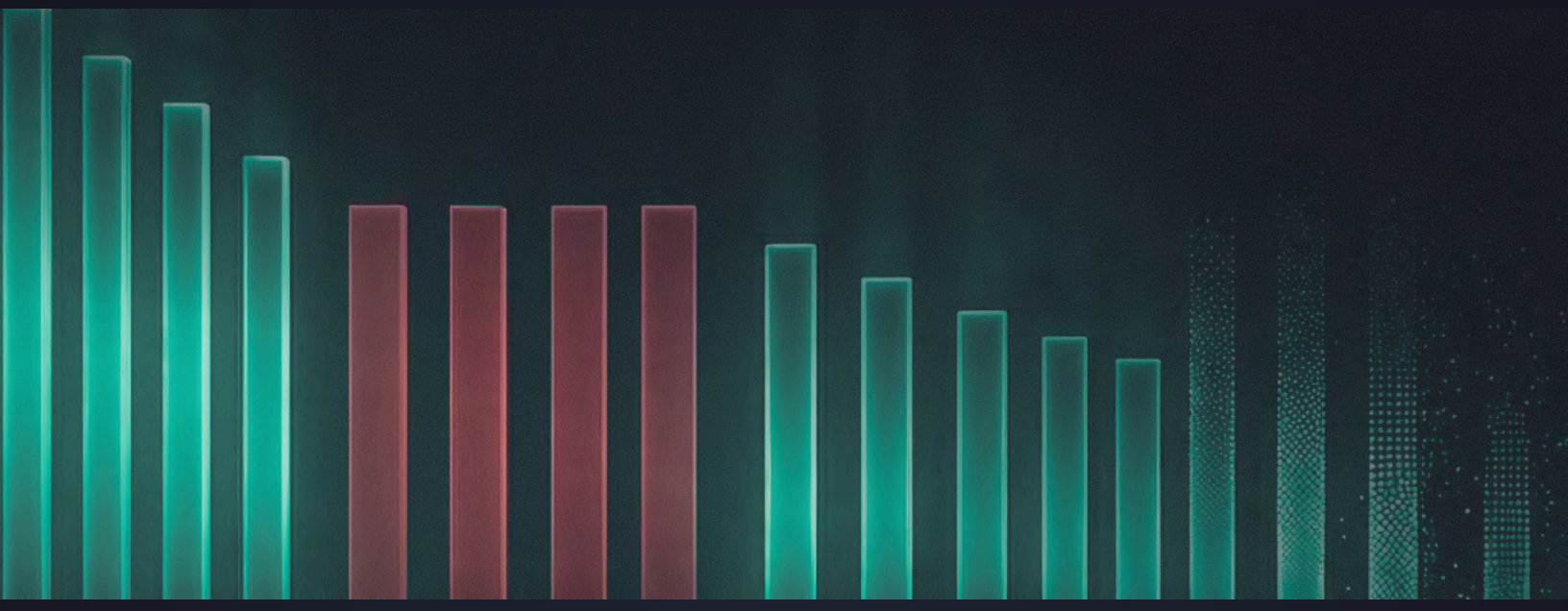
TALOS

RESEARCH

Talos Exchange Scorecard Methodology

By Victor Ramirez, Jo Suddreth and the Talos Data team



July 2026

Talos Exchange Scorecard Methodology [1]

Version 2.4

Updated July 2026

By Victor Ramirez, Jo Suddreth, and the Talos Data team

[1] Formerly Coin Metrics Trusted Exchange Framework

Table of Contents

Introduction

The Talos Exchange Scorecard (TES) is a systematic framework for assessing digital asset exchanges across their core qualities. An exchange may serve multiple uses — as a liquid trading venue, as a source for pricing and market activity, as a platform to custody assets, and as a programmable interface for automated trading and data collection. But objectively measuring quality across these dimensions is inherently challenging. Without a principled methodology, assessments of 'quality' tend to be subjective, use-case dependent, and prone to bias.

The TES addresses this by translating abstract notions of exchange quality into six measurable pillars: Data Quality, Liquidity, Regulatory Compliance, Transparency, Resilience & Security, and API Quality. Each pillar is assessed through a combination of quantitative tests and qualitative indicators, applied uniformly across each type of exchange. These scores are then translated to an interpretive grading scale to give an intuitive sense of relative quality.

The primary data source is Talos's own market data collection built and maintained across 50+ exchanges over eight years. This is supplemented by Talos's operational experience building and maintaining market data pipelines, a review of academic and industry literature on market structure, and public disclosures from exchanges.

The overall scoring is designed to be a holistic guide and can be deconstructed for different use-cases. Scores are published at the category level so users can apply their own weighting depending on context. For example, institutional trading firms looking for the most efficient trade execution may overweight Liquidity, while data intensive applications may overweight Data Quality. Talos uses the Exchange Scorecard to select constituent exchanges for our prices, indexes, and metrics.

The full results can be found here.

Methodology

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	Exchanges in this tier meet criteria thresholds across most or all of the factors assessed. Exchanges in this tier often have the highest quality data and efficient markets relative to peers.
B- to B+	Exchanges in this tier generally meet the threshold criteria across most of the factors assessed. Minor penalties across a few pillars keep these exchanges from being graded in the A tier.
C- to C+	Exchanges in this tier 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), or vice versa.
D	Exchanges in this tier score below threshold across most of the factors assessed. Exchanges in this tier may 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:

Weights	Data Quality	Liquidity	Transparency	Resilience	Regulatory Compliance	API Quality	Total
Spot	15%	15%	15%	15%	30%	10%	100%
Futures-Only	0%	20%	15%	15%	40%	10%	100%
DeFi	20%	20%	15%	35%	0%	10%	100%

Table 3. Weights.

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 quality.

Categories

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

Data Quality

The Data Quality score assesses the confidence that an exchange's reported data is accurate. Exchanges are the primary source for market data such as volume, yet crypto exchanges have historically created "fake" or non-economic volume to attract users to its platform.

Bitwise estimated in 2019 that 95% of volume reported by exchanges are fake by observing anomalous trade patterns. [1] Forbes estimated in 2022 that about 51% of bitcoin trading volume is fake by weighting the likelihood of fake volume given regulatory and web usage patterns. [2] 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 one crypto exchange accurately reports $100M of volume in one day and another 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, and identifying round trip trades.

Trusted Volume

A score of **B** or above on the Data Quality category qualifies an exchange to be part of our "trusted volume" metric, which is our best estimate of spot trading volume across all exchanges. 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 the 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 compares to a known or expected distribution. The scoring method differs by test, reflecting the nature of each signal:

Absolute threshold scoring (Benford's Law): The mean absolute deviation (MAD) of observed leading digit frequencies from the Benford's Law distribution is compared to thresholds derived from a reference population of organic crypto markets. Thresholds are calibrated to empirical percentiles of that population, so a failing exchange fails regardless of how other exchanges in the current universe perform. This replaces the prior peer-relative z-score approach.

Shape rubric scoring (Trade Sizes): The log-scale histogram of trade sizes is classified into categorical shape profiles. Passes and failures are assigned based on the error mode (shape combination) detected, with an adjudicating signal — the USD round-tooth excess — used to resolve suspect cases. This replaces the prior z-score of log-normal residuals.

Null-model MAE scoring (Trade Permutations): The mean absolute error of observed buy/sell permutation frequencies against a theoretically derived null model is computed, volume-weighted by market, and scored via z-score across the exchange universe. The null model was updated in V2.4 from a uniform 1/16 baseline to an i.i.d. model conditioned on each market's observed buy probability (see subcategory description below).

Z-score scoring (Volume Correlation): Volume correlation is computed over a restricted pair universe, volume-weighted, and scored against a z-score threshold relative to peers.

Sigmoid scoring (Round Trip Trades): A continuous score from 0 to 1 is assigned based on the proportion of volume attributable to round-trip trades, using a sigmoid function centered near the empirical threshold for concerning activity.

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 concern about data quality but does not rule out the possibility that most volume is organic, given 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 general confidence in data quality but does not imply that 100% of trading volume is organic. 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	Mean absolute deviation (MAD) of observed vs. Benford's Law distribution, calibrated against a reference population of organic crypto markets.	Pass (1.0) if MAD ≤ ~60th percentile of organic markets. Partial (0.5) if MAD ≤ ~80th percentile. Fail (0.0) if above. Also fails (0.0) if the fit is perfect: MAD below the 5th percentile of the organic population, or chi-squared below the 1st percentile.
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 of observed buy/sell flag permutation frequencies against an i.i.d. null derived from each	Pass if z-score > -1. Partial credit if -1 >= z-score > -2 . Fail if z-score < -2.
Trade Size Distribution	Shape classification of the log-scale trade size histogram. Suspect shapes (flat & decreasing, or bell & flat) are adjudicated by the USD round-tooth excess.	Hard fail (0.0) if distribution is increasing. Suspect shapes: Pass (1.0) if USD round-tooth excess ≥ 2.0; partial (0.5) if ≥ 1.3; fail (0.0) otherwise. Non-suspect shapes: Pass (1.0).
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 0.2s to 1s. Fail as the price discovery approaches and exceeds 1s.

