When Is a Cross-Venue Funding-Rate Spread Worth Trading?

Public-Data Evidence on Forecastability and Cost Coverage

Hang Liu ů humarb Research, London, United Kingdom ů https://humarb.ai ů ORCID 0009-0008-1059-3926

3 November 2026

Abstract

Cross-venue perpetual-futures funding spreads are often presented as arbitrage opportunities, yet a quoted differential is neither a forecast nor net trading income. We study when settled funding differences can cover four-fill execution costs. Public APIs supply 121,864 funding records for eight major contracts on Gate, Hyperliquid, and Pacifica, aligned into 521 eight-hour blocks from May to November 2026; current public books on those venues and Paradex provide a separate cost snapshot. The median absolute spread is 0.45 bp per eight hours and the 99th percentile is 2.99 bp. Out of sample, a capped forecast reduces MAE from 0.430 to 0.399 bp relative to last-value persistence, with significant Diebold–Mariano improvements. This statistical gain does not create economic feasibility: no candidate generates a 24-hour entry at a 34 bp hurdle, while sparse 72-hour entries at 20 bp lose money in the funding-cash proxy. Probability forecasts underperform a training-frequency baseline. The evidence supports a break-even discipline rather than a claim of persistent arbitrage.

Keywords: perpetual futures; funding-rate spreads; transaction costs; forecast evaluation; crypto market microstructure

JEL classification: G12; G13; G14; C58

Evidence status. All replay outcomes are fixed-notional funding-cash proxies, not complete trading P&L. Historical four-way order books, price/basis P&L, liquidation, transfer, and margin costs are not reconstructed. Current-book measurements are scenarios, not historical fills. This paper is research, not investment advice.

Contents

1 Introduction **1**

2 Related Literature **2**

2.1 Perpetual contracts, carry, and arbitrage 3

2.2 Execution and the economics of a round trip 3

2.3 Forecast comparison and decision evaluation 4

3 Institutional Background **4**

3.1 Funding as a venue-specific transfer 4

3.2 Discrete events and continuous accrual 5

3.3 Four executions and common quantity 5

3.4 From a quoted spread to a decision 6

4 Data **6**

4.1 Public acquisition and asset scope 6

4.2 Native records and coverage rules 7

4.3 Current order-book snapshot 8

4.4 Data quality and reproducibility boundary 8

5 Methods **9**

5.1 Paired differentials and information timing 9

5.2 A deliberately small predictor family 9

5.3 Forecast errors and Diebold–Mariano comparisons 10

5.4 Decision rules and cost scenarios 10

5.5 Probability evaluation 11

5.6 Data-snooping and robustness design 11

5.7 Notation summary 11

6 Results **11**

6.1 The gross spread is small before any forecast is made 11

6.2 Out-of-sample point forecasts improve modestly 13

6.3 The primary decision case produces no entries 13

6.4 Longer holding creates sparse signals, not profits 14

6.5 Tail-event probabilities show no improvement over the baseline 15
6.6 Current executable-cost snapshot 17

7 Robustness **18**
7.1 Chronological subsamples 18
7.2 Venue-pair and asset slices 18
7.3 Cost and holding-period sensitivity 18
7.4 Alternative estimators and selection control 19
7.5 Coverage and accounting checks 20

8 Limitations and evidence boundary **20**
8.1 What the historical replay identifies 20
8.2 Sample and external-validity limits 21
8.3 Statistical limits 21
8.4 Scope of the negative result 22

9 Conclusion **22**

10 Declarations **23**

A Replication and supplementary results **26**
A.1 Deterministic reconstruction sequence 26
A.2 Supplementary forecast comparisons 26
A.3 Complete decision grid 27
A.4 Probability-score definitions 27
A.5 Data dictionary summary 28
A.6 Checklist for interpreting a new spread 28

1 Introduction

A perpetual-futures funding rate is a visible number with an unusually strong visual claim on attention. Two venues display different rates for what appears to be the same underlying asset; taking the difference seems to reveal a cash stream available to a trader who is long one contract and short the other. The inference is attractive because the position can be approximately price-neutral in base units while funding continues to accrue. It is also incomplete. A displayed differential can refer to different accrual intervals, different stages of a settlement cycle, or different contract and collateral conventions. Even after those differences are normalized, the observation remains a description of a past or currently indicated rate rather than a promise about the rates that will prevail while a position is open. Most importantly, entering and exiting the two legs requires four executions. Fees, spread crossing, depth consumption, and timing risk can exceed the funding difference by an order of magnitude.

This paper asks a deliberately narrow question: *when does a cross-venue funding-rate spread become large enough, persistent enough, and predictable enough to cover a stated round-trip cost?* The question joins three filters that are often reported separately. The first is mechanical break-even: the absolute settled differential must be compared with costs on the same notional and horizon. The second is directional persistence: a trader must choose which venue to long and which to short before future rates are known. The third is statistical selection: a rule that appears attractive after searching many estimators, assets, venues, and horizons cannot be evaluated as though it had been specified only once. A spread can pass the first filter at a particular instant and still fail the second; a predictor can reduce average forecast error and still fail to identify the rare observations that matter economically.

The empirical design uses only public, unauthenticated exchange interfaces. The historical sample contains Gate, Hyperliquid, and Pacifica funding records for BTC, ETH, SOL, XRP, DOGE, ADA, LINK, and AVAX perpetual contracts. These eight assets were specified as established, commonly available contracts rather than selected by their replay profitability. Native eight-hour Gate records and native hourly records on the other two venues are aligned to a common UTC eight-hour grid. Hourly blocks are admitted only when all eight distinct hours are present. The raw acquisition contains 121,864 records. After pair construction and a nine-block warm-up, the panel has 12,504 pair–asset–block observations across 521 blocks from 11 May to 3 November 2026. The final 40% of blocks, comprising 5,016 panel observations, is reserved for out-of-sample evaluation. Current order books on the three historical venues and Paradex form a separate, explicitly non-historical execution-cost snapshot.

Three results organize the paper. First, gross spreads in this venue set are small relative to ordinary taker execution. The median absolute cross-venue differential is 0.45 bp per eight hours, the 90th percentile is 1.34 bp, and the 99th percentile is 2.99 bp. None of the 12,504 aligned observations reaches 20 bp. A one-settlement strategy therefore cannot cover even a 20 bp round-trip scenario anywhere in this sample, and multi-settlement holding is not an optional refinement but the central object of analysis.

Second, predictability is real in a limited statistical sense but does not imply economic feasibility. A capped rule, which retains the smaller magnitude of the last differential and its trailing three-day mean when their signs agree, lowers out-of-sample MAE from 0.430 bp under last-value persistence to 0.399 bp. RMSE falls from 0.666 to 0.634 bp. Diebold–Mariano tests using cross-sectionally averaged losses and Newey–West standard errors reject equal accuracy for both loss

functions. Yet all four candidate rules generate zero entries for the pre-specified 24-hour, 34 bp decision case. At the less demanding 72-hour, 20 bp case, only zero to seven entries occur depending on the estimator, and every entered trade has a negative funding-cash proxy after cost. A two-sided ex post reference identifies just two cost-covering windows, confirming both that opportunities are rare and that foresight cannot be relabelled as a trading rule.

Third, probability estimates do not rescue the decision. In a separate low-cost statistical scoring exercise with a 5 bp, 24-hour threshold, training-only logistic calibration yields an out-of-sample Brier score of 0.02211, slightly worse than the 0.02189 training-frequency baseline. Brier skill is -1.0%, with a moving-block 95% interval of $[-2.69\%, +0.04\%]$. Log loss is also worse, although the skill interval includes zero. Improved point forecasts of the signed differential therefore coexist with no demonstrated probability improvement for the economically relevant tail event. This distinction is central to interpreting modest forecast gains in highly anchored funding-rate data.

The paper contributes measurement rather than a trading strategy. It supplies a reproducible public-data panel; a strict settlement-coverage rule; paired forecast, decision, and probability evaluations; and a current depth-walked cost snapshot that refuses to infer historical execution from present liquidity. It also provides a negative result with a precise scope: within eight major contracts and three similarly designed funding mechanisms over roughly six months, forecastable variation is far too small to support a taker-cost cross-venue rule under the stated scenarios. This does not prove that cross-venue funding arbitrage is never profitable. It identifies conditions under which the evidence does not support the trade and shows how a small forecast improvement can be economically irrelevant.

