Research Paper

Multi-Scale Lead-Lag in High-Frequency Markets

Independent Implementation and Crypto-Market Adaptation of Hayashi and Koike (2020)

Théo Verdelhan · Leo Renault · Ziad El Arari

M2 272 Quantitative Finance – Paris-Dauphine PSL

June 2026

Abstract

This paper presents an independent implementation and empirical adaptation of the multi-scale lead-lag estimator of Hayashi and Koike (2020). The estimator is designed for non-synchronously observed high-frequency financial data and combines lagged Hayashi-Yoshida covariance with wavelet-based scale decomposition. The research question is whether price discovery in modern public crypto markets is better described by a single lead-lag coefficient or by a scale-dependent profile. Using public 2025–2026 tick data from Binance, Kraken, and Bybit, together with spot and perpetual futures contracts, we document robust multi-scale price discovery patterns. The main empirical finding is that Binance leads Kraken on BTC across all tested dyadic scales, from approximately -0.05 s at the finest scale to -1.95 s at the coarsest scale in April 2026. The direction is stable across January, February, and April 2026, while the magnitude varies with market regime and data density. Cross-product tests show that USDT-M perpetual futures lead COIN-M perpetuals and that perpetual futures lead spot markets for BTC and ETH at coarser scales. A single-scale HRY benchmark can return a near-zero estimate on the same data, illustrating the information loss caused by aggregation across frequencies. The implementation is validated through synthetic sign-convention audits, Monte Carlo diagnostics, tick-density artifact tests, and an equity sanity check on LOBSTER midpoint data. The contribution of the project is both methodological and empirical: it translates a mathematically non-trivial estimator into tested Python code and applies it to a public-data setting where exchange-level price discovery can be studied without proprietary TAQ feeds.

Keywords: lead-lag, market microstructure, Hayashi-Yoshida covariance, Hayashi-Koike estimator, wavelets, asynchronous data, crypto markets, price discovery, high-frequency finance.

Contents

1 Introduction

High-frequency financial markets do not react to information at a single time scale. At the millisecond and sub-second horizon, price changes are shaped by market data propagation, matching-engine latency, queue priority, and liquidity imbalance. At the multi-second horizon, the same price changes may reflect slower cross-venue arbitrage, risk transfer between spot and derivatives, or the inventory management of market makers. A scalar lead-lag coefficient compresses all of these channels into one number. This may be convenient, but it is often too poor a representation of price discovery.

The purpose of this project is to build a research-grade implementation of a multi-scale lead-lag estimator and to test whether the resulting scale profile reveals market structure effects that would be hidden by a single-scale method. The methodological basis is the estimator introduced by Hayashi and Koike [1]. The estimator first computes a lagged Hayashi-Yoshida cross-covariance, which is suitable for non-synchronous observations. It then filters this covariance function with wavelet autocorrelation filters, producing a separate contrast function at each dyadic scale. The estimated lead-lag at scale j is the lag at which this filtered contrast is maximized in absolute value.

The empirical application departs from the original paper in one important way. Hayashi and Koike study U.S. equity price discovery using TAQ-style quote data and micro-prices. In this project, the estimator is adapted to public high-frequency crypto data. The closest public analogue to the NASDAQ/BATS setting is same-asset cross-venue price discovery: for example BTC traded on Binance versus BTC traded on Kraken or Bybit. This setting is not a literal replication of the paper's empirical panel, but it is an economically meaningful test of the same statistical problem: two prices react to common information, are observed on irregular and non-synchronous clocks, and may not incorporate the same information at the same speed.

The implementation focuses on four objectives.

(i) **Mathematical fidelity.** The estimator should follow the Hayashi-Koike construction: lagged Hayashi-Yoshida covariance, Daubechies filters, autocorrelation wavelets, and scale-by-scale argmax estimation.

(ii) **Empirical discipline.** The sign convention must be audited and documented, because an apparently small sign mistake reverses the economic interpretation of every result.

(iii) **Public-data applicability.** The project should run on public exchange archives and APIs, not on proprietary quote feeds. This requires careful data loading, timestamp alignment, and artifact-based reporting.

(iv) **Research communication.** The final notebook, figures, tests, and paper should allow an external reviewer to understand what was implemented, why it was implemented, and what was learned from the empirical results.

The main result is clear: Binance leads Kraken on BTC across all tested scales in the April 2026 full-month sample. The estimated lag is approximately -0.05 s at scale $j = 1$ and reaches -1.95 s at scale $j = 8$. Under the codebase convention, a negative estimate means that series 1 leads series 2. When the pair is ordered as Binance first and Kraken second, the negative profile therefore indicates that Binance leads Kraken. The same direction is observed in January and February 2026, with different magnitudes.

The second important result is that the lead-lag profile depends on the market relationship. Binance and Kraken show a strong cross-venue effect. Binance and Bybit are close to synchronous on BTC, with only a small coarser-scale Bybit lead. In cross-product tests, USDT-M perpetuals lead COIN-M perpetuals, while perpetual futures lead spot markets for BTC and ETH. This is consistent with the idea that derivatives venues, where leverage and hedging demand concentrate, can incorporate information before spot markets at some horizons.

The third result is methodological. On a BTC/ETH spot pair, a single-scale HRY benchmark returns a lead-lag estimate close to zero, while the multi-scale HK estimator shows a structured profile that becomes positive at coarser scales. This does not mean that HRY is wrong; it means that a scalar average can cancel or dilute scale-dependent effects. The multi-scale estimator adds value when the research question is not only "who leads?" but also "at which horizon?".

2 Research Question and Contribution

2.1 Economic motivation

Lead-lag relationships are a direct empirical window into price discovery. If two assets or two venues are exposed to the same information, the market that updates first is likely closer to the active information-processing layer. In equity markets, this question has been studied across exchanges, quote feeds, and liquidity venues. In crypto markets, the question is equally relevant because the market is fragmented across exchanges, spot books, and derivative contracts, but the data are more accessible through public archives.

The economic interpretation depends on the pair being studied.

- For **same-asset cross-venue** pairs, a lead can be interpreted as cross-exchange price discovery. If Binance BTC leads Kraken BTC, then Binance is updating earlier to common BTC information during the sample.
- For **spot versus perpetual** pairs, a lead can be interpreted as cross-product price discovery. A perpetual futures lead suggests that the derivatives market absorbs information faster than the spot market at the relevant scale.
- For **two crypto assets** such as BTC and ETH, the result is not a pure price discovery ranking. It mixes common risk factors, index-level crypto shocks, liquidity differences, and potentially different investor clienteles.
- For **equity midpoint** pairs, the exercise is mainly a sanity check: it verifies that the implementation also operates on quote-derived price processes similar to those used in the original literature.

A central point of this project is that these interpretations are horizon-dependent. The leading venue at a 100 ms horizon need not have the same economic meaning as the leading venue at a 20 s horizon. At the finest scales, the result is closer to microstructure latency. At coarser scales, it is closer to slower arbitrage and market-wide information diffusion.

2.2 Statistical motivation

High-frequency data are irregular. Trades do not occur at deterministic intervals, and different assets or venues are rarely observed at the same time. Letting X^1 and X^2 denote two latent log-price processes, we observe samples

$$0 = t_0^1 < t_1^1 < \cdots < t_{n_1}^1 = T, \qquad 0 = t_0^2 < t_1^2 < \cdots < t_{n_2}^2 = T,$$

where the two grids are generally different. A naive regular-grid interpolation can create artificial correlation and distort lead-lag estimates. This issue is well known in the literature on realized covariance under non-synchronous trading. The Hayashi-Yoshida estimator addresses this by using overlapping observation intervals rather than forcing both series onto a common grid.

The second statistical issue is scale aggregation. A covariance function $\hat{U}_N(\tau)$ can be evaluated over a grid of lags τ, but its maximum is a single number. If the underlying dependence has multiple frequency components, the global maximum may be dominated by one band or weakened by cancellation between bands. Wavelet filtering gives a way to study the covariance function locally in scale.

2.3 Contribution of the project

The contribution of this project is not to propose a new estimator. The contribution is to implement, validate, and adapt a technically demanding estimator in a modern public-data environment. The project adds value in five ways.

(1) It gives a complete Python implementation of the lagged Hayashi-Yoshida and Hayashi-Koike estimators, including a Numba-accelerated overlap scan.

(2) It documents the sign convention through synthetic tests. This is essential because the definition of a positive lag depends on which series is shifted in the covariance calculation.

(3) It studies full-month public crypto tick data across venues and products, not only toy examples.

(4) It compares the multi-scale HK estimator to a single-scale HRY baseline and shows a case where the scalar result is economically less informative.

(5) It packages the research output as reproducible artifacts: tracked CSV summaries, figures, a narrative notebook, tests, and this paper.