Data Quality Test	Computation	Threshold
Price Anomalies	Linear ramp on anomaly frequency (proportion of timestamps where the exchange's price falls outside ±2 standard deviations of the cross-market mean).	Pass (1.0) if ≤1%. Fail (0.0) if ≥10%. Linear between 1% and 10%; midpoint 5.5% → 0.5.
Round Trip Trades	Volume-weighted proportion of round-trip trade pairs. Detection window is adaptive: calibrated to each market's prevailing half-spread rather than a fixed constant, producing consistent sensitivity across market structures.	Sigmoid function. Midpoint = 10%. Pass if < 10%. Score declines continuously as the proportion approaches and 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 amounts in base units follows Benford's Law. Benford's Law is the observation that in naturally occurring numerical data, leading digits tend to be small, with the frequency of each digit decreasing as the digit increases. 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 Appendix A1.

Volume Correlations

An assessment of how correlated an exchange's volume is relative to its peers. Exchanges with organic volume 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. Exchanges with inorganic volume deviate from this pattern in idiosyncratic ways, resulting in lower correlation with the broader exchange set [1]

Trade Permutations

An assessment of the legitimacy of the buy/sell flag distribution across trade sequences. Exchanges that fabricate volume are known to show an approximately uniform distribution of trade side permutations, likely due to wash trading or other non-economic activity generated by an artificial process. Legitimate market activity tends to skew toward consecutive buy or sell trades driven by informed traders who take liquidity in response to material new information. [5] See Appendix A2.

Expected permutation frequencies are derived from each market's observed buy probability under an i.i.d. null model, rather than a uniform baseline. This isolates order-flow autocorrelation, the actual signal for wash trading, from directional buy/sell imbalance, which is a normal property of trending markets.

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 sizes, which can be ascribed to the presence of retail and institutional participants and the effect of liquidity constraints on order sizing. Inorganic trading activity generated from artificial processes often exhibits a different distribution, including unusually high concentrations at specific size levels or a tail that decays atypically.

The log-scale trade size histogram is classified into shape profiles, with scores assigned based on the combination of shape modes detected. Exchanges that trip a suspect shape classification are adjudicated by the local density of trades at round-dollar lot sizes relative to their surrounding neighborhood, or "USD round-tooth excess". Organic markets tend to show a characteristic

round-number comb from retail order sizing; artificially generated flow does not. An increasing distribution is a hard fail regardless of the round-tooth signal. See Appendix A3.

Round Trip Trades

The estimated proportion of volume from "round-trip" trades. A pair of round-trip trades is defined as two trades with nearly identical prices and amounts executed on opposite sides (buy and sell) in close proximity in time and sequence. These trades inflate a market's volume with activity that involves zero market risk. Although not every trade matching this definition is necessarily market manipulation, a high frequency of such trades with nearly identical footprints strongly suggests non-economic automated activity.

The detection window for matching round-trip pairs is calibrated to each market's prevailing half-spread, producing consistent sensitivity across market structures with different tick sizes and liquidity profiles. See Appendix A5.

Liquidity

The Liquidity score assesses how cheaply and reliably an exchange's markets can be traded. Where Data Quality tests whether reported activity is real, Liquidity tests the execution quality behind it: the cost of crossing the book, the size the book can absorb, and whether resting liquidity is consistent with reported volume. The pillar draws on order-book snapshots, derivatives open interest, on-chain pool balances, and the lead/lag relationships between exchange prices.

For centralized spot markets, the order-book features – spread, depth, and slippage – are sampled continuously and time-averaged over the analysis window, measured on the BTC-USD, BTC-USDT, and BTC-USDC markets and rolled up to the exchange as a volume-weighted average. Open Interest extends the pillar to derivatives venues and Onchain Liquidity to decentralized venues, each of which lacks a spot order book to measure directly. The pillar score is the equal-weighted mean of the features that apply to a venue, so a venue is judged only on the measures relevant to it.

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 Appendix A4.

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.

Orderbook Depth

Orderbook depth is a measure of the resting liquidity available in an exchange's order book, calculated as the total USD value of bids and asks within ±0.1% of the mid price. Deeper books absorb larger orders with less price impact, so higher depth scores more favorably. Depth is sampled through the analysis window, time-averaged, and rolled up to the exchange as a volume-weighted average across its BTC markets. Exchanges are scored on a continuous log-relative curve. The ±0.1% band is used in place of a wider band because several venues publish only a limited number of order-book levels, which leaves the wider bands unpopulated for those exchanges. See Appendix A6.

Slippage

Slippage is an estimate of the cost of executing a market order against the order book, measured as the percentage difference between the volume-weighted fill price of a $100,000 order and the mid price, averaged across the buy and sell sides and expressed in basis points. Lower slippage indicates a book that can be traded at size without moving the price, so it scores more favorably. Slippage is sampled through the analysis window, time-averaged, and rolled up to the exchange as a volume-weighted average across its BTC markets, then mapped to a score on a piecewise scale where tighter execution costs earn higher scores. See Appendix A7.

Spread

Spread is the gap between the best bid and the best ask, expressed as a percentage of the mid price and converted to basis points. The spread is the immediate cost of crossing the book for a small order, so tighter spreads score more favorably. It is sampled hourly through the analysis window, time-averaged, and rolled up to the exchange as a volume-weighted average across its BTC markets, then mapped to a score on a piecewise scale where tighter spreads earn higher scores. See Appendix A8.

Open Interest

Open interest is the derivatives analogue of order-book depth: a venue's total futures open interest in USD, time-averaged over the analysis window. Deep, persistent open interest reflects the capital committed to a venue's contracts and the size it can support, so higher open interest scores more favorably. Open interest is taken from the exchange-level aggregate reported by Coin Metrics and covers all of a venue's futures rather than a single contract. Every venue that reports open interest is scored, including derivatives-native venues with no spot markets. Scores use the same continuous log-relative curve as Orderbook Depth: the venue with the most open interest scores 1.0, and a venue's score falls by 0.2 for each order of magnitude below that maximum. See Appendix A9.

Onchain Liquidity

The decentralized-exchange analogue of order-book depth. An AMM has no order book, so executable liquidity is measured by the value locked in its pools: the total USD held across a venue's WBTC and WETH pools quoted in USDC and USDT. For each pool, both token balances

are read on-chain and valued at the current reference rate, then summed across all of the venue's pools and across a protocol's deployments on different chains (for example, Uniswap v3 on Ethereum and Base are combined into one entity). Deeper pools absorb larger swaps with less price impact, so higher value locked scores more favorably. Unlike the order-book features, this is a current snapshot of pool balances rather than a window average. Scores use the same continuous log-relative curve as Orderbook Depth.

Depth to Volume

A check on whether an exchange's reported volume is backed by real order-book liquidity. For each market, the ±0.1% bid and ask depth in USD is divided by average daily volume to produce a support ratio, which is volume-weighted to the exchange and scored on a one-sided scale that penalizes only thin backing. Volume and the depth supporting it are difficult to fabricate at the same time, so volume far in excess of its supporting depth is a signal of wash trading. This measure is spot-only: open interest is a stock of outstanding positions rather than executable book depth, so futures venues are assessed separately. See Appendix A10.

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 meets established security standards and has avoided security incidents. 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. 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 of an exchange's ability to meet regulatory requirements across multiple jurisdictions, based on the strength of its licensing, the integrity of its public compliance footprint, and its enforcement history. The score is a percent score out of 100 built from three weighted factors: Licensing & Authorisation (65%), Footprint Integrity (15%), and Enforcement History (20%).

Disclaimer: The Regulatory Compliance category reflects Talos's assessment of each exchange's licensing status, registrations, and regulatory history based on publicly available information, including regulatory filings, enforcement actions, license registries, and disclosures made by the exchanges themselves or by relevant authorities, as of the date of publication. Talos does not independently verify the accuracy or completeness of any regulatory claim made by an exchange or third party, and scores may not reflect regulatory developments, actions, or license changes that occur after the applicable review period or that are not publicly disclosed.

This category does not constitute a legal opinion or determination as to whether any exchange is operating in compliance with applicable law in any jurisdiction, and should not be relied upon as such. It does not assess suitability, adequacy, or legality of any exchange's operations for any particular user or jurisdiction, nor does it consider every regulatory regime, jurisdiction, or license type that may apply to a given exchange. Absence of a noted regulatory issue does not imply regulatory approval or endorsement by Talos or any authority.

Subcategories

Licensing & Authorisation

This factor scores the strength of the single best license the exchange holds across a fixed basket of 12 jurisdictions: United States, United Kingdom, European Union, Japan, Singapore, Hong Kong, Switzerland, Australia, United Arab Emirates, Canada, South Korea, and Gibraltar. Only the best license counts. Holding additional, lower-tier licenses elsewhere does not add to the score.