The analysis is differentiated from two earlier humarb papers. Research 01 treats executable net carry as a measurement problem and documents route-level snapshot estimates under robust multi-day rates and depth-dependent costs (humarb Research, 2026a). Research 02 formalizes settled-history accounting, equal-base execution, and liquidation-aware net-carry framing (humarb Research, 2026b). Neither supplies an independent multi-month public-data forecast evaluation. The present paper takes their accounting distinctions as background and adds the time-series evidence: strict public settlement alignment, held-out MAE/RMSE, Diebold–Mariano tests, proper probability scores, data-snooping control, and venue, period, horizon, cost, and estimator robustness. The object is not another net-carry estimator but a test of whether the spread survives a decision process.

The remainder proceeds as follows. Section 2 locates the study in work on perpetual pricing, carry, execution, and forecast evaluation. Section 3 explains why funding formulas and settlement grids cannot be compared as raw prints. Section 4 defines the public sample and its selection limits. Section 5 states the accounting proxy, predictors, and inference. Sections 6 and 7 report results and robustness. Section 8 sets the evidence boundary before Section 9 concludes.

2 Related Literature

2.1 Perpetual contracts, carry, and arbitrage

Perpetual futures replace expiry-based convergence with recurring transfers linked to a reference price. Shiller (1993) introduces perpetual claims in a cash-settlement setting, while Ackerer et al. (2026) develops modern pricing results under alternative funding specifications. The pricing literature is important here because funding is part of contract design, not an exogenous coupon. Its level and time path depend on the premium measure, interest component, clamps, caps, and settlement convention. Two venues can therefore display different rates without creating an immediate riskless payoff, and two numerically identical rates can have different cash implications when their intervals or settlement marks differ.

The broader carry literature defines carry as the return component arising when the underlying price does not change (Koijen et al., 2018). For a cross-venue perpetual position, this interpretation is useful but incomplete. The position has funding cash flows on both legs and can also experience relative mark-price and basis movements. A neutral base quantity does not imply neutral margin, collateral, or liquidation exposure. Crypto carry has been documented as an economically meaningful return source (Schmeling et al., 2026), and arbitrage and efficiency studies show that crypto-derivative prices can contain exploitable dislocations while remaining sensitive to contract design and frictions (De Blasis and Webb, 2022; Alexander et al., 2024). These findings motivate testing funding spreads; they do not establish that every quoted spread is an implementable carry return.

Recent work moves closer to the forecasting question. Inan (2026) studies funding-rate predictability, emphasizing that future rates contain structure rather than being pure noise. Pindza (2026) analyzes centralized–decentralized funding arbitrage as a basis trade with venue and risk decomposition. The present paper shares the premise that rates can be predictable and that venue risk matters, but changes the unit of evaluation. The main object is not a full-sample strategy return. It is a chronological forecast and cost-coverage assessment in which an estimator is rewarded only if information available before the holding interval identifies the correct sign and sufficient magnitude.

2.2 Execution and the economics of a round trip

A two-leg funding position has four executions when both legs are opened and closed with marketable orders. If each leg has fixed quote notional N and the taker fee on venue v is ϕ_v, the fee component alone is approximately $2N(\phi_1+\phi_2)$ before considering differences between entry and exit prices. Base-tier schedules used in the current snapshot range from 0.040% to 0.075% per execution, producing a material fixed hurdle (Gate, 2026a; Hyperliquid, 2026a; Pacifica, 2026a; Paradex, 2026a). This component is observable, but it is not the complete cost. Marketable orders walk available levels, and depth differs by asset, venue, size, and time. Execution research shows why price impact and liquidity belong in the trading problem rather than in a footnote (Almgren and Chriss, 2001; Cont et al., 2014). Crypto microstructure evidence likewise documents substantial variation in liquidity and trading costs across venues and instruments (Almeida and Gonçalves, 2024).

Top-of-book spread is therefore an inadequate general cost measure. At a stated notional, the relevant price is a depth-walked volume-weighted average, and the exit walk must use the same base quantity as the entry. If one book cannot support the quantity, the route lacks an executable estimate at that size; it is not a zero-cost observation. This paper implements that

principle in a current public-book snapshot. Because historical order books are not available in the acquisition, the same principle cannot be applied retrospectively. The historical replay consequently subtracts explicit cost scenarios and is labelled a funding-cash proxy. Combining a current snapshot with historical rates as if both had been observed contemporaneously would create a false implementation record.

2.3 Forecast comparison and decision evaluation

Point forecast accuracy is assessed with MAE and RMSE and compared using the Diebold–Mariano framework (Diebold and Mariano, 1995). Funding blocks remain serially dependent even after cross-sectional averaging, so inference uses Newey–West covariance estimates (Newey and West, 1987). The loss comparison is kept separate from the decision result. A predictor can win on average by shrinking small, frequently observed errors while doing worse in the extreme right tail that determines whether a trade clears its cost. MAE and RMSE are thus necessary diagnostics, not substitutes for an economic loss function.

Binary cost-coverage probabilities require their own scoring rules. Brier score and log loss are strictly proper under their respective event definitions (Gneiting and Raftery, 2007). They evaluate the full predictive distribution through event probabilities rather than merely checking whether a point forecast is above a threshold. This is particularly useful when the target event is rare: a model can achieve low classification error by never predicting an opportunity and still provide uninformative probabilities. The relevant comparison is against a frequency baseline estimated only on training data, accompanied by a dependence-aware interval for skill.

Finally, searching estimators and decision settings creates a data-snooping problem. White's Reality Check evaluates the best candidate against a benchmark while accounting for the family of rules considered (White, 2000). The present candidate family is deliberately small and fixed before results are inspected: persistence, a three-day mean, a three-day median, and a capped combination. The held-out segment is not used to fit the probability model or choose a preferred rule. The Reality Check is applied to the active 72-hour decision family because the primary 24-hour case contains no entries and hence no positive performance to test. Together, these methods enforce a hierarchy: establish forecast accuracy, test the economic decision, score event probabilities, and only then consider whether any surviving result warrants a trading interpretation.

3 Institutional Background

3.1 Funding as a venue-specific transfer

A perpetual contract's funding mechanism transfers value between longs and shorts to restrain persistent departures from a reference market. Under the conventional sign, a positive rate means that longs pay shorts; a negative rate reverses the transfer. For a long position on venue l and a short position on venue s, the funding component is favourable when the long venue's rate is sufficiently below the short venue's rate. The position is not a single contract but a route with two accrual conventions, two collateral environments, and two execution stacks.

The three historical venues use related premium-and-interest designs but not identical settlement records. Gate documents minute-by-minute funding calculations, a time-weighted average

premium, an interest component, venue-specific caps, and settlement at the end of the applicable interval; contract intervals and parameters may change (Gate, 2026b). Hyperliquid pays funding hourly. Its interest component is 0.01% per eight hours, or 0.00125% per hour, and the hourly payment uses position size multiplied by oracle price and funding rate (Hyperliquid, 2026b). Pacifica also calculates and settles hourly, with an eight-hour interest reference divided by eight, frequent premium sampling, and an hourly average applied to open positions (Pacifica, 2026b). These details have two implications. Raw one-hour and eight-hour prints are not commensurate until aggregated, and a fixed-notional rate sum is not identical to exact cash on a fixed base quantity when oracle prices move within the block.

The common functional form can be written schematically as

$$F_t = P_t + \mathrm{clip}(I - P_t, -\Delta, \Delta), \tag{1}$$

where P_t is a premium measure, I an interest component, and Δ a clamp. Exact averaging, scaling, caps, and price inputs remain venue-specific. Equation (1) explains a regularity in the data: when the premium is small relative to the interest anchor and clamp, rates can remain near a baseline for long stretches. In the complete venue–asset blocks, the exact 1 bp per eight-hour anchor occurs in 15.2% of Gate observations, 29.3% of Hyperliquid observations, and 30.7% of Pacifica observations. Anchoring compresses ordinary cross-venue differences. It does not eliminate tail variation, but the tails relevant to a 20–60 bp cost hurdle are exceptionally rare in this venue set.

3.2 Discrete events and continuous accrual

Settlement timing changes the interpretation of an entry immediately before a timestamp. On a discrete venue, holding at the relevant snapshot can generate a full interval payment even if the position was opened shortly beforehand, subject to the venue's precise rules. On an hourly venue, each hour generates its own transfer. Paradex, included only in the current cost snapshot, uses continuous accrual through a funding index and publishes an eight-hour reference rate; its realized payment depends on elapsed time and index evolution (Paradex, 2026b). Treating its latest reference rate as a settled eight-hour observation would confuse an instantaneous quote with accumulated cash. This is why Paradex is excluded from the historical replay rather than forced into the same settled-record schema.