3 Mathematical Framework

3.1 Latent price processes

The starting point is a bivariate continuous-time log-price process. Following the spirit of Hayashi and Koike [1], let

$$X_t^\nu = X_0^\nu + \int_0^t \sigma_s^\nu \, dB_s^\nu, \qquad \nu \in \{1, 2\},$$

where B^1 and B^2 are Brownian components and σ_t^ν are stochastic volatility processes. The dependence between the two price processes is not summarized by a single instantaneous correlation. Instead, it can vary by frequency band. Intuitively, shocks at different time scales may be transmitted with different delays.

For a simple scalar lead-lag model, one might write that X^2 reacts to X^1 with delay θ, or that the cross-covariance between increments is largest when one process is shifted by θ. The multi-scale model generalizes this idea by allowing a separate delay θ_j at each dyadic scale j. The research target becomes the vector

$$(\theta_1, \theta_2, \ldots, \theta_J),$$

rather than a single θ.

3.2 Observation intervals and non-synchronous returns

The observed data are irregular samples of the two processes. Define observation intervals

$$I_i^1 = (t_{i-1}^1, t_i^1], \qquad I_k^2 = (t_{k-1}^2, t_k^2],$$

with increments

$$\Delta X^1(I_i^1) = X_{t_i^1}^1 - X_{t_{i-1}^1}^1, \qquad \Delta X^2(I_k^2) = X_{t_k^2}^2 - X_{t_{k-1}^2}^2.$$

The key point is that I_i^1 and I_k^2 do not need to have the same endpoints. The information used by the Hayashi-Yoshida estimator is whether these intervals overlap. Let

$$K(I, J) = \mathbf{1}\{I \cap J \neq \emptyset\}.$$

This overlap indicator replaces calendar-time matching. It is the core device that allows covariance estimation without previous-tick interpolation.

3.3 Lagged Hayashi-Yoshida covariance

For a candidate lag τ, the second observation grid is shifted by τ. The lagged Hayashi-Yoshida covariance is

$$\hat{U}_N(\tau) = \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} \Delta X^1(I_i^1) \Delta X^2(I_k^2) \mathbf{1}\{I_i^1 \cap (I_k^2 + \tau) \neq \emptyset\}. \tag{1}$$

Here $I_k^2 + \tau$ denotes the interval I_k^2 shifted by τ. In the implementation used in this project, positive τ shifts the second series forward in time. The sign convention is therefore:

$$\hat{\theta} < 0 \iff \text{series 1 leads series 2}, \qquad \hat{\theta} > 0 \iff \text{series 2 leads series 1}. \tag{2}$$

The single-scale lead-lag estimator associated with this covariance is

$$\hat{\theta}_{\mathrm{HY}} = \arg\max_{\tau \in G_N} |\hat{U}_N(\tau)|, \tag{3}$$

where G_N is a finite lag grid. In the crypto experiments, the grid is

$$G_N = \{-400, -399, \ldots, 399, 400\}\Delta_N, \qquad \Delta_N = 0.05\,\mathrm{s},$$

so that the search range is approximately $\pm 20\,\mathrm{s}$.

3.4 Why interpolation can be dangerous

A common alternative is to sample both processes on a fixed grid and fill missing values with the previous tick. This is operationally simple but statistically fragile. If one asset trades more frequently than another, previous-tick interpolation can mechanically introduce stale prices in one series and fresh prices in the other. The induced cross-covariance can look like a lead-lag even when the latent processes are synchronous.

The Hayashi-Yoshida estimator avoids this specific problem by working directly with the original observation intervals. It does not make all timestamps regular, and it does not assign a price update to a time at which no trade or quote update was observed. This is why it is a natural building block for a lead-lag estimator in asynchronous markets.

3.5 From scalar dependence to multi-scale dependence

The scalar estimator in (3) is still insufficient when the dependence is scale-dependent. Suppose that at high frequency one venue leads by 50 ms, while at lower frequency the same venue leads by 2 s. A single argmax can return one of these values, a compromise, or even zero if the covariance contributions interact destructively. To isolate horizon-specific relationships, the covariance function must be decomposed by scale.

Wavelets provide a natural way to do this. A wavelet basis decomposes a signal into localized frequency bands. In this project, the signal being filtered is not the price path itself but the lagged cross-covariance function $\hat{U}_N(\tau)$. Filtering this function gives a scale-specific contrast $\hat{\Gamma}_j(\tau)$. The lead-lag estimate at scale j is then the lag at which this filtered contrast is maximal.

3.6 Daubechies filters

Let $(h_p)_{p=0}^{L-1}$ denote a Daubechies high-pass wavelet filter of even length L. In the implementation, the default is db10, so $L = 20$. The associated low-pass scaling filter is constructed using the quadrature mirror relationship

$$g_p = (-1)^{p+1} h_{L-p-1}, \qquad p = 0, \dots, L-1. \tag{4}$$

The Daubechies family is useful because it combines compact support and vanishing moments. Compact support keeps the filters finite and computationally tractable. Vanishing moments make the wavelet insensitive to low-order local polynomial trends, which is desirable when filtering price-related objects.

The level-j wavelet filter $h_{j,p}$ is constructed recursively. At the first level,

$$h_{1,p} = h_p, \qquad p = 0, \dots, L-1.$$

For $j \geq 2$,

$$h_{j,p} = \sum_q g_{p-2q} h_{j-1,q}, \qquad p = 0, \dots, L_j - 1, \tag{5}$$

where coefficients outside the support of g are treated as zero, and

$$L_j = (2^j - 1)(L - 1) + 1. \tag{6}$$

The support length grows with j, which is expected: coarser scales require wider filters in the time domain.

3.7 Autocorrelation wavelets

The Hayashi-Koike estimator uses the autocorrelation of the level-j wavelet filter:

$$\Psi_j(l) = \sum_p h_{j,p} h_{j,p+|l|}, \qquad l = -(L_j - 1), \dots, L_j - 1. \tag{7}$$

The sequence $\Psi_j(l)$ is symmetric and finite. It is the discrete filtering kernel applied to the lagged covariance function. Its transfer function is

$$H_{j,L}(\lambda) = \sum_{l=-(L_j-1)}^{L_j-1} \Psi_j(l) e^{-il\lambda}. \tag{8}$$

In the ideal Littlewood-Paley intuition, $H_{j,L}(\lambda)$ behaves like a band-pass filter for a dyadic frequency band. In finite samples, the filter is not a perfect indicator, but it localizes the covariance around a specific time scale. With finest resolution Δ_N, the empirical band associated with scale j is reported as

$$\big[2^j \Delta_N,\ 2^{j+1} \Delta_N\big]. \tag{9}$$

For crypto experiments, $\Delta_N = 0.05\,\mathrm{s}$, so the first scale corresponds to 0.10–0.20 s, while the eighth scale corresponds to 12.80–25.60 s.

3.8 Hayashi-Koike filtered contrast

The scale-j filtered covariance contrast is

$$\hat{\Gamma}_j(\tau) = \sum_{l=-(L_j-1)}^{L_j-1} \hat{U}_N(\tau - l\Delta_N)\Psi_j(l). \tag{10}$$

This equation is the computational center of the estimator. First, the lagged Hayashi-Yoshida covariance $\hat{U}_N(\tau)$ is evaluated on an extended lag grid, because the filter requires values at $\tau - l\Delta_N$. Second, for each scale j, the covariance sequence is convolved with the reversed autocorrelation wavelet. Third, the valid segment corresponding to the target grid G_N is extracted.

The scale-specific lead-lag estimator is then

$$\hat{\theta}_j = \arg\max_{\tau \in G_N} \left| \hat{\Gamma}_j(\tau) \right|. \tag{11}$$

The absolute value is important. The estimator searches for the strongest dependence at that scale, regardless of whether the covariance is positive or negative. In the empirical crypto pairs studied here, the economic interpretation is based primarily on the location of the maximum, not on the sign of the covariance peak.

3.9 Interpretation of scale-specific estimates

The output of the HK estimator is a vector

$$(\hat{\theta}_1, \ldots, \hat{\theta}_J).$$

It should not be read as eight independent experiments. The scales are connected by the same underlying price processes and the same observation windows. A smooth monotone profile, such as the Binance-Kraken BTC result, is more convincing than an isolated nonzero value at one scale. Conversely, a noisy profile with alternating signs is a warning that either the relationship is weak or the data are not sufficiently informative at that scale.

The contrast peak is also informative. A larger value of $\max_\tau |\hat{\Gamma}_j(\tau)|$ indicates a stronger filtered covariance at that scale. In the full-month BTC Binance-Kraken experiment, the contrast peak rises substantially with j, consistent with stronger coarser-scale co-movement. However, contrast magnitudes should not be compared mechanically across very different pairs without considering volatility, tick density, and sample length.

4 Implementation

4.1 Repository architecture

The codebase is organized as a small research package. The core estimators live in `hk_leadlag/estimators`. The wavelet construction is implemented in `hk_leadlag/wavelet.py`. Exchange-specific data loaders are separated from the estimator logic, and experiment runners in `scripts/` write artifact directories under `outputs/`. This separation is important: estimation code should not depend on a particular data source, and empirical artifacts should be reproducible from explicit configurations.