Each license is classified by regulatory depth into one of four classes:

- Class 1: full market or prudential authorisation (e.g. a national trust charter, a CFTC-registered DCM/DCO, FCA Part 4A authorisation, MAS MPI, an ASIC AFSL, FINMA banking license, SFC Type 1/7).
- Class 2: comprehensive crypto-specific or payments licensing (e.g. NYDFS BitLicense, MiCA CASP authorisation, VARA full VASP license, Korea FSC/FIU VASP license, MAS Recognised Market Operator).
- Class 3: registration or AML-only regimes with no prudential or market-conduct authority (e.g. FinCEN MSB registration, AUSTRAC DCE registration, FINTRAC MSB registration).
- Class 4: No regulatory licensing, authorization, or registration

Jurisdiction	WGI Reg. Quality (Percentile)	Class 1	Class 2	Class 3
United States	91.04	CFTC DCM/DCO; national trust charter (OCC)	NYDFS BitLicense; state MTL network	FinCEN MSB registration only
United Kingdom	93.87	FCA Part 4A full authorisation	FCA cryptoasset registration (AML)	—
European Union	83.02	MiCA CASP full authorisation; MiFID II investment firm	EMI/PI licence (payments-only)	—
Japan	89.62	FSA crypto-asset exchange registration	—	—

Jurisdiction	WGI Reg. Quality (Percentile)	Class 1	Class 2	Class 3
Singapore	100.00	MAS Major Payment Institution (DPT)	MAS Recognised Market Operator/RCH	—
Hong Kong	96.70	SFC Type 1/7 licence; VATP licence	"Deemed licensed" transitional status	TCSP licence only
Switzerland	97.17	FINMA banking licence	FIINMA-supervised SRO membership (VQF, PolyReg)	—
Australia	96.23	ASIC Australian Financial Services Licence (AFSL)	—	AUSTRAC DCE registration (AML only)
UAE	86.79	VARA full VASP licence; DFSA/ADGM FSRA licence	SCA licence (retail, UAE-wide)	—
Canada	91.51	OSC Restricted Dealer registration	—	FINTRAC MSB registration only
South Korea	82.55	FSC/FIU VASP licence (full scope)	—	—
Gibraltar	93.87	—	GFSC DLT Provider Licence	—

Table 5a. Jurisdiction weights and licence classes. *World Bank, Worldwide Governance Indicators, Regulatory Quality (2024 estimate), Data360 (data360.worldbank.org/en/dataset/WB_WGI). Percentile ranks derived from the raw regulatory quality estimate across 213 economies.

Footprint Integrity

This factor deducts for indicators that an exchange's public-facing compliance posture is inconsistent with holding a strong license — regardless of the license itself. Only flags tied to one of the 12 basket jurisdictions apply. The score floors at zero.

Flag	Deduction (%)	Cap
No mandatory KYC on the platform	–4	n/a (single flag)
Named on a basket-jurisdiction regulator's public warning / unregistered-entity list	–5 per jurisdiction	-10 total
Licence suspended or revoked in a basket jurisdiction	–3 per jurisdiction	-6 total
Corporate domicile in a FATF grey-listed jurisdiction	-4	n/a
Corporate domicile in a FATF black-listed jurisdiction	-15	n/a

Table 5b. Footprint Integrity deductions.

Enforcement History

This factor deducts for regulatory or criminal enforcement actions from the preceding five years, tiered by severity. Where a single course of conduct draws action from multiple agencies (e.g. a coordinated DOJ/FinCEN/OFAC settlement), it is counted once, at its highest applicable tier. Custody and security breaches, including exchange hacks, are not assessed here and are instead covered under Resilience & Security.

Tier	Deduction (%)	Examples
Tier 1	–20	Criminal conviction; sanctions evasion; fraud or market-manipulation finding; licence revocation for cause
Tier 2	–10	AML programme breakdown; unlicensed/unregistered operation
Tier 3	–4	Disclosure violations; deceptive marketing

Table 5c. Enforcement History tiers.

Deductions decay with time: actions resolved within the past year apply in full, actions resolved one to three years ago apply at half weight, and actions older than three years no longer count. The score floors at zero.

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.

Revisions

Changelog

DATE	NOTES
July 30, 2026	V2.4 released.
August 19, 2025	API Quality scores revised due to data collection error.
July 31, 2025	V2.3 released.
April 30, 2024	V2.2 released.
March 6, 2024	Added a new rule to remove failed exchanges ahead of the next official release 3 months after trading ceases.
October 31, 2023	V2.1 released. Revamped Transparency, Resilience & Security, Regulatory scores; added grading scale, added rate limiting score.
March 1, 2023	V2.0 released.

Release Notes

2.4 (July 2026)

- **Renamed to "Talos Exchange Scorecard"** from "Trusted Exchange Framework".
- **Exchange universe expanded.** Added Balancer V2, Bitget, Coinbase Derivatives, Curve, Hyperliquid, Sushiswap V1, and Uniswap V1, V2, V3, V4.
- **Introduced a new category: Liquidity.**
    - The Liquidity score factors execution quality directly from order-book data: Spread, Orderbook Depth, Slippage, and Depth-to-Volume, as well as Open Interest from futures, and On-Chain Liquidity for DEXes. Price Discovery and Price Anomalies have been moved from Data Quality into this pillar.
        - **Why we made this change**: These tests measure how efficiently a market prices and executes trades. Separating them lets Data Quality serve a single, focused purpose: wash-trading detection.
- **Regulatory Compliance:**
    - Replaced the active-license-count methodology with a weighted composite score built from three factors: Licensing & Authorisation, Footprint Integrity, and Enforcement History.
        - **Why we made this change:** The new composite better captures the strength and quality of an exchange's regulatory standing by accounting for regulatory circumstances beyond raw license count. Licensing & Authorisation weighs the caliber of an exchange's best license, not the count, so a single rigorously supervised market authorization is more representative than having several less rigorous registrations. Footprint Integrity looks past the license itself to how an exchange behaves elsewhere, surfacing red flags a license alone wouldn't catch. Enforcement History adds a record of regulatory and criminal actions, scaled by severity and recency, that a license count could never reflect.
    - Table 5 (whitelisted license codes) has been replaced by Table 5a (jurisdiction weights and license classes) and Table 5b/5c (Footprint Integrity and Enforcement History deduction schedules). See the updated Regulatory Compliance section.

- **Recalibrated Data Quality Tests**:
    - **Benford's Law**: Scoring changed from a peer-relative z-score to absolute, crypto-calibrated MAD thresholds. A fabrication detection step was also added which creates a small penalty when a market fits Benford too well.
        - **Why we made this change**: Benford's Law can be trivially gamed because its mathematical properties are known. Enforcing a penalty when the fit is too perfect discourages wash-trading venues that exploit this property.
    - **Trade Permutations**: The null model changed from a uniform 1/16 baseline to an i.i.d. model conditioned on each market's observed buy fraction.
        - **Why we made this change**: The uniform baseline penalized directional buy/sell imbalance, which is normal in trending markets. The new null isolates autocorrelation in order flow.
    - **Trade Sizes**: The score is now a shape rubric adjudicated by whether trades cluster at round-dollar amounts.
        - **Why we made this change**: The prior round-number share could not distinguish organic retail flow from artificial flow, as both had similar aggregate round-share values. USD round-lot density can.
    - **Round Trip Trades**: The matching window is now calibrated to each market's half-spread rather than a fixed constant.
        - **Why we made this change**: A fixed window over-detected in wide-spread markets and under-detected in tight-spread ones.

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 Exchange Scorecard 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 Exchange Scorecard 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 quality.

Appendix

A1. 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 markets consistent with Benford's Law (Coinbase BTC-USD, Kraken BTC-USDC, Bybit BTC-USDT) and a market that violates this law (Poloniex BTC-USDT). The x-axis represents the leading digits while the y-axis represents the frequency of the leading digits.

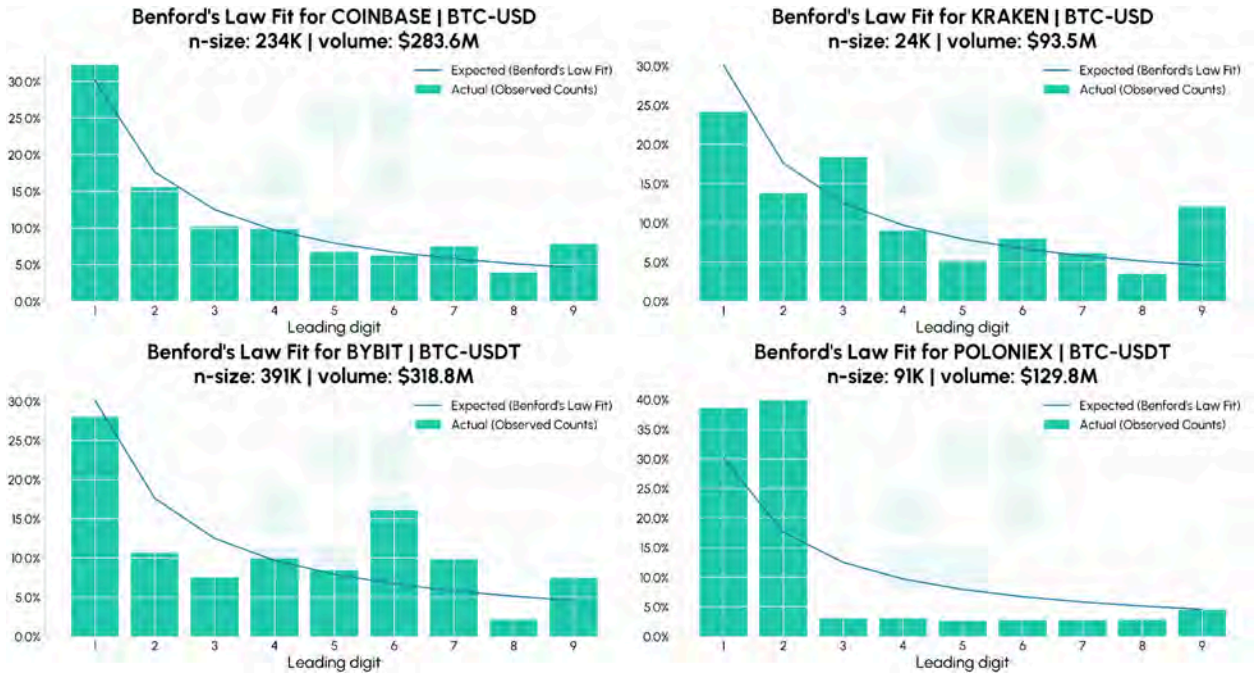


Figure 1a: Leading digit amounts for trades plotted against Benford's Law. Taken from a random sample of trades from 2026-01-01 to 2026-07-01.