The common grid in this paper is an accounting grid, not an assertion that all venues settle every eight hours. Gate supplies one settled record per eight-hour event. Hyperliquid and Pacifica supply eight settled hourly records, whose rates are summed after coverage checks. The grid enables a consistent fixed-notional comparison while preserving the native data in the replication files. It does not reproduce exact account cash, which would additionally require base quantity, settlement or oracle marks, collateral conversion, and event-level position timing.

3.3 Four executions and common quantity

Consider opening a long on venue 1 and a short on venue 2, then closing both. The entry consumes asks on venue 1 and bids on venue 2; the exit consumes bids on venue 1 and asks on venue 2. Let q be the common base quantity and m_v a contemporaneous mid-price. For asks A_v and bids B_v, define $VWAP(A_v, q)$ and $VWAP(B_v, q)$ by walking levels until quantity q is

filled. The implementation-shortfall component is

$$C^{IS} = q[VWAP(A_1, q) - m_1] + q[m_1 - VWAP(B_1, q)] + q[m_2 - VWAP(B_2, q)] + q[VWAP(A_2, q) - m_2]. \quad (2)$$

Fees are charged on actual executed notional at each of the four fills, not on a single mid-price approximation. If any book cannot fill q, the route has no complete four-fill estimate at that size. Reducing q to the largest commonly fillable amount is permissible only if the resulting smaller size is reported; silently retaining the original notional would understate cost.

The snapshot uses USD 10,000 as an example per-leg budget, not as a production threshold. Gate's contract sizes are converted to base quantity using the public contract multiplier. Hyperliquid, Pacifica, and Paradex books are stored in base units directly. Entry and exit reuse the same current book by assumption, which isolates displayed depth from unobserved time variation. The result is a scenario for four fills, not a prediction of future exit liquidity. Books were requested sequentially, with a maximum within-route timestamp skew of 14.4 seconds, another reason not to describe the numbers as synchronized executable quotes.

3.4 From a quoted spread to a decision

Let $r_{a,v,t}$ be the settled eight-hour-equivalent rate for asset a on venue v in block t. For an ordered pair (u, v), define $d_{a,u,v,t} = r_{a,v,t} - r_{a,u,t}$. A positive realized d favours a long position on u and short on v; a negative value favours the reverse. A trader deciding before block t knows neither d_t nor its future cumulative sum. The economically relevant forecast is therefore signed and horizon-specific. Multiplying a favourable absolute spread by a holding period without forecasting sign can convert a noisy quote into fictitious carry.

The break-even condition for an H-block hold and cost fraction c is

$$\operatorname{sign}(\widehat{d}_{t,H}) \sum_{j=0}^{H-1} d_{t+j} > c, \quad (3)$$

where $\widehat{d}_{t,H}$ is a direction forecast formed before the first outcome block. The left side is a fixed-notional funding proxy. Complete trade P&L would add relative price and basis terms and subtract all remaining frictions. Equation (3) is consequently a necessary funding-cash condition under the accounting assumptions, not a sufficient condition for total profitability.

4 Data

4.1 Public acquisition and asset scope

The acquisition script uses public REST endpoints without credentials or private account data. Gate supplies settled funding rates, contract metadata, and order books through API v4 (Gate, 2026a). Hyperliquid supplies funding history and current level-two books through its public information endpoint. Pacifica supplies hourly funding history with cursor pagination and current books through its public market-data interface (Pacifica, 2026a). Paradex supplies current books for the supplementary snapshot. Requests record timestamps, parameters, response hashes, row counts, and failures. The acquisition completed on 3 November 2026 with zero recorded failures.

The compressed normalized funding file contains 121,864 rows and has the following SHA-256 hash:

`2b88d0723287f121c503f6c7d80288b80d31486e1fcb0d986bdcaad51cc3fb3d`.

The historical asset set is BTC, ETH, SOL, XRP, DOGE, ADA, LINK, and AVAX. These are eight established contracts available across the historical venues at acquisition, chosen to represent major and comparatively liquid markets rather than to maximize replay spread. This is not a point-in-time listing universe. Contracts that failed, were delisted, or were absent on 3 November cannot enter, and the three venues are survivors by construction. The paper therefore does not claim universe-wide estimates or freedom from survivorship bias. The narrow asset set is a deliberate trade-off: it permits complete native-history coverage and transparent reproduction while limiting external validity.

Gate's public funding-history interface provides only a trailing interval of approximately 180 days. The common aligned sample is bounded by that interface, even though the hourly venues return older observations. After nine blocks are reserved for predictor warm-up, the analysis spans 521 eight-hour blocks from 16:00 UTC on 11 May 2026 through 08:00 UTC on 3 November 2026. The chronological split occurs at 00:00 UTC on 25 August 2026: 312 blocks for training and 209 for testing. Every panel contains eight assets and three unordered venue pairs, so the test segment has $209 \times 8 \times 3 = 5{,}016$ observations.

4.2 Native records and coverage rules

The normalized history stores venue, asset, exchange timestamp in milliseconds, native rate, native period, and record kind. Gate rows are settled eight-hour rates. Hyperliquid and Pacifica rows are settled hourly rates. For hourly sources, timestamps are assigned to the hour containing the exchange event and then to the eight-hour block ending at 00:00, 08:00, or 16:00 UTC. Small milliseconds-after-hour offsets do not move a record into the next hour. Duplicate native timestamps and duplicate hour keys are removed before aggregation.

Coverage is fail-closed. An hourly venue–asset block requires exactly eight distinct native hourly observations; a Gate block requires one settled eight-hour observation. Missing hours are not interpolated, zero-filled, or replaced by current rates. Pair panels retain only blocks with valid rates on both venues. These controls prevent two subtle errors: counting duplicate events as extra cash and treating absent history as a zero rate. The complete native table contains 18,904 venue–asset eight-hour blocks before pair construction and warm-up. After lagged predictors become available, the three-pair panel contains 12,504 observations.

Summing hourly rates is a fixed-notional normalization. If N is the per-leg quote notional and $r_{a,v,h}$ is an hourly rate, the proxy for a block is $N \sum_{h=1}^{8} r_{a,v,h}$. Exact funding on a constant base position can instead depend on hourly oracle prices and quote or settlement currency. The replication data therefore preserve native rates and do not describe the sum as an account statement. The distinction is small for many quiet blocks but can become material during volatile periods, precisely when a trader would most want an exact reconstruction.

Table 1: Sample construction and descriptive statistics. The historical window is 11 May–3 November 2026; spread percentiles use the 12,504 warmed-up pair–asset–block observations.

Item	Value
Native public funding records	121,864
Complete venue–asset eight-hour blocks	18,904
Warmed-up pair–asset–block observations	12,504
UTC eight-hour blocks	521
Training / test blocks	312 / 209
Test pair–asset–block observations	5,016
Median absolute differential (bp per 8h)	0.450
90th percentile absolute differential (bp per 8h)	1.340
99th percentile absolute differential (bp per 8h)	2.989
Observations above 20 bp	0

4.3 Current order-book snapshot

The supplementary snapshot contains four venues, eight assets, and all six unordered venue pairs per asset, for 48 route measurements. Each book is converted to price–base-quantity levels. For each route, the script evaluates both venues' asks and bids, determines the common quantity supported at the USD 10,000 example budget, and computes four depth-walked executions and their actual-notional taker fees. Base-tier fee assumptions are 0.075% for all eight Gate contracts in the archived metadata, 0.045% for Hyperliquid, 0.040% for Pacifica, and 0.045% for Paradex Pro 0 (Gate, 2026a; Hyperliquid, 2026a; Pacifica, 2026a; Paradex, 2026a). Paradex retail orders can have zero explicit fees, but the Pro assumption keeps the API-oriented scenario conservative; depth dominates that venue's measured costs in any case.

This snapshot has a different estimand from the historical replay. It asks what the displayed books imply at acquisition, not what trades would have cost during each historical funding window. No current cost is applied to a past date as though it were observed then. Likewise, no historical funding spread is combined with a current book to claim an executable opportunity. The two datasets answer complementary but separate questions: the replay measures settled funding variation and predictability, while the snapshot illustrates the scale and cross-sectional dispersion of current execution cost.

4.4 Data quality and reproducibility boundary

The replication package includes the acquisition script, normalized public histories, aligned panel, out-of-sample decisions, contract metadata, current books, analysis script, and table-generating outputs. It also includes an offline accounting component with tests for equal-base four-fill accounting, insufficient exit depth, actual exit fees, settlement-mark requirements, event duplication, and missing coverage. Those tests check accounting logic; they do not validate exchange completeness or remove the sample limitations.