The main components are:

- `NonSyncSeries`: container for two irregular timestamp arrays and two price arrays.
- `HayashiYoshidaEstimator`: lagged HY covariance and scalar argmax.
- `WaveletLeadLagEstimator`: HK multi-scale estimator.
- `DaubechiesFilter`: high-pass and low-pass filters, recursive level filters, autocorrelation wavelets.
- `ArtifactStore`: standardized experiment outputs, including configuration, summary tables, metadata, and figures.

4.2 Computational algorithm

The computational bottleneck is the lagged HY covariance. For each candidate lag, one could naively loop over all pairs of intervals, but that would be prohibitively expensive. The implementation uses a two-pointer overlap scan. For each lag τ, the algorithm walks through the two sorted interval lists and advances the interval that ends first. This reduces the single-lag calculation to linear complexity in the number of intervals. The implementation then evaluates all lags in a Numba-compiled loop.

Table 1: Estimator pipeline used in the project.

Step	Description
Input construction	Load two tick series, sort timestamps, construct log-price arrays, and align the common observation window.
Lag grid	Define $G_N = \{-G, \dots, G\}\Delta_N$ with $\Delta_N = 0.05\,\mathrm{s}$ and $G = 400$ for crypto runs.
Extended HY grid	Evaluate $\hat{U}_N(\tau)$ on a wider grid to allow convolution with the largest wavelet support.
Wavelet filters	Build db10 level filters $h_{j,p}$ and autocorrelation kernels $\Psi_j(l)$ for $j = 1, \dots, 8$.
Filtered contrasts	Compute $\hat{\Gamma}_j(\tau)$ by convolving $\hat{U}_N$ with Ψ_j.
Argmax	Estimate $\hat{\theta}_j = \arg\max_{\tau \in G_N} \lvert\hat{\Gamma}_j(\tau)\rvert$.
Artifacts	Save summary CSV files, metadata, figures, and notebook-ready outputs.

4.3 Sign-convention audit

The sign convention is not left implicit. The script `scripts/audit_sign_convention.py` creates synthetic lagged series and checks the sign returned by both HY and HK estimators. With the implementation's definition of lag, the convention is:

$$\hat{\theta}_j < 0 \quad \text{means series 1 leads series 2.}$$

This convention is used consistently throughout the empirical analysis. For example, in the pair "BTC Binance vs Kraken", Binance is series 1 and Kraken is series 2. Therefore a negative estimate means Binance leads Kraken.

4.4 Data sources and price proxies

The project uses public high-frequency data. Crypto experiments are based on transaction prices from exchange archives or APIs. The equity sanity check uses LOBSTER order-book data and midpoint-derived prices.

Table 2: Data sources used in the empirical analysis.

Source	Data type	Role in the project
Binance Vision	Spot and futures aggregate trades	Main public source for BTC/ETH spot, USDT-M perpetuals, and COIN-M perpetuals.
Kraken REST	Trades	Cross-venue benchmark against Binance for BTC and ETH.
Bybit public archive	Trades	Alternative cross-venue benchmark for BTC.
LOBSTER sample	Message and order-book data	Quote-based equity sanity check using midpoint prices.

The crypto application therefore uses transaction log-prices rather than micro-prices. This is a practical adaptation, not a claim that trades are a perfect substitute for quotes. Trade prices are public and easy to audit, but they mix market order direction, bid-ask bounce, and exchange-specific trade aggregation rules. The LOBSTER exercises help verify that the estimator also works in a quote-based environment.

4.5 Artifact mode

Full raw tick archives are not committed to the repository. Instead, the final notebook and this paper rely on lightweight tracked artifacts: `summary.csv`, `series_stats.json`, and final figures. This is a deliberate research-engineering choice. A reviewer can inspect the numerical results and reproduce the paper figures without downloading tens of gigabytes of raw data. At the same time, scripts are included to regenerate the main experiments when the raw data are available locally.

5 Empirical Design

5.1 Main design choices

The core crypto experiments use the following common settings:

- finest lag step $\Delta_N = 0.05\,\mathrm{s}$;
- scales $j = 1, \dots, 8$;
- db10 wavelet filter;
- lag search half-width $400\Delta_N = 20\,\mathrm{s}$;
- full calendar-month windows for the main cross-venue and cross-product runs.

The dyadic scale mapping is shown in Table 3.

Table 3: Dyadic scale mapping for crypto experiments with $\Delta_N = 0.05\,\mathrm{s}$.

Scale j	$2^j\Delta_N$	$2^{j+1}\Delta_N$	Interpretation
1	0.10 s	0.20 s	Sub-second microstructure band
2	0.20 s	0.40 s	Fast cross-venue adjustment
3	0.40 s	0.80 s	Sub-second adjustment
4	0.80 s	1.60 s	One-second information diffusion
5	1.60 s	3.20 s	Short arbitrage horizon
6	3.20 s	6.40 s	Multi-second reaction band
7	6.40 s	12.80 s	Coarser liquidity adjustment
8	12.80 s	25.60 s	Slowest band in the main design

5.2 Core empirical questions

The empirical section is organized around four questions.

(1) Does Binance lead Kraken on BTC, and is the result stable by scale?

(2) Is the Binance-Kraken BTC result stable across months?

(3) How do lead-lag profiles differ across venues and products?

(4) What is lost when the multi-scale estimator is replaced by a single-scale benchmark?

These questions are deliberately more modest than a trading-strategy claim. The goal is not to show a deployable arbitrage signal. A trading signal would require latency modeling, fees, queue position, market impact, and exchange-specific execution constraints. The goal here is to measure scale-dependent price discovery in public high-frequency data.

5.3 Sample sizes and tick densities

The full-month runs involve millions of ticks. Table 4 summarizes the main empirical samples. The difference between median and mean inter-arrival times is substantial, which is typical for trade data: activity clusters during information events, while quiet periods create long gaps.

Table 4: Sample sizes and observation densities for the main full-month runs.

Run	n_1	n_2	Median dt1	Median dt2
BTC Binance vs Kraken, Apr. 2026	12,615,384	864,522	32.49 ms	1388.18 ms
ETH Binance vs Kraken, Apr. 2026	11,712,941	529,204	13.87 ms	2280.30 ms
BTC USDT-M perp vs COIN-M perp, Apr. 2026	16,081,100	1,918,505	75.00 ms	81.00 ms
BTC spot vs USDT-M perp, Apr. 2026	12,615,384	16,081,100	32.49 ms	75.00 ms
ETH spot vs USDT-M perp, Apr. 2026	11,712,941	17,913,017	13.87 ms	64.00 ms
BTC Binance vs Bybit, Apr. 2026	12,615,384	5,513,595	32.49 ms	22.00 ms

6 Core Result: BTC Binance vs Kraken

6.1 Full-month April 2026 result

The central experiment compares BTCUSDT on Binance spot to XBTUSD on Kraken over April 2026. Binance is series 1 and Kraken is series 2. The estimates are negative at every scale, so the conclusion is that Binance leads Kraken across the full dyadic range considered in the project.

Table 5: BTC Binance vs Kraken, April 2026 full month. Negative values mean Binance leads Kraken.

Scale j	Period min	Period max	$\hat{\theta}_j$	Interpretation
1	0.10 s	0.20 s	-0.05 s	Binance leads Kraken
2	0.20 s	0.40 s	-0.15 s	Binance leads Kraken
3	0.40 s	0.80 s	-0.20 s	Binance leads Kraken
4	0.80 s	1.60 s	-0.30 s	Binance leads Kraken
5	1.60 s	3.20 s	-0.50 s	Binance leads Kraken
6	3.20 s	6.40 s	-0.80 s	Binance leads Kraken
7	6.40 s	12.80 s	-1.25 s	Binance leads Kraken
8	12.80 s	25.60 s	-1.95 s	Binance leads Kraken

The profile is monotone in magnitude. This matters. A single isolated negative estimate could be a local artifact, but a smooth scale profile is more consistent with a real cross-venue information transmission pattern. The coarser-scale estimates indicate that the Binance-Kraken adjustment is not only a millisecond-level phenomenon. It persists into the multi-second horizon.