Although most markets with organic volume fit Benford's Law (with some tolerance to account for idiosyncratic exchange activity), there are cases where some markets fit Benford's Law too perfectly. This was the case for MEXC in 2025.

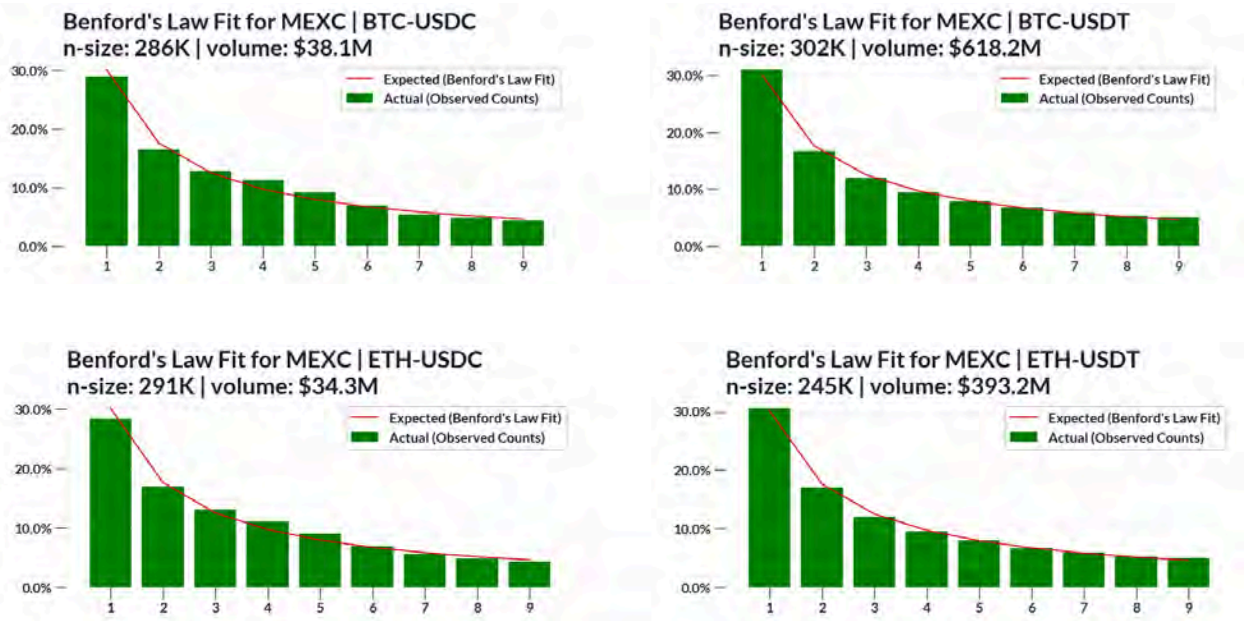


Figure 1b: Leading digit amounts for MEXC trades plotted against Benford's Law. Taken from a random sample of trades from 2025-01-01 to 2025-07-01.

In V2.4, scoring was changed from a peer-relative z-score to absolute, crypto-calibrated MAD thresholds derived from a reference population of organic crypto markets. The prior z-score approach ranked exchanges relative to each other, meaning that if the entire universe had poor Benford fits, exchanges could still receive passing scores. Absolute thresholds ensure that a failing exchange fails regardless of who else is in the universe.

A fabrication detection step was also added to flag markets that fit Benford's Law too perfectly. Under the prior approach, an exchange like MEXC in 2025, whose markets hugged the Benford distribution far more tightly than any organic market at comparable trade counts, passed the test despite this being a strong signal of artificial order generation. In V2.4, a fit is flagged as suspect if its MAD falls below the 5th percentile of the organic market reference population, or if its chi-squared statistic falls below the 1st percentile. Either condition is sufficient to fail the test.

Benford's Law Tests (2026 H1, v2.4)

MEXC

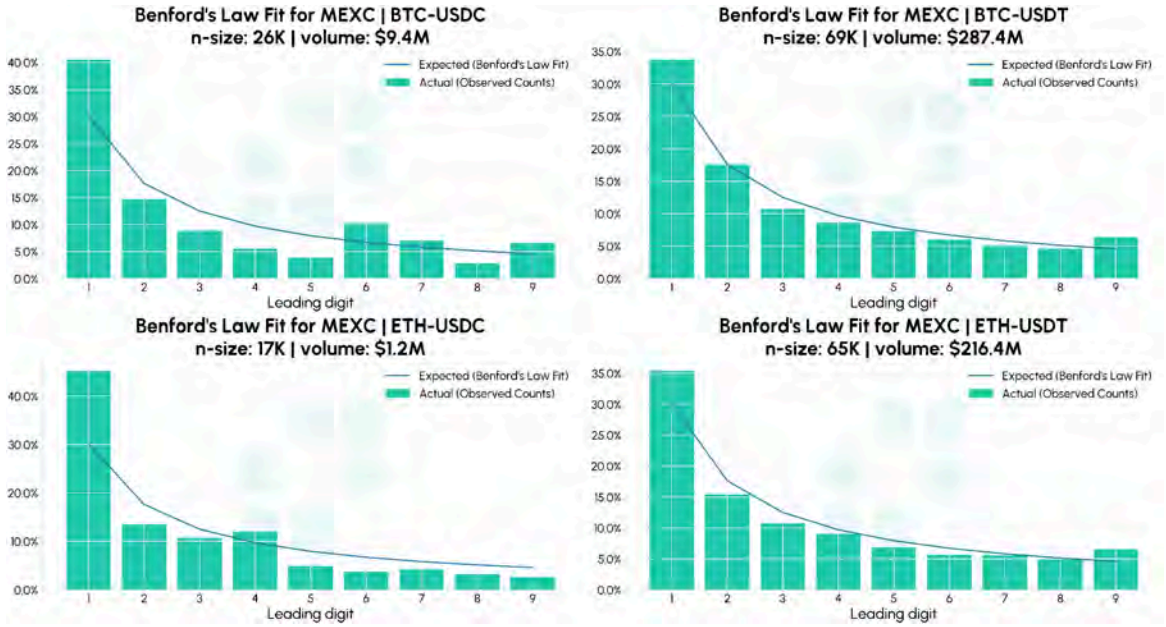


Figure 1c: Leading digit amounts for MEXC trades plotted against Benford's Law. Taken from a random sample of trades from 2025-01-01 to 2025-07-01.

LBANK

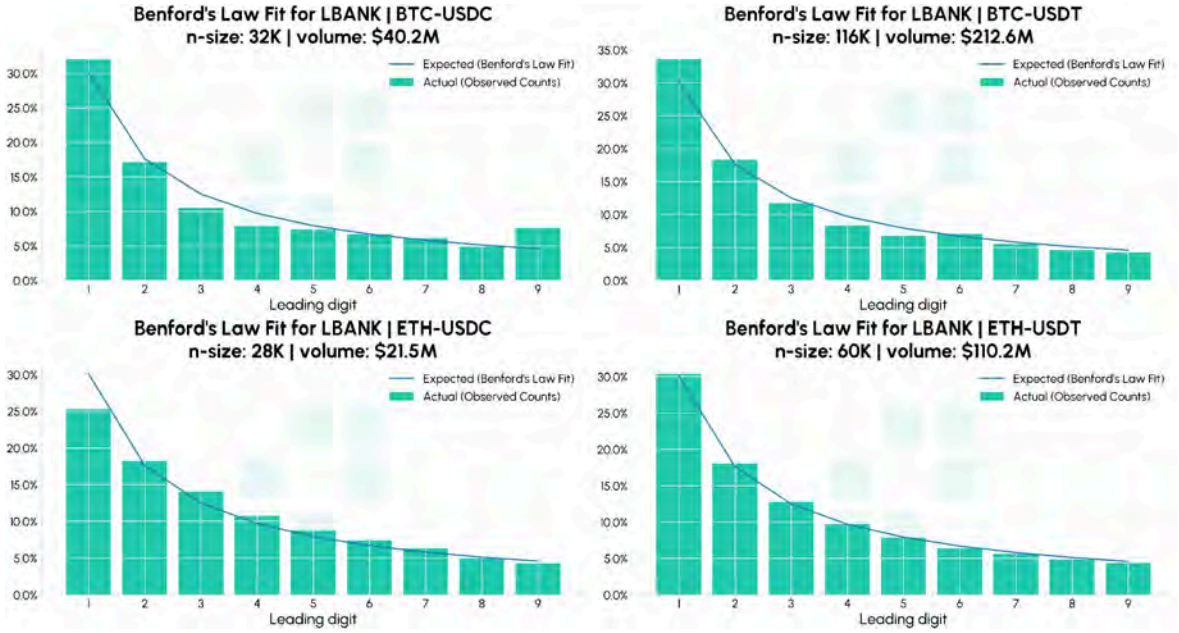


Figure 1d: Leading digit amounts for LBANK trades plotted against Benford's Law. Taken from a random sample of trades from 2026-01-01 to 2026-07-01.

A2. Trade Permutations

Most markets with organic volume follow trade permutation patterns with several consecutive buy or sell orders. By contrast, markets with heavy wash trading show a near-uniform distribution of trade permutations. Below we illustrate the difference between an organic market (coinbase-btc-usdt) and a market whose trade permutation patterns are consistent with heavy wash trading (mexc-btc-usdc).