Three remaining data issues are substantive. First, public endpoints can revise history or change availability, so hashes and acquisition timestamps are necessary for exact reproduction. Second, the selected venues have similar premium-and-interest mechanisms, which limits the range of institutional designs and may compress differentials relative to a more heterogeneous set. Third, public books reveal displayed liquidity rather than hidden, proprietary, or immediately replenished depth. The snapshot can overstate or understate realized fills. These constraints are

reported as identification limits, not repaired with synthetic data.

5 Methods

5.1 Paired differentials and information timing

For each asset a and unordered venue pair (u, v), with venues alphabetically ordered, the realized differential is

$$d_{a,u,v,t} = r_{a,v,t} - r_{a,u,t}. \tag{4}$$

A forecast $\widehat{d}_{a,u,v,t}$ is constructed from differentials strictly before block t. If its sign is positive, the decision rule longs u and shorts v; if negative, it reverses the legs. A zero forecast means no position. The outcome for an H-block hold is

$$G_{a,u,v,t,H} = \operatorname{sign}(\widehat{d}_{a,u,v,t}) \sum_{j=0}^{H-1} d_{a,u,v,t+j}, \tag{5}$$

expressed as a rate per eight-hour-equivalent fixed notional. Multiplying by $N = 10{,}000$ gives the funding-cash proxy in US dollars under the fixed-notional assumption.

This timing convention prevents the just-settled block from predicting itself. Predictors at row t use observations through $t-1$, while outcomes begin at t. The decision is interpreted as occurring immediately after the $t-1$ settlement and before the first outcome block. Multi-block outcomes are non-overlapping within each train or test segment by evaluating entries only on every Hth block. The same entry grid is used for every estimator, preventing one rule from receiving a more favourable calendar.

5.2 A deliberately small predictor family

Four fixed predictors are evaluated. Persistence uses the last settled differential,

$$\widehat{d}_t^{\,\text{last}} = d_{t-1}. \tag{6}$$

The three-day mean and median use the previous nine eight-hour blocks,

$$\widehat{d}_t^{\,\text{mean}} = \frac{1}{9}\sum_{j=1}^{9} d_{t-j}, \qquad \widehat{d}_t^{\,\text{median}} = \operatorname{median}(d_{t-1}, \ldots, d_{t-9}). \tag{7}$$

The capped rule combines persistence and the three-day mean:

$$\widehat{d}_t^{\,\text{cap}} = \begin{cases} \operatorname{sign}(d_{t-1}) \min(|d_{t-1}|, |\bar{d}_{t-1:9}|), & d_{t-1}\bar{d}_{t-1:9} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The cap responds to disagreement by abstaining and otherwise shrinks magnitude toward the smaller of the recent and smoothed estimates. It is an example estimator chosen for its conservative interpretation, not a production rule or a claim of optimality. Nine blocks, four estimators, three holding horizons, and three cost scenarios are example research settings disclosed in full. The zero forecast is included as a point-error benchmark but not as an active trading candidate.

The predictor family is intentionally small to limit data mining. No lag length, threshold, or shrinkage weight is tuned on the test segment. The chronological split is made before evaluation, and the logistic event model is fitted on training observations only. The paper reports all candidates rather than selecting the most attractive decision result ex post.

5.3 Forecast errors and Diebold–Mariano comparisons

For predictor m, the signed error is $e_t^m = d_t - \widehat{d}_t^m$. Out-of-sample MAE and RMSE are

$$MAE_m = \frac{1}{n}\sum_t |e_t^m|, \qquad RMSE_m = \sqrt{\frac{1}{n}\sum_t (e_t^m)^2}, \tag{9}$$

reported in basis points per eight hours. Sign accuracy is descriptive because zero and near-zero rates make it an incomplete economic measure.

Diebold–Mariano comparisons use loss differentials against persistence. For each timestamp, the loss differential is first averaged across the 24 pair–asset panels, reducing the influence of cross-sectional duplication. The resulting time series is tested with a Newey–West variance estimate and 18 lags, corresponding to six days of eight-hour observations (Diebold and Mariano, 1995; Newey and West, 1987). Both absolute and squared losses are reported. A significant difference is interpreted as forecast accuracy only; it is not evidence of positive net cash.

5.4 Decision rules and cost scenarios

For holding blocks $H \in \{1, 3, 9\}$, corresponding to 8, 24, and 72 hours, the forecasted gross magnitude is $H|\widehat{d}_t|$. A candidate enters only when

$$H|\widehat{d}_t| > c, \tag{10}$$

where $c \in \{20, 34, 60\}$ bp is a round-trip cost scenario. The realized proxy after cost is

$$\pi_{t,H,c} = \begin{cases} G_{t,H} - c, & H|\widehat{d}_t| > c \text{ and } \widehat{d}_t \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The 24-hour, 34 bp case is designated primary because it requires more than one settlement while remaining close to the scale of four-fill costs observed in the current core-venue snapshot. The other cases are sensitivity analyses, not alternative definitions selected to obtain a favourable result.

Reported decision statistics include the number of entries, entry share, win rate among entries, mean gross and net basis points per entry, and descriptive net-cash sums across independent routes. These sums are not portfolio returns. Routes share legs, overlap in economic exposure, and require capital and margin on multiple venues; no account-level equity curve or capital denominator is inferred. No-trade decisions are included in Reality Check means so that abstention is not treated as a successful active trade.

A two-sided ex post reference enters whenever $|D_{t,H}| > c$ and assigns the favourable realized direction, where $D_{t,H} = \sum_{j=0}^{H-1} d_{t+j}$. This reference deliberately uses future information. It measures how many windows could cover cost in principle and is never called a feasible strategy,

an oracle return, or an attainable upper bound after accounting for all risks.

5.5 Probability evaluation

The economically relevant event is $Y_{t,H,c} = \mathbf{1}(G_{t,H} > c)$. Probability calibration uses the three-day median predictor, a 24-hour horizon, and a 5 bp threshold. This low-cost threshold is necessary because the 34 bp primary case has zero training events, making a two-class probability model unidentified. The 5 bp exercise is therefore a statistical diagnostic, not an executable-cost claim.

A one-feature logistic model is fitted on training data:

$$\Pr(Y_t = 1 \mid x_t) = \text{logit}^{-1}(\alpha + \beta x_t), \qquad x_t = \log \frac{H|\widehat{d}_t| + \varepsilon}{c}. \tag{12}$$

Out-of-sample Brier score and log loss are compared with a constant probability equal to the training event frequency. Brier skill is $1 - BS/BS_{base}$. A moving-block bootstrap uses 21 eight-hour blocks, 2,000 replications, and a fixed seed to obtain a 95% interval for skill. Decision-level Brier losses are averaged by timestamp before resampling, preserving the same target as the reported score. Calibration bins are descriptive and do not affect model fitting.

5.6 Data-snooping and robustness design

White's Reality Check is applied to the four active estimators in the 72-hour, 20 bp case. The benchmark is zero net cash for all decisions. Pair–asset net outcomes are averaged by timestamp, and a centered moving-block bootstrap preserves short-run dependence. The reported p-value accounts for selecting the best member of this small candidate family. The 24-hour primary case is not used for this test because all candidates abstain and there is no active performance to compare.

Four robustness dimensions accompany the main results. Chronological subsamples split the test segment into early and late halves. Venue-pair slices re-estimate decision outcomes for each of the three pairs. Parameter sensitivity varies holding horizon and cost hurdle. Alternative estimators compare persistence, mean, median, and capped forecasts on both statistical and economic criteria. Asset slices are reported as descriptive diagnostics with small entry counts, not as independent proofs. The replication tables retain empty and undefined statistics rather than replacing them with zero.

5.7 Notation summary

6 Results

6.1 The gross spread is small before any forecast is made

Figure 1 shows the distribution of absolute paired differentials on the common eight-hour grid. The median is 0.45 bp, the 75th percentile is 0.88 bp, and the 90th percentile is 1.34 bp. Even the upper tail remains narrow in execution-cost units: the 95th percentile is 1.735 bp and the 99th percentile is 2.989 bp. No aligned observation exceeds 20, 34, or 60 bp. The result is stronger than a statement that the average spread is unattractive. In this sample, the entire

Table 2: Principal symbols used in the empirical design.