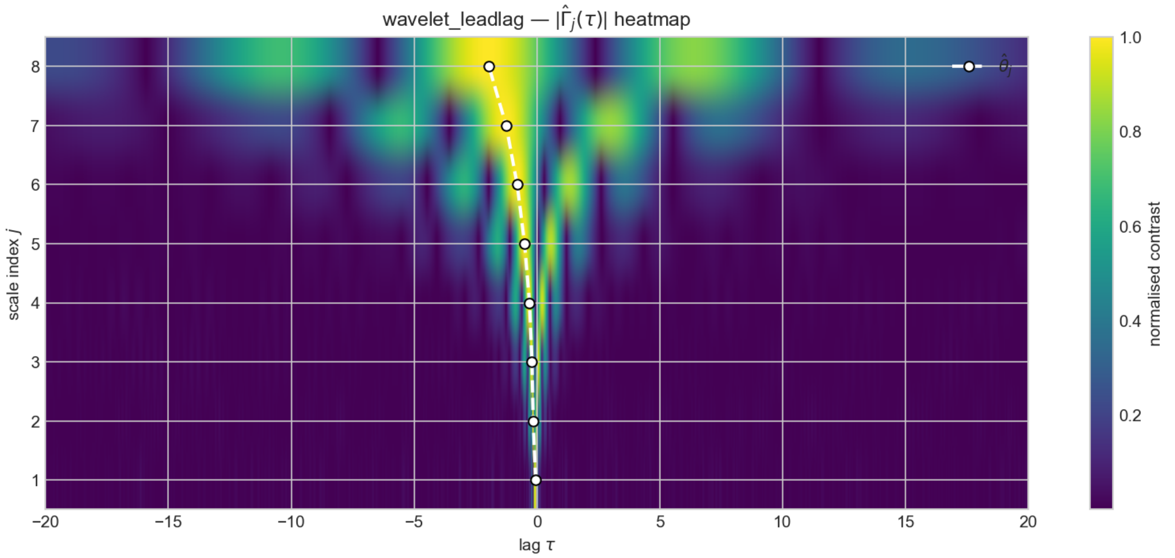


Figure 1: Filtered HK contrast heatmap for BTC Binance vs Kraken, April 2026. The ridge of maxima is located at negative lags across scales, which means Binance leads Kraken under the project sign convention.

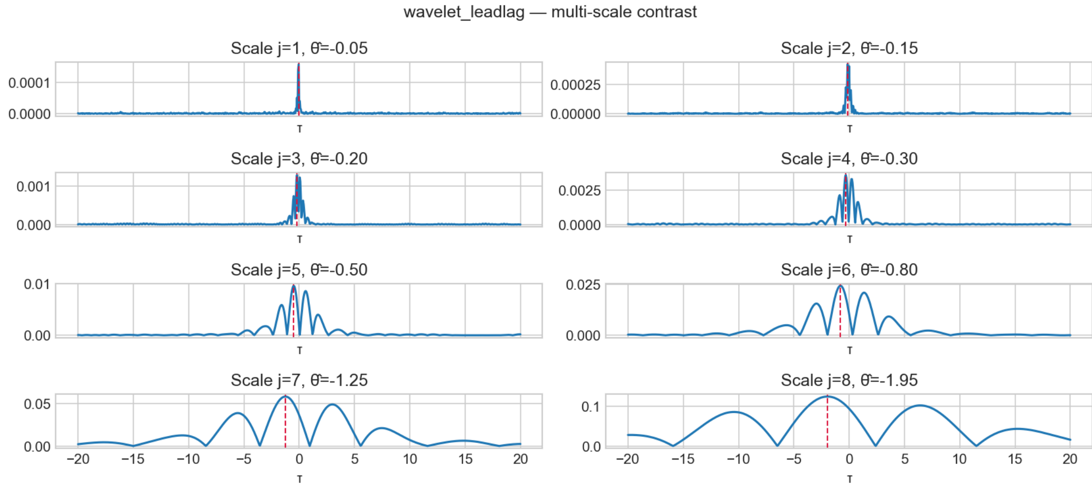


Figure 2: Scale-by-scale contrast curves for BTC Binance vs Kraken. The coarser scales display stronger and wider contrast peaks, consistent with slower cross-venue adjustment components.

6.2 Interpretation

Binance is the more active venue in this sample. In April 2026, the Binance BTC series has more than 12.6 million observations, while the Kraken series has about 0.86 million. The median inter-arrival time is approximately 32 ms on Binance and 1.39 s on Kraken. This difference does not by itself prove a lead-lag relationship, but it is economically consistent with Binance being closer to the active information layer.

The result should not be interpreted as a risk-free arbitrage opportunity. A 1.95 s coarser-scale lead does not mean that a trader can mechanically buy or sell Kraken after observing Binance. The estimator measures statistical dependence in historical public data. Deployability would depend on fees, latency, queue priority, slippage, inventory constraints, and the ability to trade both venues under the same capital and operational constraints.

7 Temporal Robustness

7.1 January, February, and April 2026

To test whether the BTC Binance-Kraken result is specific to April 2026, the same experiment is run on January and February 2026. The direction is invariant: all estimates are negative at all scales in all three months. The magnitude is not invariant, which is expected. Market activity, volatility, and data density change across months.

Table 6: Temporal robustness for BTC Binance vs Kraken. All values are in seconds.

Scale j	Jan. 2026	Feb. 2026	Apr. 2026
1	-0.05	-0.05	-0.05
2	-0.15	-0.15	-0.15
3	-0.20	-0.20	-0.20
4	-0.30	-0.30	-0.30
5	-0.50	-0.40	-0.50
6	-0.80	-0.55	-0.80
7	-1.30	-0.75	-1.25
8	-2.00	-1.15	-1.95

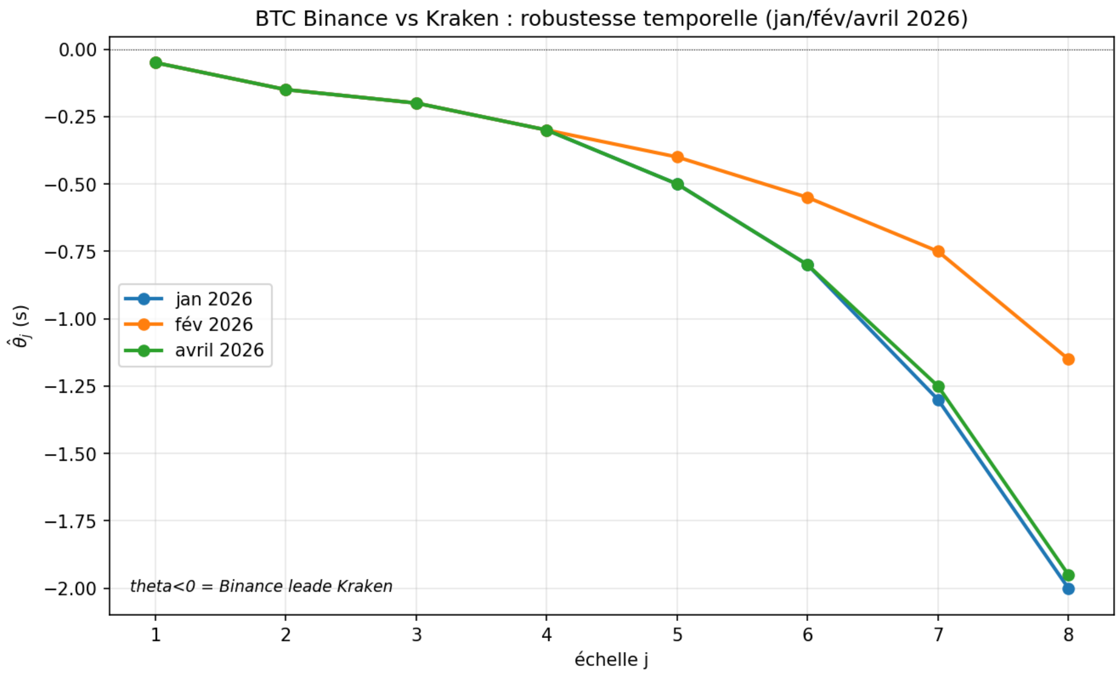


Figure 3: BTC Binance-Kraken lead-lag profiles across January, February, and April 2026. Direction is stable, while magnitude is regime-dependent.

7.2 What robustness means here

Robustness does not mean that the estimated lead is constant. It means that the economic ranking is stable under reasonable changes in the sample window. The coarser-scale magnitude is smaller in February than in January and April. This could reflect different volatility conditions, different exchange participation, or changes in trade intensity. The estimator reveals these differences rather than averaging them away.

8 Cross-Venue and Cross-Product Results

8.1 Full-month April 2026 panel

The next experiment compares multiple market relationships using the same settings. Table 7 reports the coarsest-scale estimate ($j = 8$), and Table 8 reports the full scale profiles.

Table 7: Cross-venue and cross-product panel, April 2026 full-month runs.