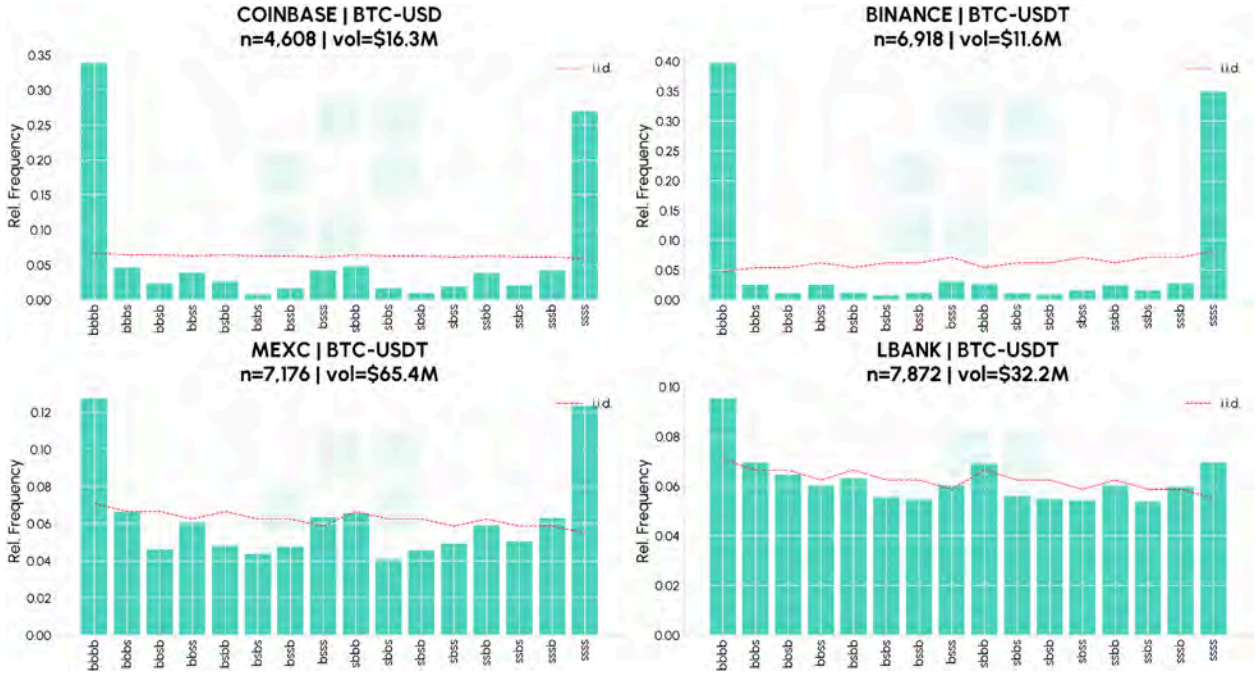


Figure 2a: 4-Trade buy-sell patterns from select markets. Taken from a random sample of trades from 2026-01-01 to 2026-07-01.

A3. Trade Size Distributions

Organic trading activity tends to cause the distribution of trade sizes to follow a bell-curve distribution with some peaks at round trade size in either base or quote units. Markets with significant levels of inorganic or spurious trades deviate from this distribution significantly. Organic markets (Coinbase BTC-USD, Bitstamp BTC-USD) exhibit

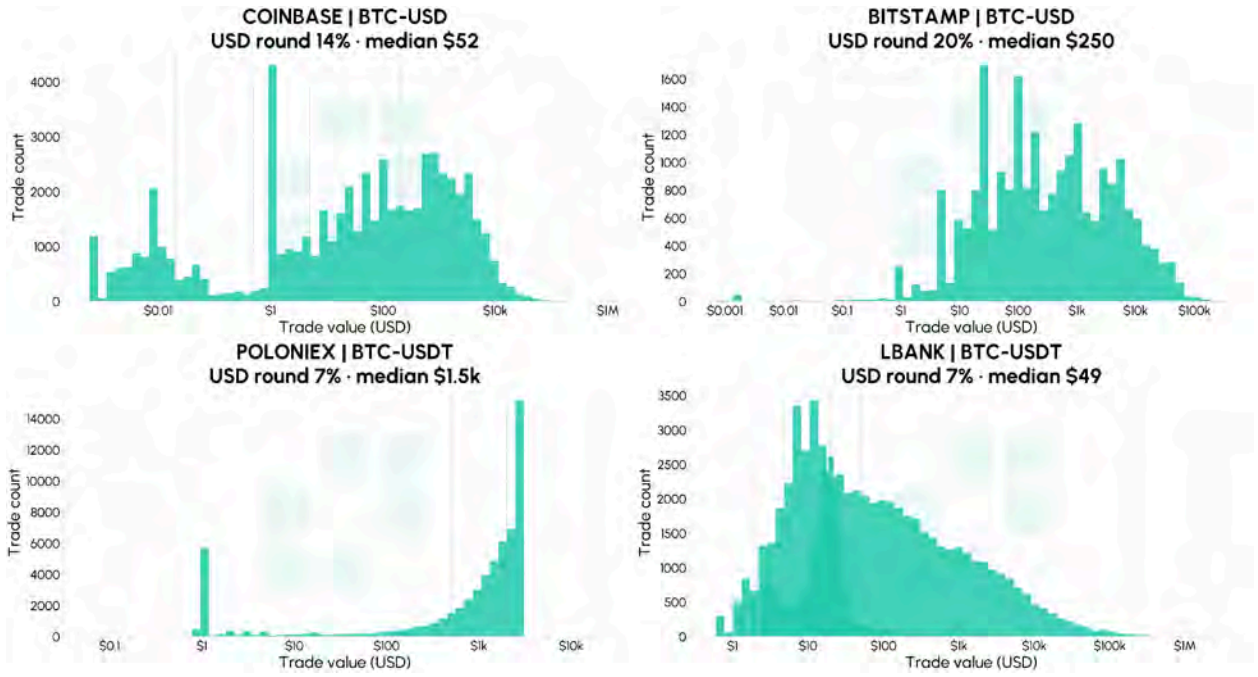
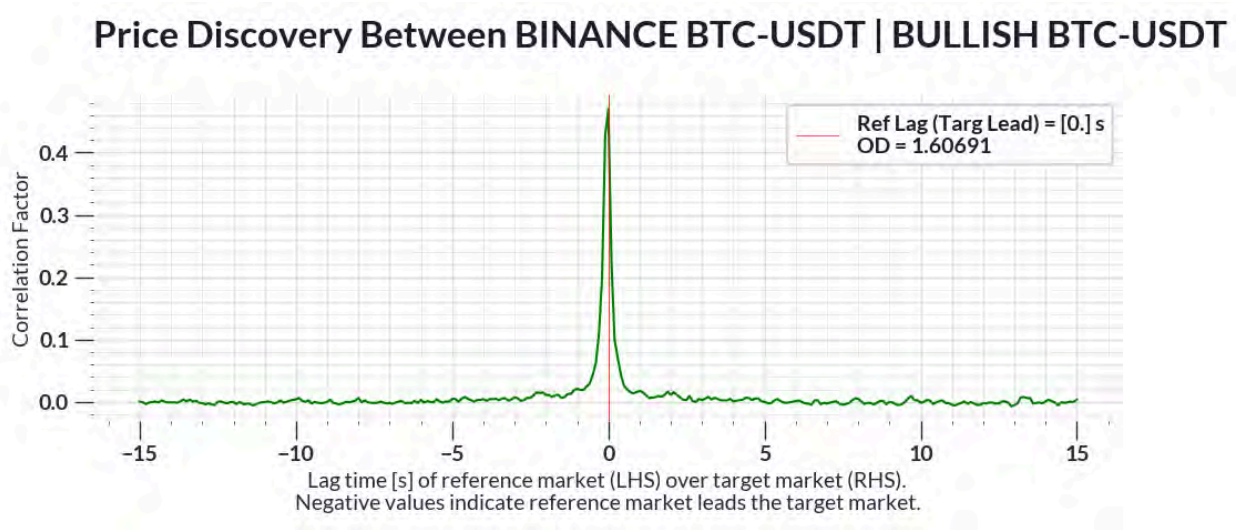


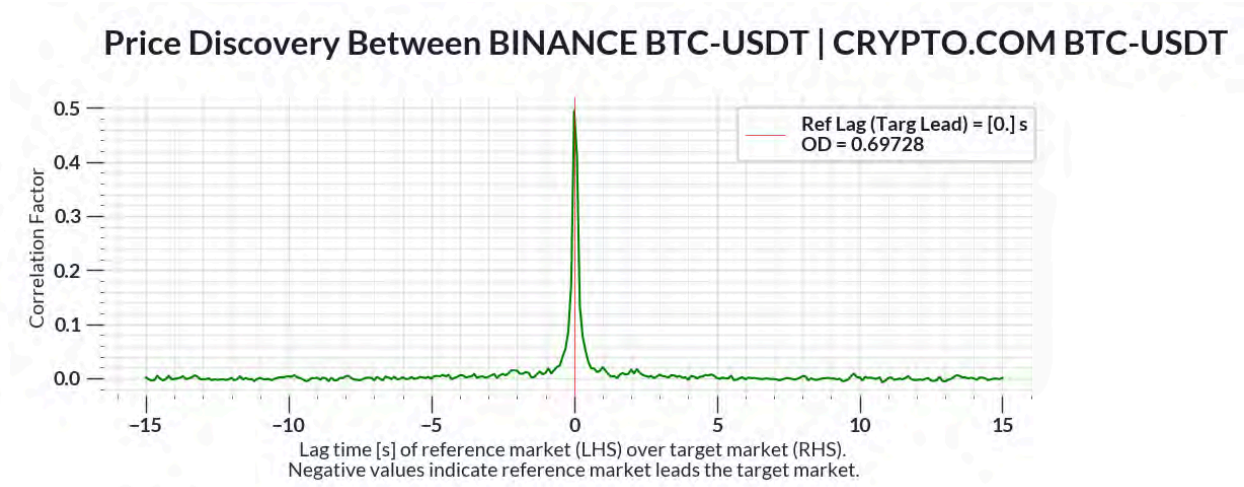
Figure 3: Trade size distribution from venues with more organic trading activity (top row) versus less (bottom row). Taken from a random sample of trades from 2025-01-01 to 2025-07-01.