Symbol	Definition
$r_{a,v,t}$	Settled eight-hour-equivalent rate for asset a at venue v
$d_{a,u,v,t}$	Paired differential $r_{a,v,t} - r_{a,u,t}$
$\widehat{d}_t$	Forecast formed from blocks strictly before t
H	Non-overlapping holding horizon in eight-hour blocks
N	Example fixed notional, USD 10,000 per leg
c	Round-trip cost scenario in basis points
$G_{t,H}$	Signed realized cumulative funding differential over H blocks
$\pi_{t,H,c}$	Funding-cash proxy after the scenario cost
$Y_{t,H,c}$	Indicator that correctly directed cumulative funding exceeds cost

observed distribution lies below the lowest round-trip cost scenario for a one-block holding period. For scale only, mechanically repeating the median absolute differential over 1,095 eight-hour blocks gives 4.93% of fixed notional per year before direction, interruption, and all costs; the corresponding 99th-percentile scale is 32.73%. These annualized magnitudes are not returns, because absolute spreads have no assured direction and the fixed-notional proxy has no capital denominator.

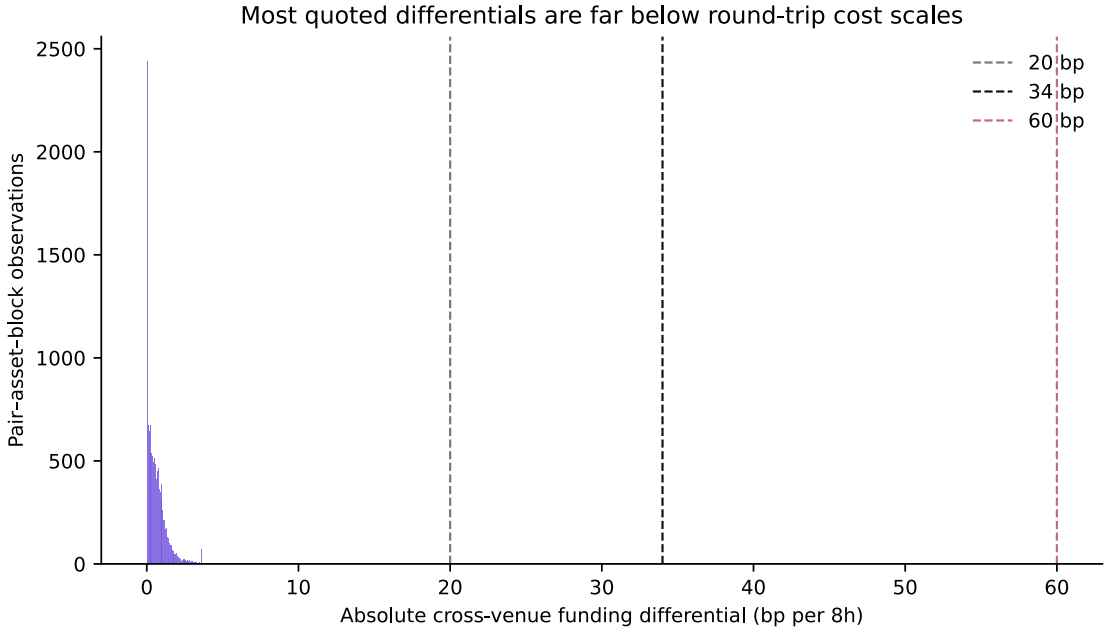


Figure 1: Distribution of absolute cross-venue funding differentials. Rates are aligned to UTC eight-hour blocks and expressed in basis points per eight hours. Vertical references mark the 20, 34, and 60 bp cost scenarios; all observed absolute spreads lie below them.

The settlement mechanisms help explain this compression. Exact 1 bp per-eight-hour-equivalent rates account for 15.25% of Gate venue–asset blocks, 29.30% of Hyperliquid blocks, and 30.73% of Pacifica blocks. These shares are descriptive rather than proof that a particular formula component binds at every such observation. They nevertheless show why a large fraction of paired observations is small, discrete, and clustered around a common anchor. When both sides of a pair sit at the same rate, the differential is exactly zero; when one side moves away, the first increment is still tiny relative to four-fill execution.

This evidence changes the economic question. A scanner that ranks contemporaneous rate differences is largely ranking variation within a few basis points, not identifying stand-alone arbitrage. A position would need either much lower execution costs than the taker scenarios, a much larger future spread than the observed current one, or persistent collection over several settlements. The first possibility is outside the stated taker-cost experiment. The second and third are forecast questions and are tested directly below.

6.2 Out-of-sample point forecasts improve modestly

Table 3 reports held-out point-forecast performance over 5,016 pair–asset–block observations. Last-value persistence is a strong benchmark: its MAE is 0.430 bp and RMSE is 0.666 bp, both better than a zero forecast. The trailing three-day mean and median do not improve on persistence under either loss. The capped estimator does. Its MAE is 0.399 bp and RMSE is 0.634 bp, reductions of approximately 7.1% and 4.8% relative to persistence. These gains are useful but small in levels: about 0.031 bp of MAE and 0.032 bp of RMSE per eight hours.

Table 3: Out-of-sample point-forecast accuracy, in basis points per eight hours. The test segment contains 209 blocks and 5,016 panel observations. Bias is the mean forecast minus realization; sign accuracy is descriptive.

Estimator	MAE	RMSE	Bias	Sign accuracy	n
Zero	0.5397	0.8118	−0.3383	0.1946	5,016
Last value	0.4296	0.6658	−0.0001	0.7444	5,016
Three-day mean	0.4589	0.6768	−0.0008	0.6820	5,016
Three-day median	0.4493	0.6963	−0.0173	0.6842	5,016
Capped	**0.3988**	**0.6340**	−0.0985	0.6974	5,016

The Diebold–Mariano results distinguish this improvement from sampling noise under the stated dependence correction. Against persistence, the capped estimator has an absolute-loss statistic of -4.510 with $p = 0.0000065$, and a squared-loss statistic of -2.561 with $p = 0.0104$. Negative statistics favour the capped rule. The three-day mean and median are less consistent: their absolute-loss comparisons significantly favour persistence, while their squared-loss comparisons are not significant. The capped result is therefore not simply an artifact of reporting one convenient loss function, although it remains specific to this estimator, sample, and test design.

Two features limit the economic interpretation. First, persistence has the highest sign accuracy at 74.44%, even though the capped rule has lower point errors. Shrinking magnitude and abstaining under sign disagreement can reduce loss without improving directional classification. Second, a predictor can be most accurate near the centre of a tightly anchored distribution while remaining uninformative about the rare large cumulative differentials that determine trade entry. The decision and probability results show that both concerns matter.

6.3 The primary decision case produces no entries

The 24-hour, 34 bp case makes the separation between forecast accuracy and economic content especially clear. Each estimator faces 1,656 non-overlapping test decisions. Last value, the three-day mean, the three-day median, and the capped rule all produce zero entries. The ex post two-sided reference also finds zero windows whose realized cumulative absolute differential exceeds 34 bp. Thus no forecast-based entry is suppressed merely by an excessively conservative

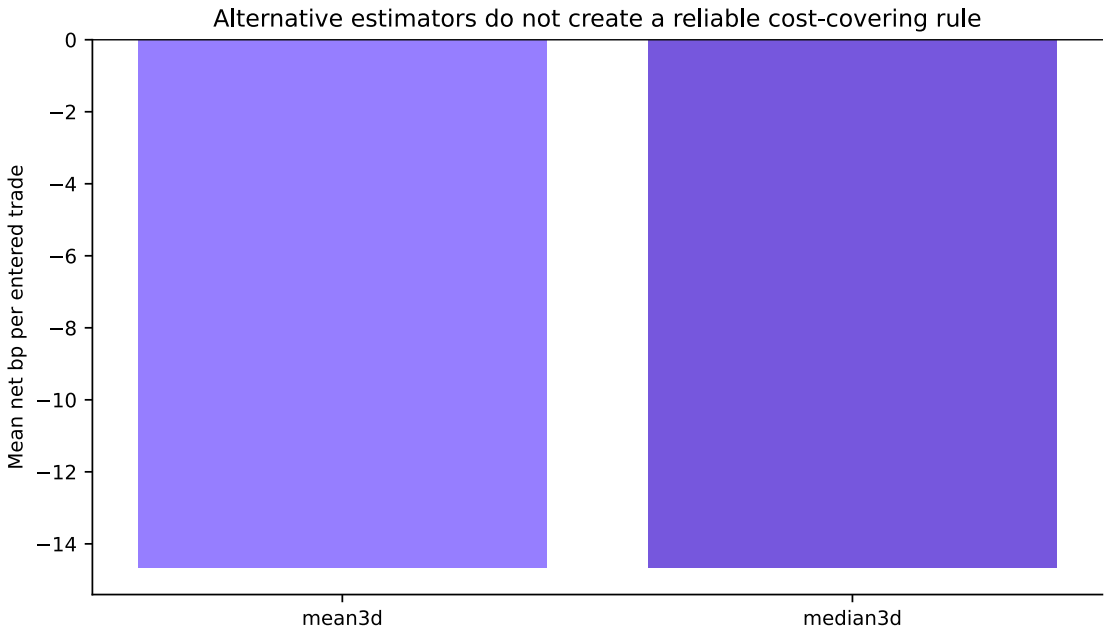


Figure 2: Mean net funding-cash proxy per entered trade by estimator in the 72-hour, 20 bp case. The capped estimator has no entries and therefore no defined per-trade bar. Statistical forecast improvements do not become a reliable cost-covering rule.

rule: there is no realized cost-covering window in this decision case to identify.