Setup	$\hat{\theta}_{j=8}$	Reading
BTC Binance vs Kraken	-1.95 s	Binance leads Kraken
ETH Binance vs Kraken	-0.30 s	Binance leads Kraken
BTC USDT-M perp vs COIN-M perp	-1.70 s	USDT-M perp leads COIN-M perp
BTC spot vs USDT-M perp	+0.35 s	USDT-M perp leads spot
ETH spot vs USDT-M perp	+0.40 s	USDT-M perp leads spot
BTC Binance vs Bybit	+0.20 s	Small Bybit lead at the coarsest scale

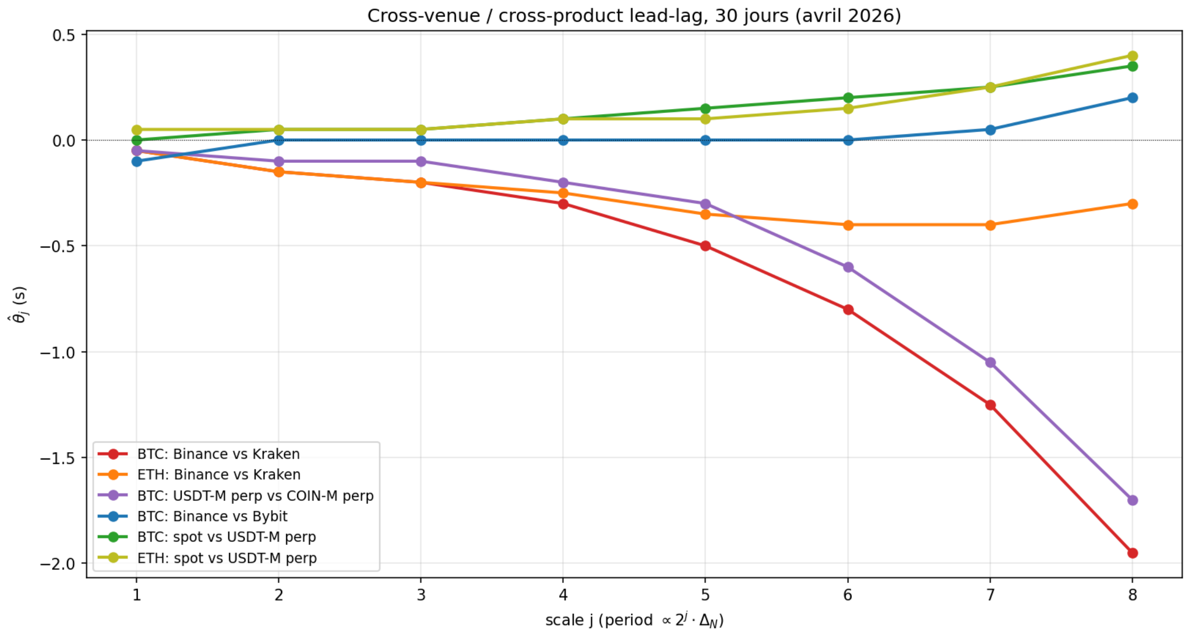


Figure 4: Full-month April 2026 comparison panel. The sign and slope of the scale profile differ by market relationship.

Table 8: Full scale profiles for the April 2026 panel. Values are $\hat{\theta}_j$ in seconds.

Setup	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$
BTC Binance vs Kraken	-0.05	-0.15	-0.20	-0.30	-0.50	-0.80	-1.25	-1.95
ETH Binance vs Kraken	-0.05	-0.15	-0.20	-0.25	-0.35	-0.40	-0.40	-0.30
BTC USDT-M perp vs COIN-M perp	-0.05	-0.10	-0.10	-0.20	-0.30	-0.60	-1.05	-1.70
BTC spot vs USDT-M perp	0.00	0.05	0.05	0.10	0.15	0.20	0.25	0.35
ETH spot vs USDT-M perp	0.05	0.05	0.05	0.10	0.10	0.15	0.25	0.40
BTC Binance vs Bybit	-0.10	0.00	0.00	0.00	0.00	0.00	0.05	0.20

8.2 Binance versus Kraken: BTC and ETH

The Binance-Kraken result appears for both BTC and ETH, but BTC displays a much larger coarser-scale lag. In ETH, the profile remains negative but flattens at coarser scales, ending at -0.30 s. This suggests that exchange-level price discovery is not only a function of the venue pair. It also depends on the asset, its liquidity, and the relative activity of market participants.

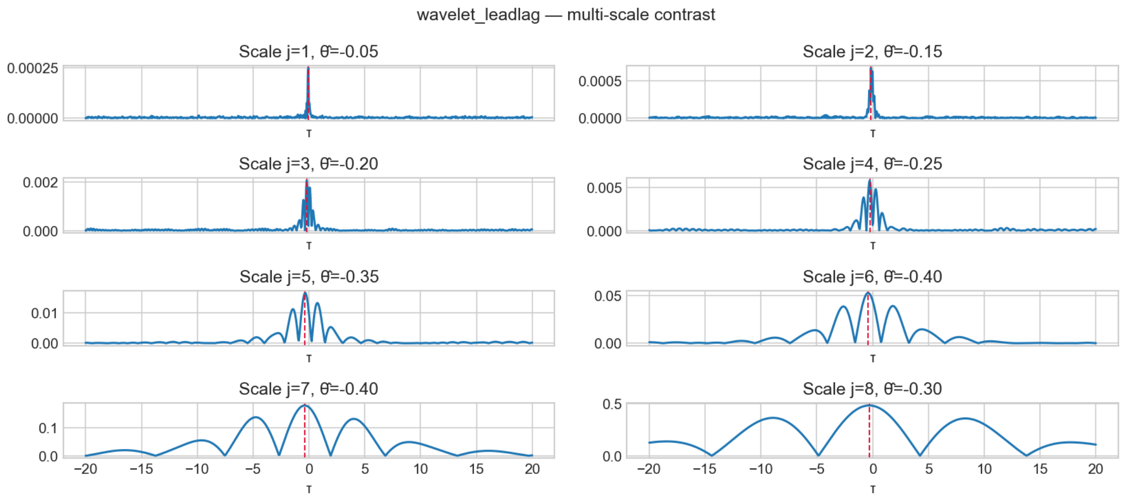
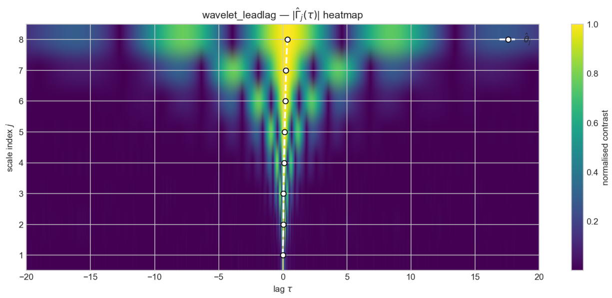


Figure 5: Filtered contrast curves for ETH Binance vs Kraken, April 2026. The direction is the same as BTC, but the coarser-scale magnitude is smaller.

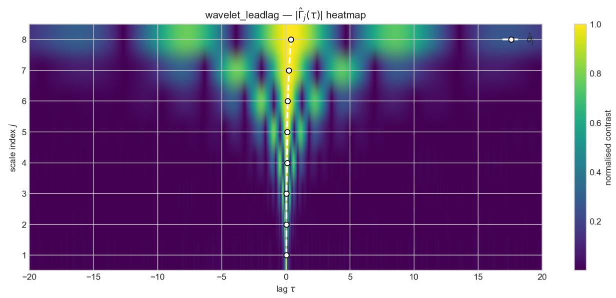
8.3 Perpetual futures versus spot

For BTC and ETH, the spot-versus-perpetual setup is ordered as spot first and perpetual second. Positive estimates therefore mean that the perpetual futures series leads the spot series. The estimates are small at fine scales and become larger at coarser scales. For BTC, the coarsest estimate is 0.35 s; for ETH it is 0.40 s.

This result is economically plausible. Perpetual futures are heavily used for leveraged directional exposure and hedging. They often concentrate informed or fast-moving order flow. A derivatives lead at multi-second horizons is therefore consistent with price discovery passing from leveraged markets to spot markets.



(a) BTC spot vs perpetual.



(b) ETH spot vs perpetual.

Figure 6: Spot-versus-perpetual heatmaps. Positive lags indicate that the second series, the perpetual contract, leads spot.

8.4 USDT-M versus COIN-M perpetuals

The BTC USDT-M versus COIN-M perpetual result is ordered with the USDT-M contract first and the COIN-M contract second. The estimates are negative across all scales and become more negative at coarser horizons, reaching -1.70 s at $j = 8$. This means the USDT-M perpetual leads the COIN-M perpetual. A natural interpretation is that the more liquid, stablecoin-margined contract absorbs information before the coin-margined contract.

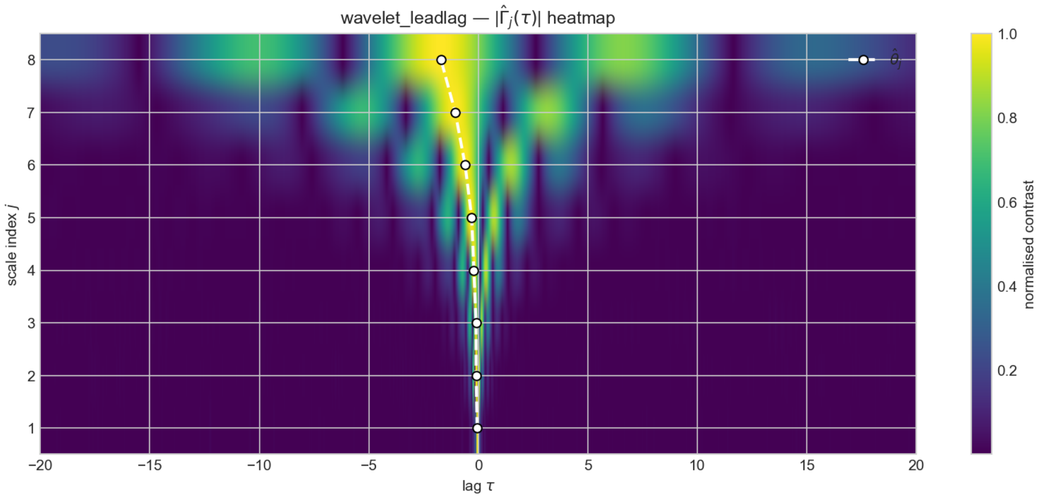


Figure 7: BTC USDT-M perpetual versus COIN-M perpetual, April 2026. The negative profile indicates that the USDT-M contract leads the COIN-M contract.