A4. 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 you to observe how much a given market should be lagged for its prices to most strongly correlate with the adjacent reference market.

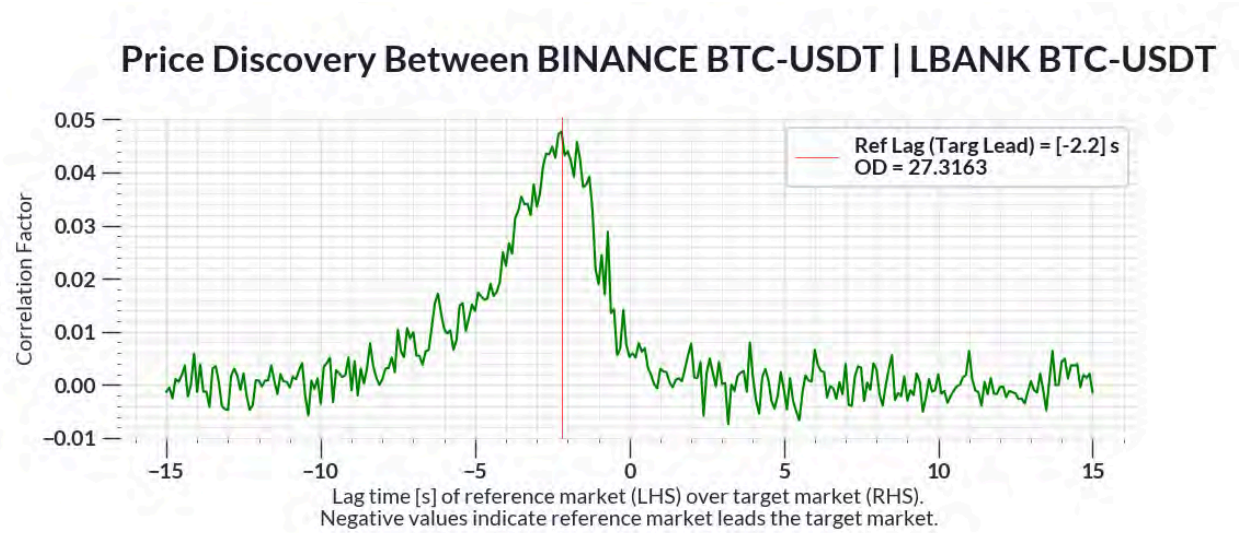
Consider the lag curves on a random sample during 2025 Q2 for the example BTC-USDT markets below, taking Binance as a reference exchange. Bitcoin prices between Binance and Bullish as well as Binance and Crypto.com have a max correlation at t = 0s, meaning the prices are in sync. These markets are generally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



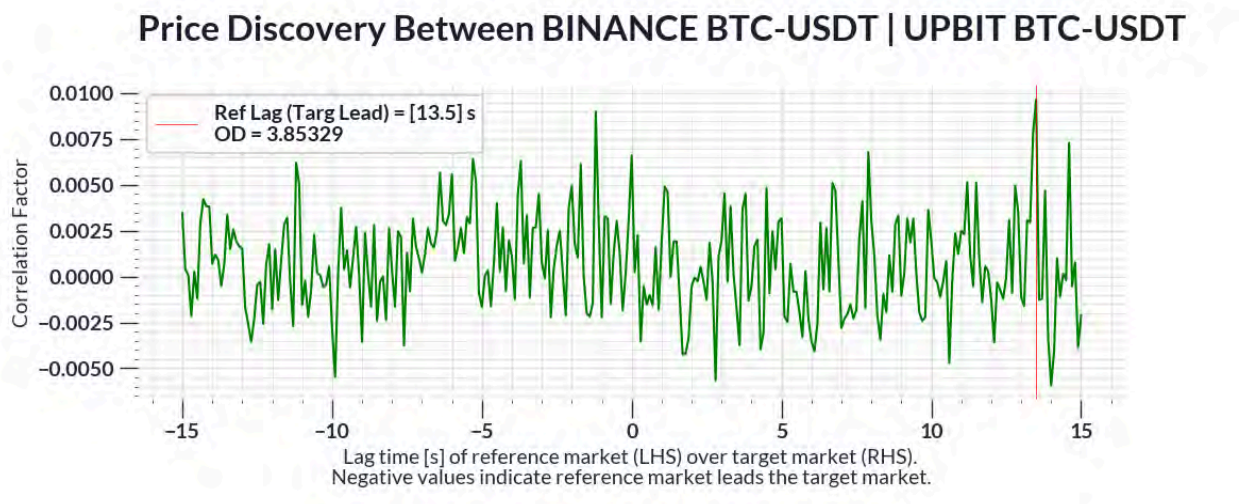


There are two common cases for markets to be inefficient: one where the target market clearly lags the reference market, and the other where there is no clear relationship between the prices of the target and reference markets.

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.



A5. Round Trip Trades

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.

Below are a set of executed trades by market in a randomly sampled time period.

BTC-USD

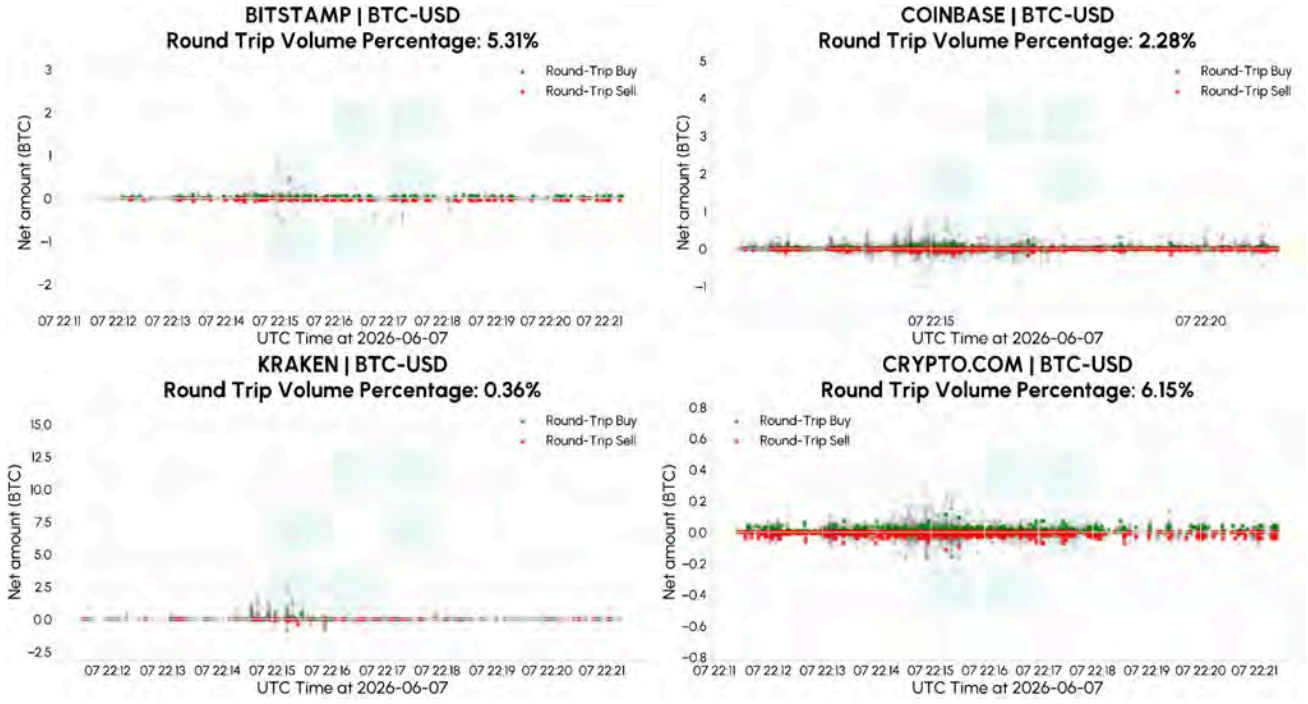
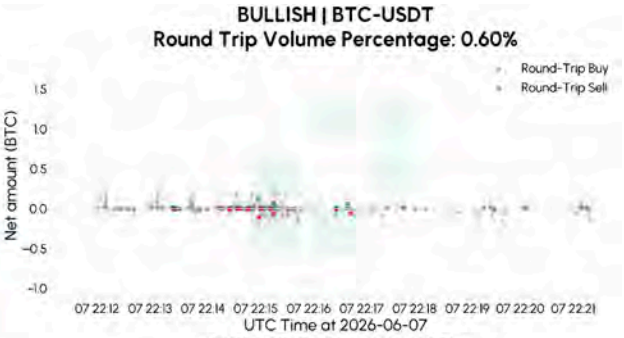
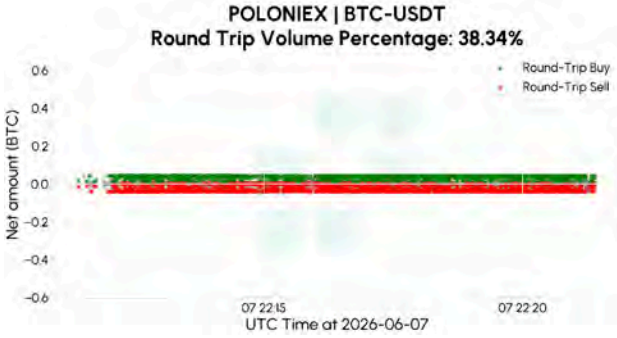


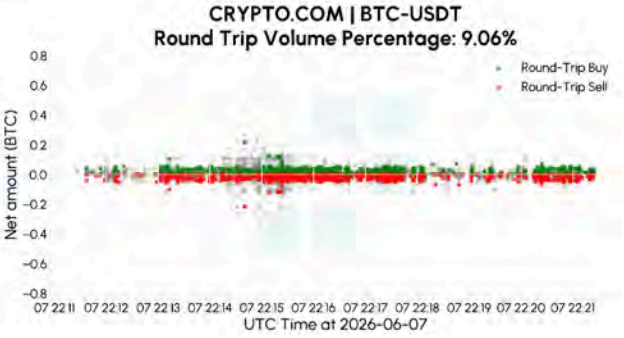
Figure 5a. Taken from a random sample of trades from 2025-01-01 to 2025-07-01.