All 8-hour cases at 20, 34, and 60 bp likewise have zero entries and zero ex post cost-covering windows. The same is true of every 24-hour case, including the lowest 20 bp hurdle. These are not missing results. They are the empirical answer under the pre-specified break-even rule. Assigning a zero return to an abstention is appropriate for cash accounting, but it must not be misread as a successful trading strategy or as evidence of positive risk-adjusted performance.

6.4 Longer holding creates sparse signals, not profits

Relaxing the holding period to 72 hours and lowering the scenario cost to 20 bp generates a small number of entries. Persistence enters seven times among 552 decisions, the three-day mean enters six times, and the median enters three times; the capped rule abstains throughout. Every entered position has a negative funding-cash proxy after cost. Mean net results per entry are −8.25, −14.67, and −14.67 bp for persistence, mean, and median, respectively. Their descriptive sums across independent route observations are −$57.72, −$88.04, and −$44.01 for fixed $10,000 notional per leg. These sums are not portfolio P&L and have no common capital denominator.

The ex post reference enters only two windows, both with the correct direction imposed after realization, for a descriptive proxy sum of $4.28 after the 20 bp scenario cost. The contrast is instructive. Rare favourable windows exist in this relaxed case, but the information available at decision time does not select them reliably. Some forecast-selected windows realize little or adverse cumulative carry, while some favourable windows are not selected. This is a directional timing failure, not merely an insufficiently precise estimate of a known cost.

At a 34 bp hurdle over 72 hours, only the three-day mean enters, with two trades, a zero win rate, and mean net proxy of −28.53 bp. At 60 bp, all candidates abstain. White's Reality Check

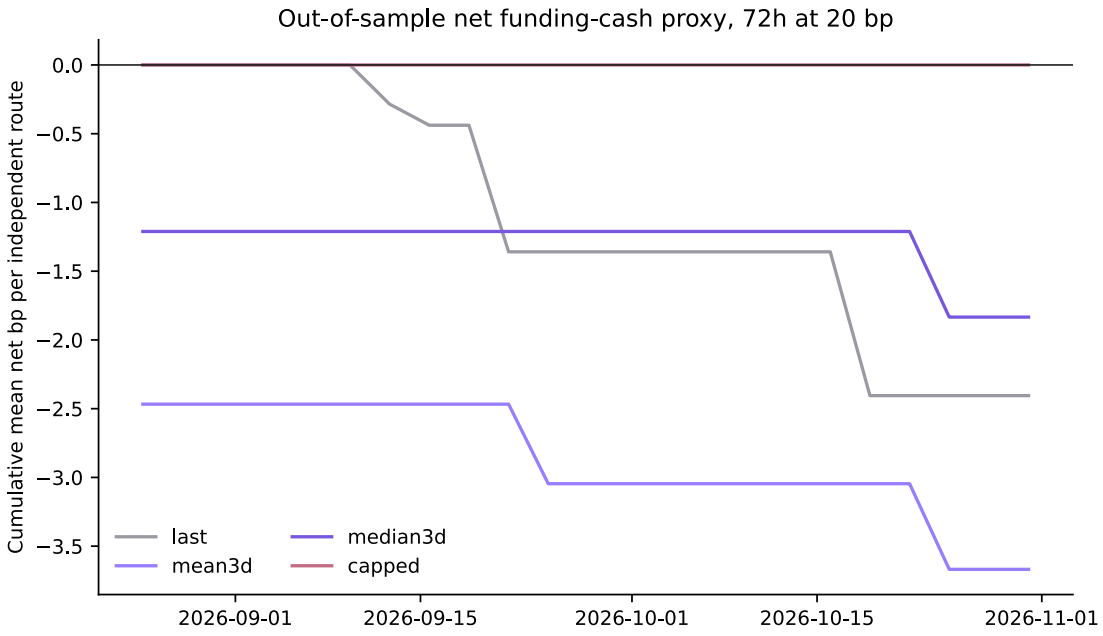


Figure 3: Out-of-sample cumulative funding-cash proxy in the 72-hour, 20 bp case. Lines show the cumulative cross-sectional mean net basis points per independent route for each estimator. Values are fixed-notional proxies, not complete trading P&L or portfolio returns.

for the 72-hour, 20 bp candidate family yields a best mean net outcome of zero, attained by the always-abstaining capped candidate, and a bootstrap p-value of 1.0 against the zero-cash benchmark. There is accordingly no evidence that selecting the best of the four active rules produces improvement after accounting for this limited search.

6.5 Tail-event probabilities show no improvement over the baseline

The 24-hour, 5 bp statistical scenario contains 37 positive events among 1,656 test decisions, an event rate of 2.234%. The logistic model produces a Brier score of 0.022107, compared with 0.021886 for the constant training-frequency forecast. Brier skill is −0.01011, and the moving-block 95% interval is $[-0.02688, 0.00044]$. Log loss is 0.11929 for the model and 0.10784 for the baseline. Both proper scores therefore prefer the simple frequency benchmark on their point estimates (Gneiting and Raftery, 2007). The Brier-skill interval includes zero, so the evidence does not statistically separate the model's skill from no improvement.

Calibration is especially thin outside the lowest predicted-probability bin. Of 1,656 decisions, 1,633 receive probabilities no greater than 0.1; their mean prediction is 0.02458 and observed frequency is 0.02205. Only 23 decisions fall in the 0.1–0.2 bin, where the mean prediction is 0.11884 but the observed frequency is 0.04348. With one observed event in that bin, the bin-level discrepancy should not be overinterpreted by itself. Together with the aggregate scores and interval, however, it provides no support for using the model as a reliable tail-probability filter.

The probability result also clarifies why lower MAE is insufficient. Point-error metrics average over the whole distribution, whereas trade selection depends on a threshold crossing and subsequent directional persistence. A rule can improve typical signed-rate estimates while assigning

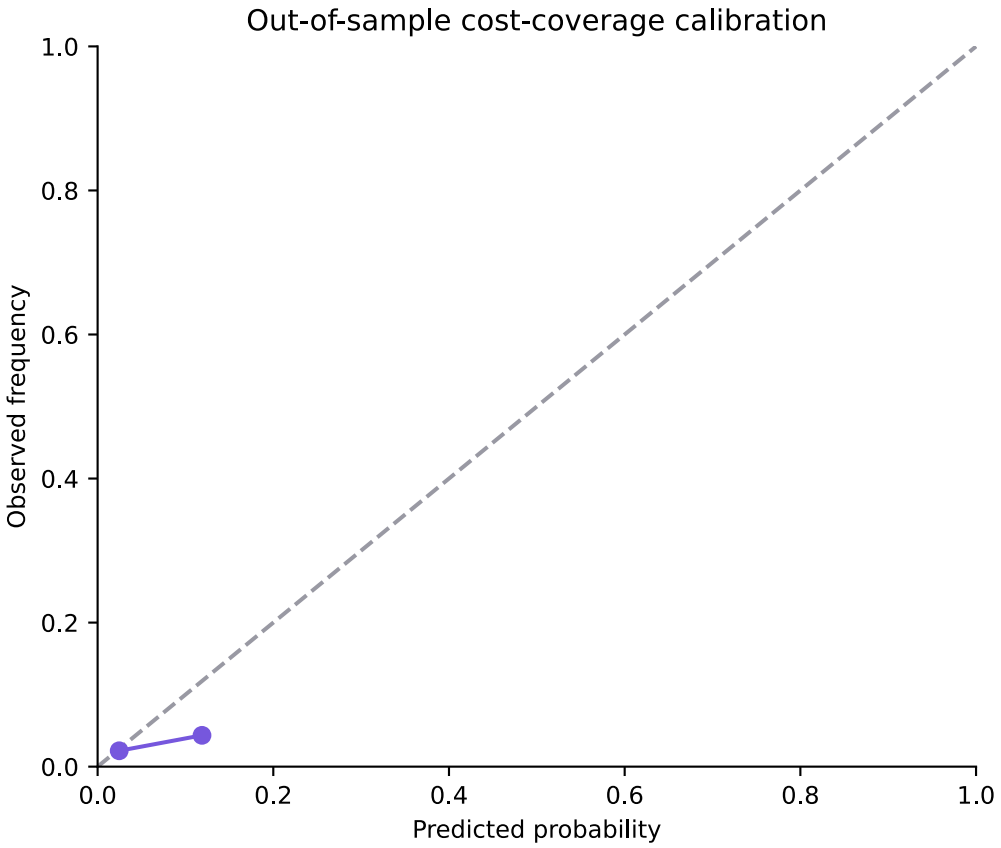


Figure 4: Out-of-sample probability calibration in the explicitly low-cost 24-hour, 5 bp statistical scenario. The exercise diagnoses event prediction; it is not an executable transaction-cost assumption.