8.5 Binance versus Bybit

The Binance-Bybit BTC result is much closer to synchronous than the Binance-Kraken result. Most scales are exactly zero or close to zero, with a small Bybit lead at the coarsest scale. This contrast is useful because it shows that the estimator does not mechanically assign Binance a lead against every venue. The relationship depends on the pair.

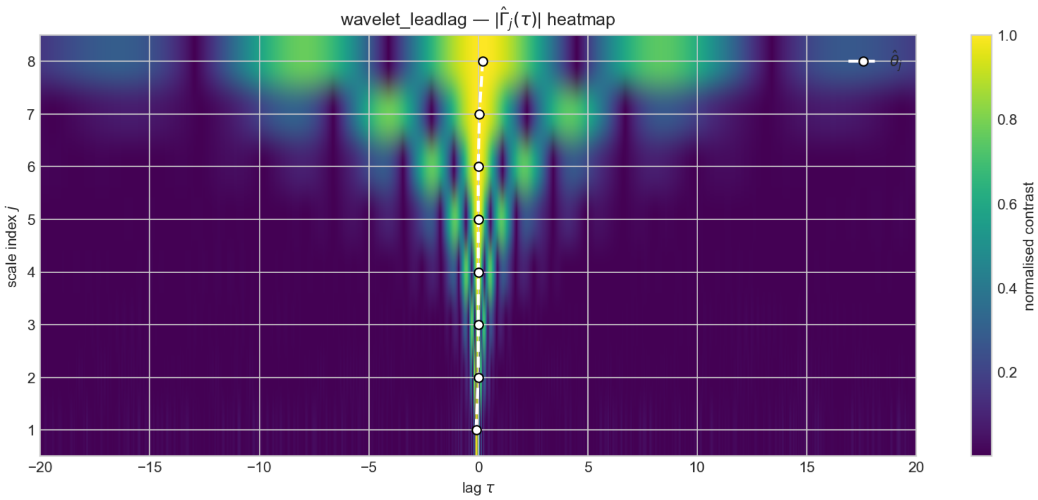


Figure 8: BTC Binance versus Bybit, April 2026. The profile is near zero through most scales, unlike the Binance-Kraken result.

9 Why Multi-Scale Estimation Matters

9.1 Comparison with HRY

The HRY estimator is a single-scale benchmark. It is useful because it tests whether the multi-scale estimator is providing information beyond a scalar lead-lag estimate. On the BTC/ETH spot pair over April 13–19, 2026, the HRY estimate is 0.00 s. The HK profile, however, is not uniformly zero: it becomes positive at coarser scales, reaching 0.25 s at $j = 8$.

Table 9: BTC/ETH spot pair, April 13–19, 2026: HK versus HRY.

Estimator	Scale	Period band	Estimate
HRY	single	aggregate	0.00 s
HK	$j = 1$	0.10–0.20 s	0.00 s
HK	$j = 2$	0.20–0.40 s	0.00 s
HK	$j = 3$	0.40–0.80 s	0.00 s
HK	$j = 4$	0.80–1.60 s	0.05 s
HK	$j = 5$	1.60–3.20 s	0.05 s
HK	$j = 6$	3.20–6.40 s	0.10 s
HK	$j = 7$	6.40–12.80 s	0.15 s
HK	$j = 8$	12.80–25.60 s	0.25 s

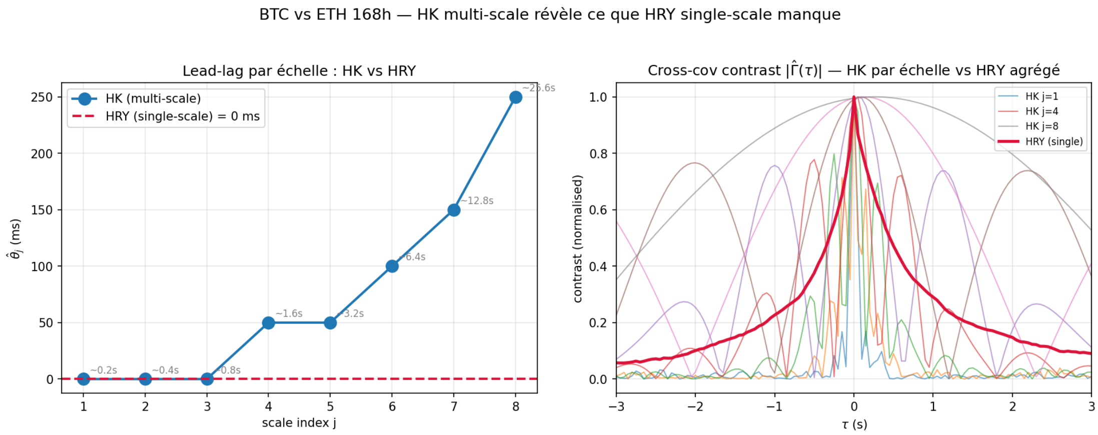


Figure 9: HK versus HRY on the BTC/ETH spot pair. The scalar HRY estimate is close to zero, while HK reveals a coarser-scale positive profile.

9.2 Interpretation

The BTC/ETH result illustrates the central methodological lesson. If the research object is a single trading delay, HRY may be sufficient. If the research object is information diffusion across horizons, HRY can be too compressed. A zero scalar estimate can hide the fact that one series leads at some scales and not at others. The multi-scale estimator acts like a diagnostic lens: it does not guarantee a stronger signal, but it reveals whether the scalar signal is stable across horizons.

10 Validation and Robustness Checks

10.1 Synthetic sign-convention tests

The sign-convention test generates two synthetic series where the true lead is known by construction. The goal is not to prove asymptotic validity; it is to ensure that the code's economic interpretation matches the implemented lag definition. This test is simple but crucial. Without it, one could report every venue ranking in the wrong direction.

The result of the audit is the convention already stated in (2): a negative estimate means that series 1 leads series 2. This convention is then used consistently in all result tables.

10.2 Monte Carlo diagnostic

The project includes a reduced-scale Monte Carlo diagnostic inspired by the simulation logic in the original paper. It compares the non-synchronous HK/HY implementation with a previous-tick interpolation approach under constant volatility and Heston-type volatility scenarios. The goal is to verify that the non-synchronous estimator behaves more sensibly than interpolation in controlled settings.

Table 10: Selected Monte Carlo diagnostic results. Values are in grid steps.

Scenario	Estimator	Scale	True θ_j	Median	Correct sign
constant, Lo-MacKinlay high	HK/HY non-sync	1	1	1	0.95
constant, Lo-MacKinlay high	HK/HY non-sync	2	1	1	1.00
constant, Lo-MacKinlay high	previous-tick	1	1	1	0.75
constant, Lo-MacKinlay high	previous-tick	4	2	-1	0.00
constant, moderate	HK/HY non-sync	1	1	1	1.00
constant, moderate	HK/HY non-sync	4	2	1	1.00
constant, moderate	previous-tick	6	5	0.5	0.50
Heston, Lo-MacKinlay high	HK/HY non-sync	5	3	2.5	1.00
Heston, Lo-MacKinlay high	previous-tick	5	3	-0.5	0.15

The diagnostic is intentionally modest in scale. It is not presented as a full replication of the paper's simulation section. Its role is to catch major implementation failures and to demonstrate the qualitative danger of previous-tick interpolation. In several selected cases, interpolation reverses the sign or produces much larger dispersion.

10.3 Tick-density artifact test

A possible objection to the Binance-Kraken result is that Binance is much denser than Kraken. The estimator could, in principle, be capturing a tick-density artifact rather than true price discovery. The project therefore includes a synthetic thinning test. A known lead is imposed, and the two series are randomly thinned at different keep probabilities.

Table 11: Tick-density artifact test. The true median lead is -3 grid units.