BTC-USDT


BINANCE | BTC-USDT
Round Trip Volume Percentage: 0.48%
Round-Trip Buy
Round-Trip Sell
Net amount (BTC)
20
15
10
5
0
−5
07 22:15
07 22:20
UTC Time at 2026-06-07


BULLISH | BTC-USDT
Round Trip Volume Percentage: 0.60%
Round-Trip Buy
Round-Trip Sell
Net amount (BTC)
1.5
1.0
0.5
0.0
−0.5
−1.0
07 22:12
07 22:13
07 22:14
07 22:15
07 22:16
07 22:17
07 22:18
07 22:19
07 22:20
07 22:21
UTC Time at 2026-06-07


POLONIEX | BTC-USDT
Round Trip Volume Percentage: 38.34%
Round-Trip Buy
Round-Trip Sell
Net amount (BTC)
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
07 22:15
07 22:20
UTC Time at 2026-06-07


CRYPTO.COM | BTC-USDT
Round Trip Volume Percentage: 9.06%
Round-Trip Buy
Round-Trip Sell
Net amount (BTC)
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
07 22:11
07 22:12
07 22:13
07 22:14
07 22:15
07 22:16
07 22:17
07 22:18
07 22:19
07 22:20
07 22:21
UTC Time at 2026-06-07

A6. Orderbook Depth

Order-book depth measures the USD value resting within ±0.1% of the mid price, summed across bid and ask. Figure 6 shows each exchange's depth by side (bid above the axis, ask below) for BTC-USD, BTC-USDT, and BTC-USDC, time-averaged over the window and rolled up to the exchange as a volume-weighted average across its BTC markets. The ±0.1% band is used instead of a wider band because several venues, including Crypto.com and LMAX, publish only a limited number of order-book levels, leaving wider bands unpopulated for those exchanges.

Depth is scored on a continuous log-relative curve: the deepest venue is given a perfect score and decrements by ~2 per order of magnitude below it.

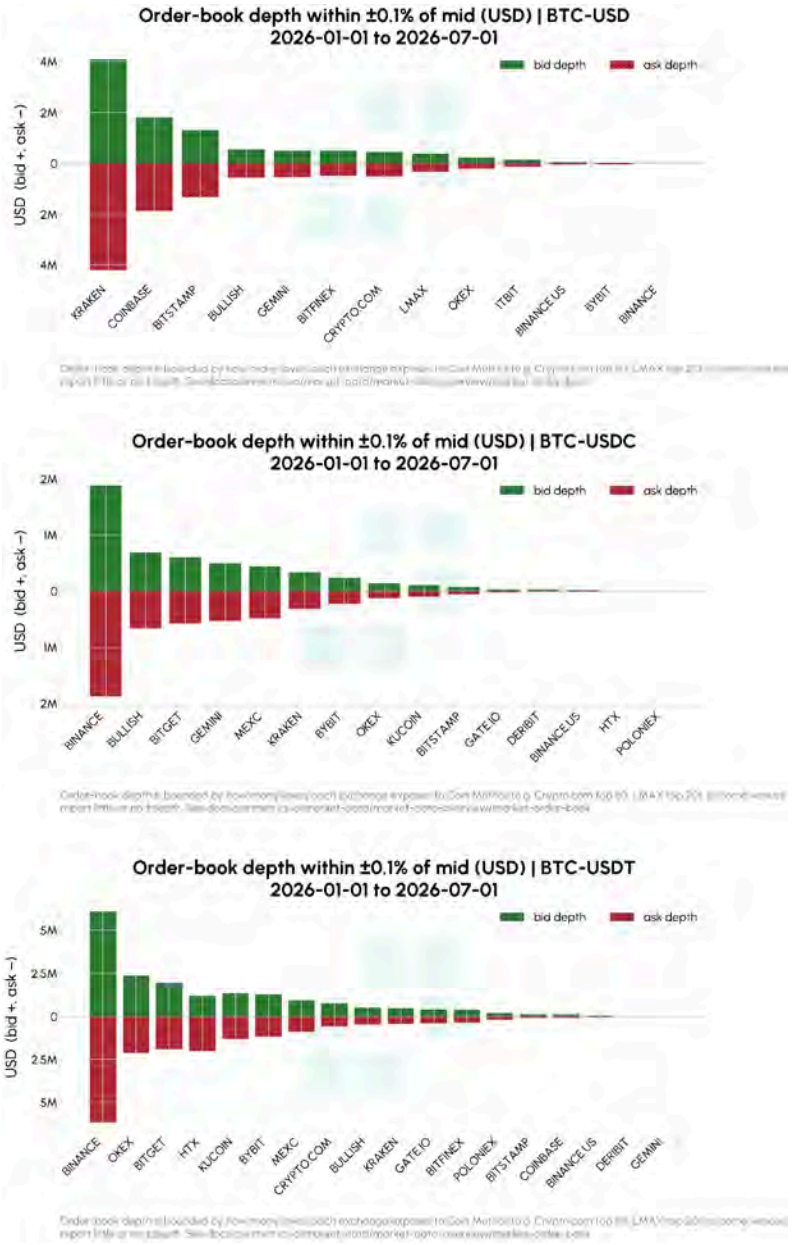


Figure 6. Mean 0.1% Orderbook Depth for BTC markets from 2026-01-01 to 2026-07-01.

A7. Slippage

The Slippage test estimates the cost of executing a $100,000 market order. the percentage difference between the volume-weighted fill price and the mid price, averaged across the buy and sell sides and converted to basis points. Figure 7 ranks exchanges on this measure for BTC-USD, BTC-USDT, and BTC-USDC, faceted by pair and shown on a log scale, with lower bars indicating cheaper execution.

Figure 7. Mean slippage for BTC markets from 2026-01-01 to 2026-07-01.

A8. Spread

Spread is the gap between the best bid and best ask, expressed as a percentage of the mid price and converted to basis points. Figure 8 shows the mean spread for each exchange's BTC-USD, BTC-USDT, and BTC-USDC markets over the analysis window, faceted by pair and shown on a log scale.

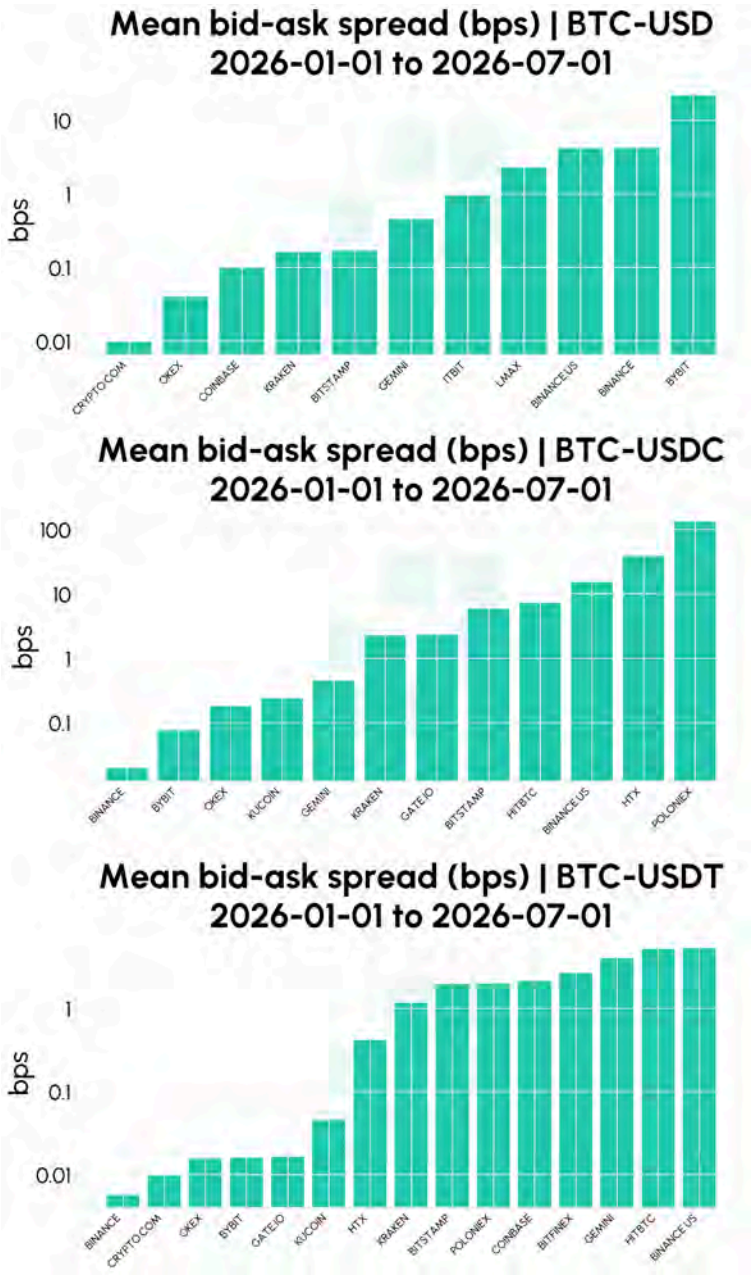


Figure 8. Mean bid-ask spread for BTC markets from 2026-01-01 to 2026-07-01.