excessive probability to some non-events or failing to rank the few genuine events. In this sample, neither the point forecasts nor the probability model establishes a bridge from statistical predictability to cost coverage.

6.6 Current executable-cost snapshot

The current-book exercise provides institutional scale rather than historical execution evidence. Across the 24 routes involving only Gate, Hyperliquid, and Pacifica, median four-fill round-trip cost is 28.72 bp for a $10,000 quote budget per fill. The minimum is 17.61 bp and the maximum is 47.60 bp. These values include taker fees and depth-walked slippage at the snapshot timestamps. The 34 bp primary hurdle lies inside this observed range, making it a useful central scenario without claiming that it was the historical cost of any replay decision.

Routes involving Paradex are reported separately because their current books are much thinner. Their median round-trip scenario is approximately 2,204 bp, which dominates the pooled 48-route median of 85.77 bp and makes a pooled "average venue cost" misleading. The maximum cross-book timestamp skew is 14.392 seconds. That skew is small enough for a descriptive snapshot but still means the four legs are not simultaneous executable quotes. Neither the core nor the Paradex estimates include price/basis P&L, funding during execution, transfer delays, margin financing, or liquidation losses.

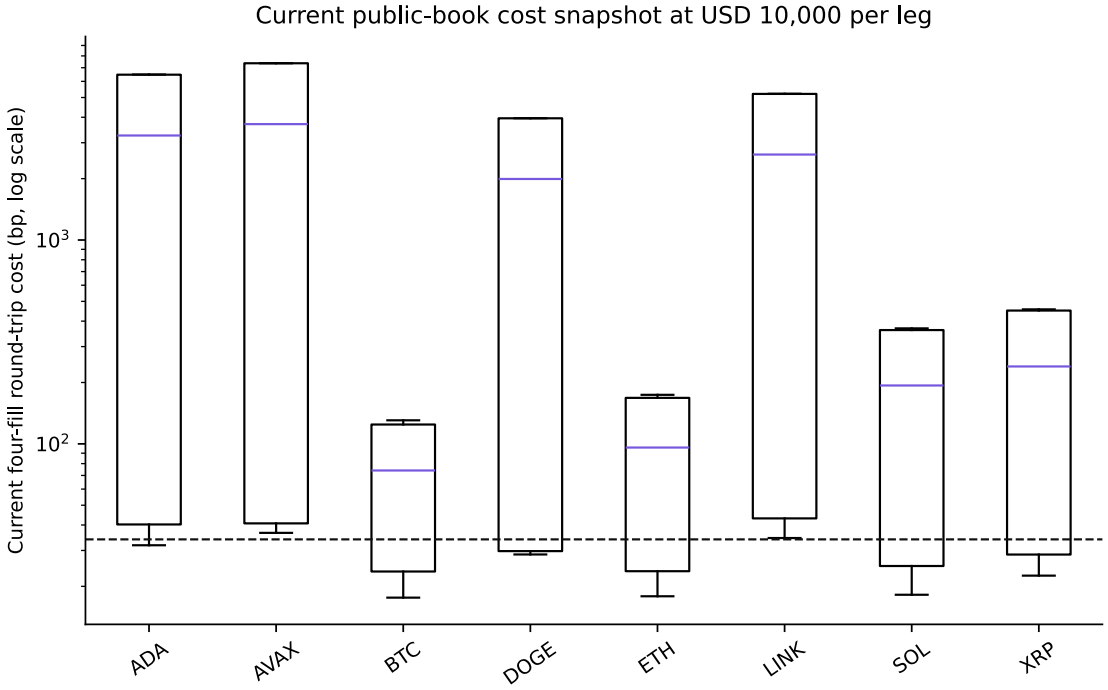


Figure 5: Current depth-walked four-fill cost scenarios by asset on a logarithmic scale. Boxplots pool venue routes within each asset; the dashed reference is 34 bp. Snapshot quotes are not historical fills.

Combining the historical and current evidence therefore supports a narrow conclusion. The observed cost scale is one to two orders of magnitude above the typical eight-hour spread, while multi-day persistence is too weak and too poorly directed to close that gap under the tested rules. The appropriate output is abstention in almost all test decisions, not a repackaging of

small forecast gains as expected arbitrage profit.

7 Robustness

The robustness analysis asks whether the main conclusion depends on one venue pair, one part of the test period, a single estimator, or a convenient cost–horizon combination. It does not expand the candidate search until a profitable specification appears. Empty cells remain undefined when no trades occur, and slices with only one or two entries are treated as diagnostics rather than independent statistical discoveries.

7.1 Chronological subsamples

For the 72-hour, 20 bp case using the three-day median, the early test half contains 264 decisions and two entries; the late half contains 288 decisions and one entry. Both entries in the early half lose, with mean net proxy of -14.53 bp, and the late entry also loses, with net proxy of -14.95 bp. The ex post reference identifies one cost-covering window in each half. Hence the contrast between sparse feasible-looking forecasts and sparse realized windows is not confined to a single market episode.

This split does not create two fully independent experiments. The subsamples are adjacent, assets and venues are shared, and the number of active entries is tiny. Their role is to detect an obvious dependence of the descriptive result on one endpoint or short interval. No such reversal appears: neither half supports positive net funding cash under the stated rule.

7.2 Venue-pair and asset slices

Venue slices show the same pattern. Gate–Hyperliquid produces one median-rule entry in the relaxed 72-hour, 20 bp case, with a net proxy of -11.52 bp. Gate–Pacifica produces no entries. Hyperliquid–Pacifica produces two entries, both unsuccessful, with mean net proxy of -16.25 bp. The ex post reference finds one favourable window for Gate–Hyperliquid and one for Hyperliquid–Pacifica, and none for Gate–Pacifica. Thus no venue pair supplies a hidden cluster of forecast-selected successful trades.

Asset slices are even sparser. ADA has one median-rule entry and LINK has two; all three lose. AVAX, BTC, DOGE, ETH, SOL, and XRP have no entries under this rule. DOGE has one ex post cost-covering window despite no forecast entry, while LINK has two forecast entries but no ex post cost-covering window. These disagreements are consistent with weak directional timing, but the counts are too small for asset-level ranking or inference. In particular, the results do not justify a claim that one asset is inherently more arbitrageable than another.

7.3 Cost and holding-period sensitivity

Figure 6 shows how the median estimator's entry rate and mean net proxy change with the cost hurdle at 72 hours. The complete grid produces no entries in every 8-hour and 24-hour case. Activity appears only at 72 hours: persistence and the mean enter at 20 bp, the mean also enters at 34 bp, and the median enters only at 20 bp. The capped estimator never enters. No active specification has a positive win rate or positive mean net proxy. Increasing the hurdle eliminates entries rather than improving selection; reducing it to 20 bp admits a handful of unsuccessful

Table 4: Robustness slices for the three-day median estimator at 72 hours and 20 bp. Net values are per entered trade. Undefined statistics for zero-entry slices are omitted rather than set to zero.

Slice	Decisions	Entries	Win rate	Mean gross bp	Mean net bp
Gate–Hyperliquid	184	1	0.000	8.483	−11.517
Gate–Pacifica	184	0	–	–	–
Hyperliquid–Pacifica	184	2	0.000	3.752	−16.248
Early test half	264	2	0.000	5.466	−14.534
Late test half	288	1	0.000	5.055	−14.945
ADA	69	1	0.000	5.055	−14.945
LINK	69	2	0.000	5.466	−14.534
Other six assets	414	0	–	–	–

trades.