Keep series 1	Keep series 2	n_1	n_2	Median estimate
1.0	1.0	30,001	30,001	-3.0
1.0	0.2	30,001	5,967	-3.0
0.5	1.0	14,923	30,001	-3.0
0.2	0.2	5,967	6,034	-3.0
0.1	0.2	2,932	6,034	-2.5
0.1	0.1	2,932	2,996	2.0

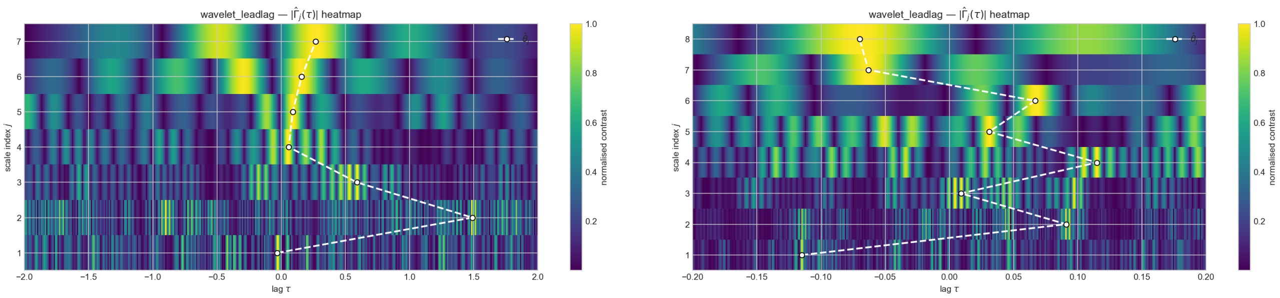
The sign remains correct under realistic thinning levels. It deteriorates only in the most extreme sparse-sparse case. This does not prove that all empirical results are immune to data-density effects, but it supports the interpretation that moderate density imbalances do not mechanically flip the estimator.

10.4 Equity sanity check

The project also runs the estimator on LOBSTER sample data using midpoint prices. This is important because the original Hayashi-Koike empirical setting is closer to quote data than to crypto transaction data. The equity results are not the main economic contribution of the project, but they verify that the code works on quote-derived price processes.

Table 12: LOBSTER equity sanity checks. Values are in seconds.

Pair	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$
SPY vs AAPL	-0.038	-0.053	-0.058	-0.059	-0.046	-0.119	-0.046	-0.085
AAPL vs MSFT	-0.115	0.091	0.009	0.115	0.031	0.067	-0.063	-0.070



(a) SPY vs AAPL.

(b) AAPL vs MSFT.

Figure 10: LOBSTER midpoint sanity checks. These exercises confirm that the estimator is not tied to crypto transaction data.

11 Additional Empirical Diagnostics

11.1 BTC/ETH cross-asset and proxy checks

The project also studies BTC/ETH spot relationships and synthetic proxy constructions. These checks are less directly interpretable than same-asset cross-venue tests because BTC and ETH are different assets. A measured lead can reflect common crypto-market shocks, differences in liquidity, or asset-specific news. Nevertheless, the exercise is useful because it tests whether the estimator produces coherent profiles on highly active but non-identical assets.

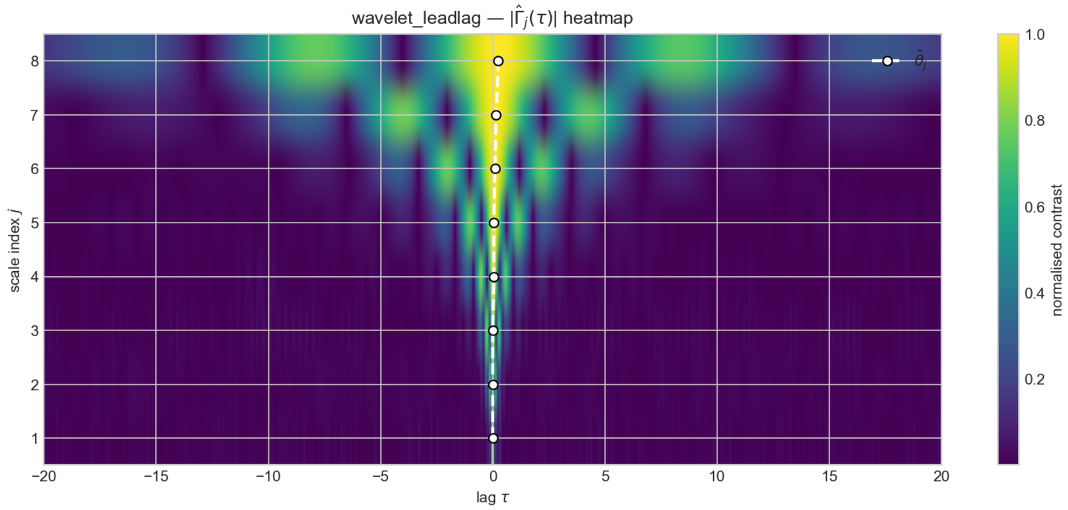


Figure 11: BTC/ETH spot heatmap over a 168-hour window. The profile is near zero at fine scales and positive at coarser scales.

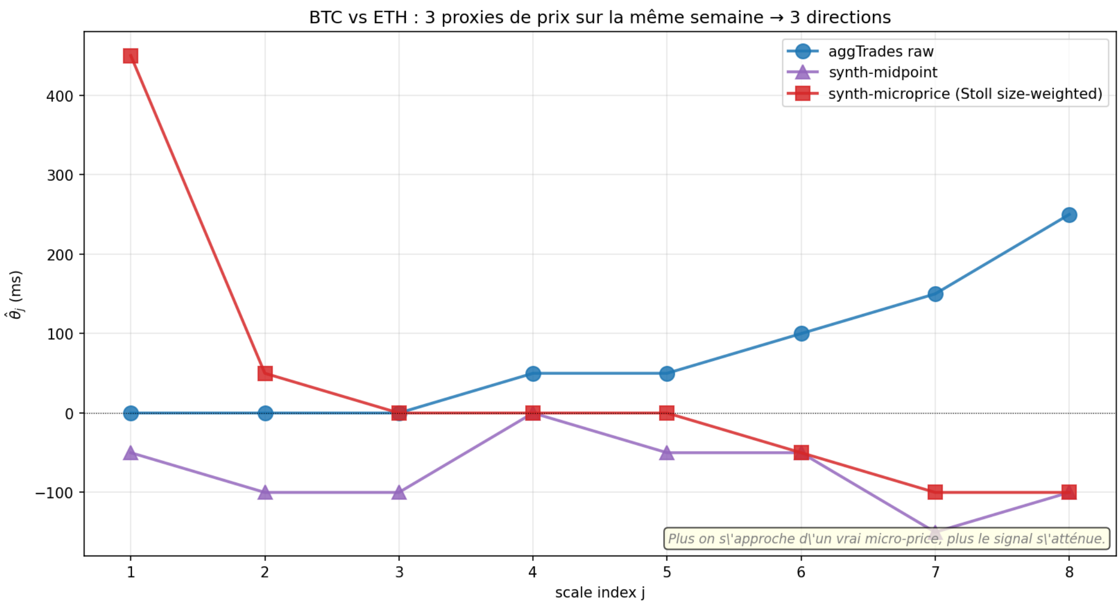


Figure 12: Proxy comparison for BTC/ETH experiments. Proxy choices affect magnitude and should be treated as part of the empirical design.

11.2 Market regime comparison

The repository includes a calm-versus-FOMC comparison. The purpose is not to make a causal macro announcement claim, but to show that lead-lag magnitudes can change across regimes. High-frequency relationships are conditional objects. They depend on volatility, participation, liquidity, and the flow of public information.

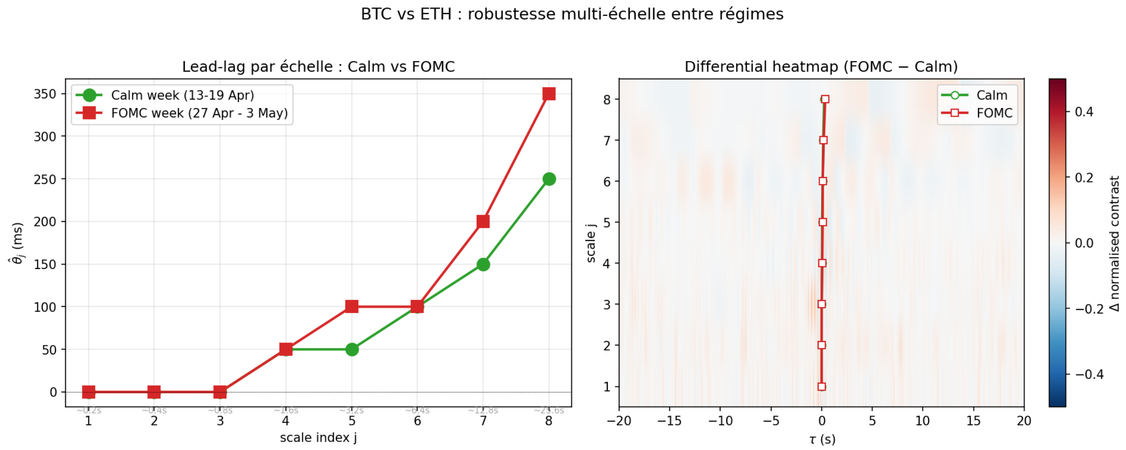


Figure 13: Calm versus FOMC-period comparison. Lead-lag profiles should be interpreted conditionally on market regime.