A9. Open Interest

Open interest is the derivatives analogue of order-book depth: a venue's total futures open interest in USD, time-averaged over the analysis window. Figure 9 ranks exchanges on this measure on a linear scale. Unlike the order-book features, open interest is pulled at the exchange level: the aggregate across every futures market a venue lists, covering every asset, quote currency, and contract type, rather than a single BTC contract. Derivatives-native venues with no spot markets are scored alongside spot exchanges as a result. Scores use the same continuous log-relative curve as Orderbook Depth: the venue with the most open interest gets a perfect score, and each order of magnitude below it is decremented by a letter grade.

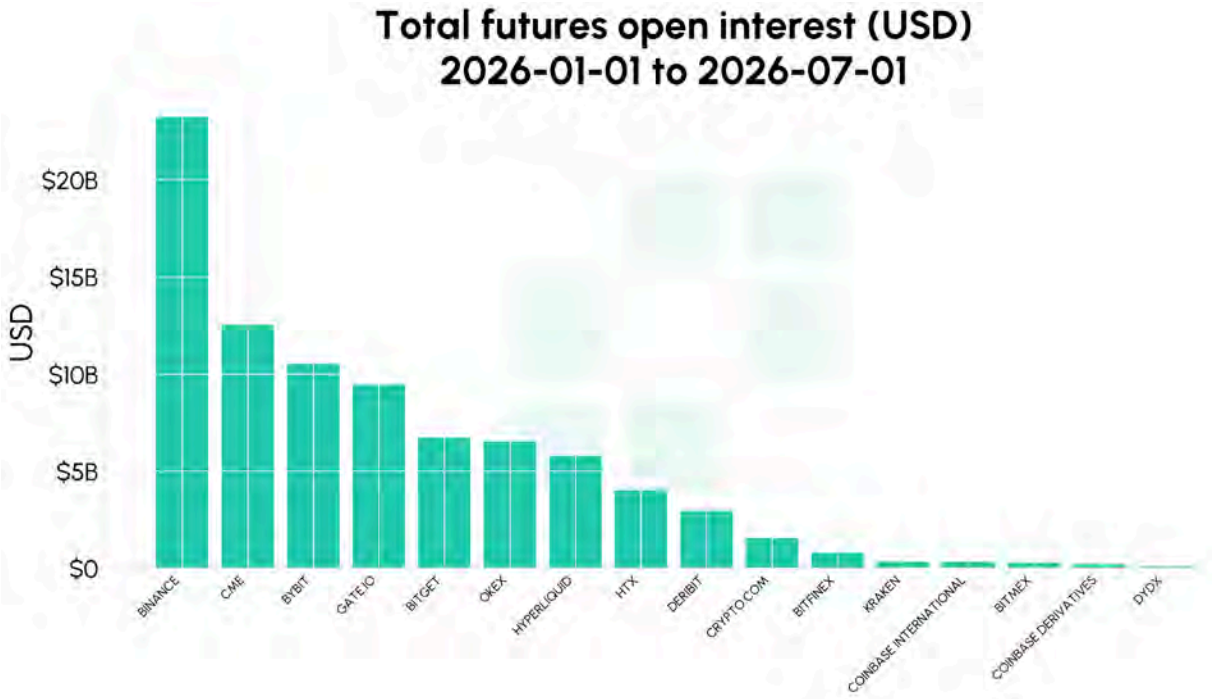


Figure 9. Open interest for exchanges from 2026-01-01 to 2026-07-01.

A10. Depth to Volume

Depth-to-Volume checks whether an exchange's reported spot volume is backed by real order-book liquidity. For each market, ±0.1% bid and ask depth in USD is divided by average daily volume to produce a support ratio; the ratio is volume-weighted to the exchange and scored on a one-sided curve that penalizes only thin backing, since volume and the depth backing it are difficult to fabricate at the same time.

Figure 10 plots each exchange's ±0.1% book depth (bid and ask summed across its BTC-USD, BTC-USDT, and BTC-USDC markets) against its average daily reported volume, faceted by pair alongside a fourth panel totaling all three. Green, yellow, and red bands mark the acceptance regime: a curved threshold that requires proportionally less depth-backing at higher volume, so a ratio that would be thin for a $10 million market is not for a $10 billion one. A venue's score is a sigmoid function of its log-distance between the yellow and green thresholds, so points in the green tier score highest and points below the red threshold score lowest.

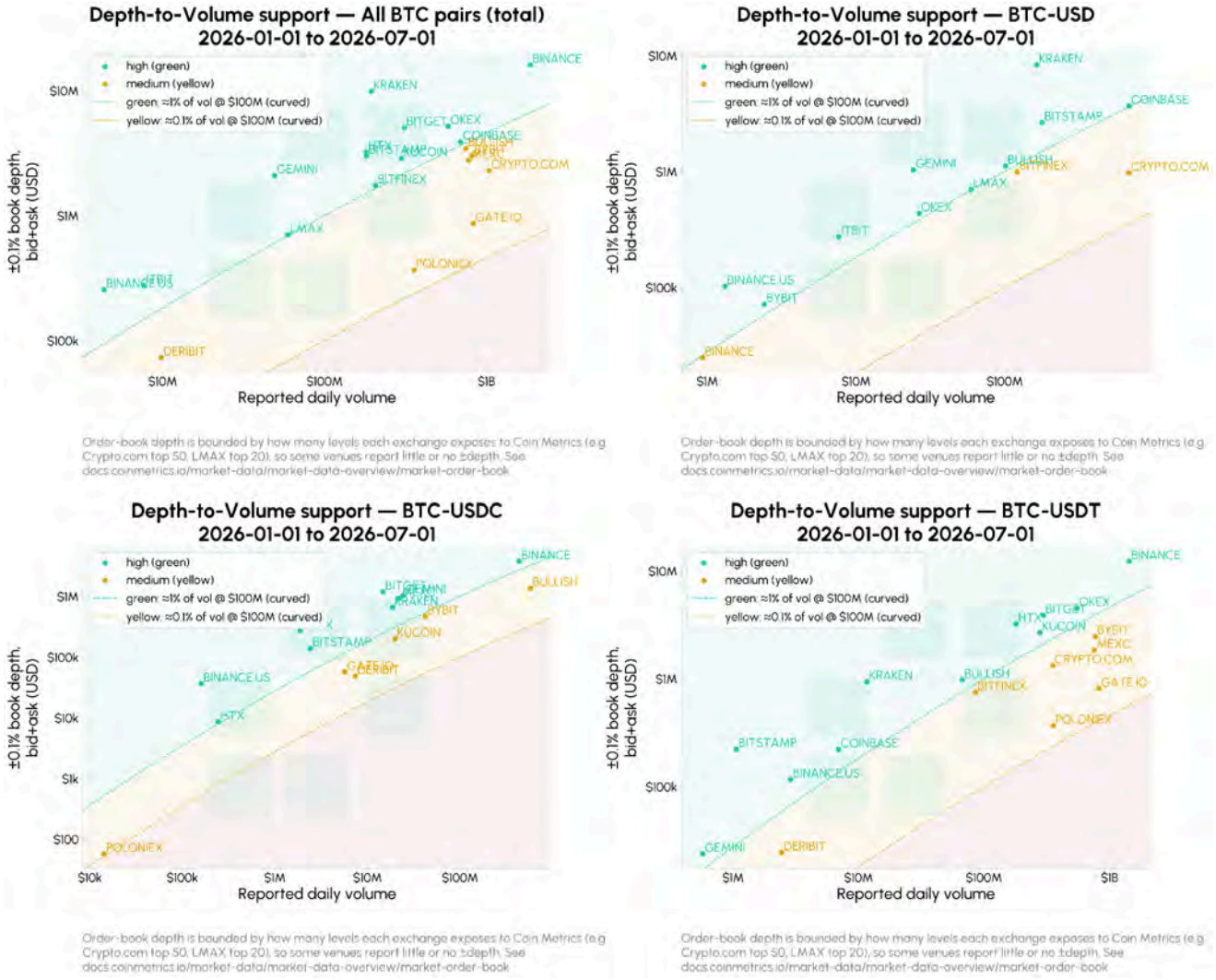


Figure 10. Ratio of orderbook depth to spot volume for BTC markets from 2026-01-01 to 2026-07-01.

References

1. 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.

2. "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.

3. 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.

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).

8. World Bank. "Worldwide Governance Indicators: Regulatory Quality (2024 Estimate)." *Data360*, https://data360.worldbank.org/en/dataset/WB_WGI. Accessed 22 July 2026.

talos.com