12 Discussion

12.1 What the results say

The strongest empirical conclusion is the Binance-Kraken BTC result. Across three full-month samples and eight dyadic scales, the sign is stable and the profile is economically coherent. Binance leads Kraken in the public transaction data studied here. The lead is small at fine scales and larger at coarser scales. This is precisely the kind of result that motivates a multi-scale estimator.

The cross-product results also have a clear interpretation. USDT-M perpetuals lead COIN-M perpetuals, and perpetual futures lead spot markets for both BTC and ETH at coarser scales. This aligns with the intuition that derivative markets can be the first layer of price discovery when leverage and hedging demand are concentrated there.

The Binance-Bybit result is useful as a negative control. It shows that the estimator does not always produce a large Binance lead. The profile is close to zero through most scales, with only a small coarser-scale Bybit lead. This supports the idea that the Binance-Kraken result is pair-specific rather than a mechanical artifact of putting Binance first.

12.2 What the results do not say

The results do not prove that a live trading strategy exists. A statistical lead in historical data is only one component of an executable strategy. A real strategy would need:

- exchange connectivity and latency measurements;
- taker and maker fee schedules;
- order-book depth and queue position;
- market impact and adverse selection estimates;
- funding rates and inventory constraints for derivatives;
- capital controls and operational constraints across venues.

The results also do not prove that the leading venue is causally generating price changes. The estimator measures timing of common information incorporation, not structural causality. A third latent factor can drive both markets, with one venue reacting earlier.

12.3 Limitations

The main limitations are the following.

(1) **Transaction prices rather than quotes.** Crypto experiments use trades because public quote histories are harder to standardize across venues. Trades can contain bid-ask bounce and order-flow effects.

(2) **Exchange timestamps.** Public data rely on exchange or API timestamps, not co-located capture timestamps. Network and publication delays can affect very fine-scale interpretation.

(3) **Public-data venue set.** The study covers a selected set of venues and pairs. A complete crypto price discovery study would include more exchanges and stablecoin pairs.

(4) **No execution model.** The project measures lead-lag profiles but does not model transaction costs, fill probabilities, or market impact.

(5) **Finite-sample filtering.** Wavelet filters are finite and the dyadic bands are approximate. Edge effects and support length matter, especially at coarser scales.

12.4 Possible extensions

Several extensions would make the project closer to a publishable empirical microstructure paper.

- Estimate confidence intervals by block bootstrap or subsampling across days.
- Use public order-book snapshots or proprietary quote data to compare trade-price and quote-price lead-lags.
- Add more venues and construct a network of scale-dependent price discovery.
- Condition estimates on volatility, funding rates, volume imbalance, and macro announcement windows.
- Compare HK to alternative estimators such as Dobrev-Schaumburg or spectral lead-lag methods.

13 Conclusion

This project implements and applies the Hayashi-Koike multi-scale lead-lag estimator in a public crypto-market setting. The methodological value of the estimator is that it respects two central features of high-frequency data: non-synchronous observation times and scale-dependent dependence. The empirical value is that it reveals economically interpretable price discovery profiles that a scalar estimator can hide.

The main empirical conclusion is that Binance leads Kraken on BTC across all tested scales in the public samples studied here. The lead is stable in sign across January, February, and April 2026, and it grows in magnitude at coarser scales. Cross-product results indicate that perpetual futures lead spot markets for BTC and ETH at multi-second horizons, while USDT-M perpetuals lead COIN-M perpetuals. A single-scale HRY benchmark can return a near-zero estimate where HK reveals a structured scale profile, confirming the importance of multi-resolution analysis.

The project should be read as a research implementation and empirical microstructure study, not as a trading system. Its main achievement is to turn a mathematical estimator from the literature into tested code, validated artifacts, and interpretable empirical results on modern public markets.

A Appendix A: Mathematical Details

A.1 Relationship between overlap covariance and lead-lag

If two latent price processes have maximal dependence when one is shifted by a lag θ, then the covariance of increments should peak near that shift. In regular sampling, one might estimate this by ordinary lagged covariance. In irregular sampling, the same idea must be written in terms of overlapping intervals. Equation (1) is the non-synchronous analogue. It sums the product of increments only when the shifted intervals overlap.

The estimator is robust to non-synchronicity because the overlap rule is endogenous to the sampling times. It does not require inventing unobserved prices. This is particularly important when one venue is much more active than another.

A.2 Filter support and extended lag grid

The level-j autocorrelation wavelet has length $2L_j - 1$. Therefore $\hat{\Gamma}_j(\tau)$ requires $\hat{U}_N(\tau - l\Delta_N)$ for $l = -(L_j - 1), \ldots, L_j - 1$. To compute all scales on the same search grid, the implementation extends the HY grid by the maximum support length:

$$L_{\max} = (2^J - 1)(L - 1) + 1.$$

For db10 and $J = 8$, $L = 20$ and $L_{\max}$ is large enough to cover the support of all filters used in the convolution. This is why the implementation first computes $\hat{U}_N$ on an extended grid and only then slices the valid segment for each scale.

A.3 Why use absolute contrast?

The argmax in (11) uses $|\hat{\Gamma}_j(\tau)|$ rather than $\hat{\Gamma}_j(\tau)$. This follows the objective of finding the strongest scale-specific dependence. A negative covariance peak can be just as informative as a positive one if the research question is timing. The sign of the lead-lag estimate comes from the lag location, not from the covariance sign.

B Appendix B: Reproducibility

B.1 Environment

The repository uses a standard Python environment with NumPy, pandas, PyWavelets, Numba, matplotlib, and pytest. The package can be installed in editable mode:

```
python -m venv .venv
source .venv/bin/activate
pip install -r requirements.txt
pip install -e .
```

The test suite is run with:

```
python -m pytest -q
```

At the time this paper was generated, the local test suite passed with 27 tests.

B.2 Core experiment command

The April 2026 BTC Binance-Kraken experiment can be rerun with:

```
python scripts/run_cross_exchange_kraken.py \
  --start 2026-04-01 --end 2026-04-30 \
  --binance-symbol BTCUSDT --kraken-pair XBTUSD \
  --name 2026-04_full_month_btc_binance_vs_kraken
```

B.3 Notebook

The main narrative notebook is `main.ipynb`. It is designed to run in artifact mode: it reads tracked summaries and figures, and it does not require raw tick archives for review.

C Appendix C: Additional Result Tables

Table 13: Selected contrast peaks for April 2026 full-month runs.

Setup	$j=2$	$j=4$	$j=6$	$j=8$
BTC Binance vs Kraken	0.000421	0.003624	0.024304	0.123796
ETH Binance vs Kraken	0.000672	0.005832	0.052992	0.484910
BTC USDT-M perp vs COIN-M perp	0.000455	0.003096	0.013928	0.053342
BTC spot vs USDT-M perp	0.000241	0.002362	0.009297	0.018291
ETH spot vs USDT-M perp	0.000652	0.004023	0.015553	0.035963
BTC Binance vs Bybit	0.000258	0.002768	0.014450	0.037150

Table 14: Per-day BTC Binance versus Bybit distribution, January 2026.

Scale	Median	Mean	Q25	Q75	Negative	Zero	Positive
1	-0.05	-0.250	-0.075	-0.050	93.55%	6.45%	0.00%
2	-0.05	-0.027	-0.050	0.000	61.29%	35.48%	3.23%
3	-0.05	-0.026	-0.050	0.000	54.84%	41.94%	3.23%
4	0.00	-0.023	-0.050	0.000	45.16%	48.39%	6.45%
5	-0.05	-0.031	-0.050	0.000	58.06%	35.48%	6.45%
6	-0.05	-0.047	-0.100	0.000	67.74%	16.13%	16.13%
7	-0.10	-0.045	-0.150	0.025	64.52%	9.68%	25.81%
8	0.00	-0.042	-0.200	0.125	45.16%	12.90%	41.94%

D Appendix D: Relation to the Original Paper

Hayashi and Koike [1] study multi-scale lead-lag relationships in high-frequency financial markets with an emphasis on equity market data. This project follows the estimator but adapts the empirical setting.

Table 15: Comparison between Hayashi-Koike (2020) and this project.

Dimension	Hayashi-Koike (2020)	This project
Market	NASDAQ/BATS equities	Crypto venues and products
Price proxy	Quote micro-price	Transaction log-price for crypto; midpoint for LOBSTER checks
Observation issue	Non-synchronous quote updates	Non-synchronous trade timestamps and quote-derived samples
Resolution	Ultra-high-frequency equity data	50 ms crypto lag grid; 1 ms LOBSTER grid
Main objective	Study multi-scale price discovery in equities	Implement estimator and test public crypto price discovery
Validation	Simulation and empirical equity panel	Synthetic audits, Monte Carlo diagnostics, tick-density tests, equity sanity checks

References

[1] Hayashi, T. and Koike, Y. (2020). *Multi-scale analysis of lead-lag relationships in high-frequency financial markets*. Journal version and arXiv preprint arXiv:1708.03992.

[2] Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2), 359–379.

[3] Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, 19(2), 426–461.

[4] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM.

[5] Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B*, 62(2), 271–292.

[6] Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1–2), 181–211.