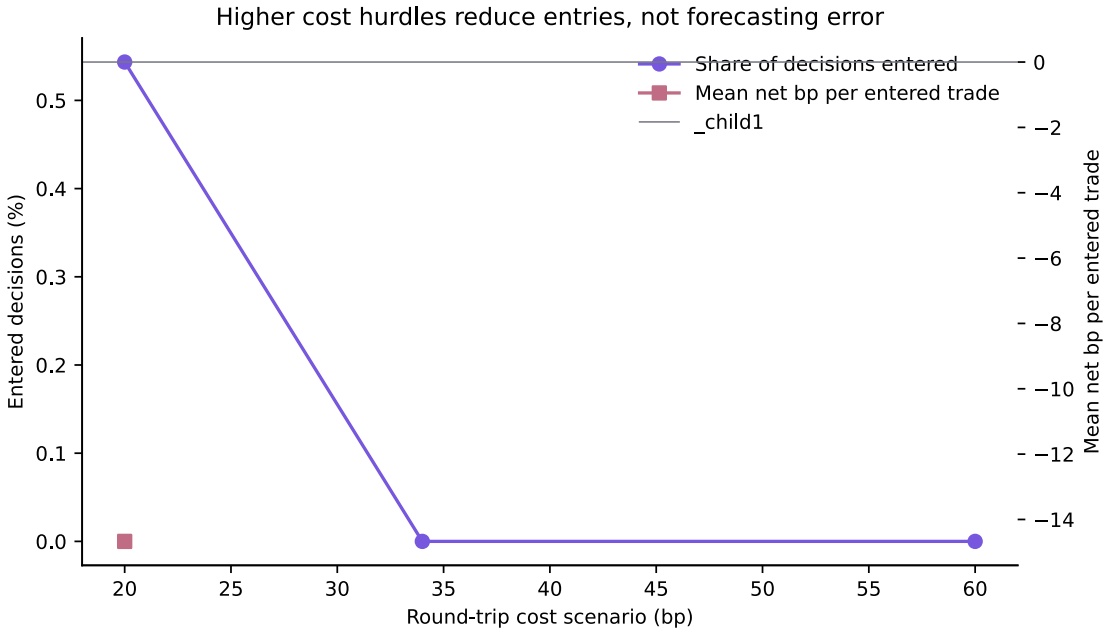


Figure 6: Cost sensitivity for the three-day median estimator at a 72-hour horizon. The left axis reports entered decisions and the right axis reports mean net basis points per entered trade. Zero entries at 60 bp make the per-trade statistic undefined.

This pattern is economically coherent with the spread distribution. A one-block outcome cannot exceed any cost hurdle in the sample. Longer holds allow cumulative differentials to become larger, but they also expose the position to sign reversals and require the initial direction to remain correct for more settlements. Cost sensitivity therefore does not trace a smooth profitability frontier. It mostly moves the system between abstention and a few adversely selected entries.

7.4 Alternative estimators and selection control

The estimator comparison does not produce a conflict between statistical and economic criteria that can be resolved by choosing whichever result looks best. The capped rule has the lowest MAE and RMSE and the strongest Diebold–Mariano evidence, yet it generates no trades in

any cost–horizon case. Persistence has the highest sign accuracy and the most entries in the relaxed 72-hour, 20 bp case, but all seven entries lose. The mean and median lie between these extremes and likewise fail economically. Each estimator reveals a different aspect of the same gap between centre-of-distribution accuracy and tail-event selection.

The small, fixed candidate family and chronological holdout reduce, but do not eliminate, researcher discretion. The White Reality Check specifically addresses selection among the four active estimators in the only case with several entries. Its p-value of 1.0 reflects that the best mean across all decisions is the zero produced by abstention; no active candidate dominates the zero-cash benchmark. This is preferable to reporting the least negative active rule as though it were a successful discovery.

7.5 Coverage and accounting checks

The accounting tests and bias checks support the mechanical interpretation of the replay. Equal-quantity round trips charge exit fees on realized exit notional, asks are consumed from low to high price and bids from high to low, insufficient depth fails closed, and duplicate or missing settlement events are not silently converted into cash. Settlement marks and native event intervals are kept explicit in the reusable accounting component. Eight unit tests cover these behaviours. The panel construction separately requires eight distinct native hourly observations for Hyperliquid and Pacifica and one native settled record for Gate in every admitted eight-hour block.

These checks cannot repair information that the public data do not contain. They instead prevent a precise but false implementation from manufacturing feasibility through look-ahead, duplicated funding, favourable book ordering, or zero-filled gaps. Across time, venue, asset, parameter, and estimator slices, the qualitative conclusion remains unchanged: the tested public-data signals do not support a positive net funding-cash rule under the stated execution scenarios.

8 Limitations and evidence boundary

8.1 What the historical replay identifies

The historical evidence identifies settled funding-rate differentials under a strict common-grid construction and the funding-cash consequences of pre-specified forecast rules under scenario costs. It does not identify a historically executable trade. No archive of synchronized four-way historical order books is available in the replication package. Consequently, the replay cannot establish that both entry legs and both exit legs would have filled at the required quantities, at the assumed times, and at the scenario costs. Applying current depth to past timestamps would create false precision and is deliberately avoided.

The fixed-notional funding-cash proxy also omits the remaining components of complete trading P&L. Equal-base long and short legs can still generate price and basis P&L when entry and exit marks differ, when contracts use different multipliers or settlement currencies, or when oracle and settlement-mark conventions diverge. Margin interest, collateral yield, cross-venue transfers, withdrawal costs, borrowing or margin constraints, auto-deleveraging, and liquidation are not reconstructed. These omissions are not assigned a zero economic value. They define the boundary of the result: a negative funding-cash proxy is sufficient to reject a candidate under the stated proxy, whereas a positive proxy would still require full execution and risk accounting

before being called profitable.

Hourly funding on Hyperliquid and Pacifica is summed into eight-hour fixed-notional equivalents. This normalization makes native records comparable, but it is not a claim that every venue uses identical intrahour oracle prices, position quantities, or cash-crediting mechanics. Exact cash reconstruction would require historical position and settlement-mark data at every event. The paper's paired-rate object is therefore appropriate for comparing settled rates, not a substitute for venue-account ledgers.

8.2 Sample and external-validity limits

The eight assets are currently available, pre-specified major common contracts. Delisted assets, failed venues, contracts that disappeared during the window, and assets without complete common coverage are absent. This creates survivorship and common-availability selection. The sample may understate opportunities in episodic, thin, or newly listed markets, but studying those markets would also make historical execution and manipulation concerns more severe. The findings should not be extrapolated to the entire perpetual-futures universe.

Gate's public history limit restricts the common sample to approximately six months. The period contains variation across assets and time but cannot represent every volatility, funding, or liquidity regime. Moreover, Gate, Hyperliquid, and Pacifica have meaningfully similar funding anchors in the observed data. Evidence of compression among these mechanisms does not rule out larger dislocations between materially different contract designs, during exchange stress, or around market-wide shocks. Such episodes require new data and should be studied prospectively rather than inferred from this sample.

The current order-book snapshot has a different limitation. It measures public quotes at one acquisition time, with at most 14.392 seconds of cross-book timestamp skew. It does not guarantee simultaneous execution, include queue position or hidden liquidity, or represent other market states. Paradex's thin current books make its separate cost estimate especially regime-dependent. The snapshot establishes the scale and heterogeneity of a current taker-cost scenario; it neither dates historical costs nor forecasts future costs.

8.3 Statistical limits

The chronological holdout prevents direct test-set fitting, and the candidate family is deliberately small. Nevertheless, the choice of eight assets, three historical venues, nine-block smoothing, three horizons, and three cost scenarios remains a research design rather than a theorem. Alternative lags, nonlinear models, venue-specific estimators, or regime classifiers could be proposed. They would require a fresh holdout or prospective evaluation and an explicit multiplicity correction; they cannot be validated by repeatedly searching this test segment.

Diebold–Mariano tests average losses across pair–asset panels at each timestamp before applying Newey–West inference. This reduces cross-sectional duplication but does not make every economic observation independent. The decision replay similarly avoids treating shared routes as independent portfolio bets. Small entry counts preclude reliable trade-level inference, while zero-entry cases cannot support estimates of win rates, average trade returns, or calibration at the primary 34 bp threshold. The 5 bp probability exercise addresses a statistical event only and should never be quoted as evidence that a five-basis-point round trip is executable.

Finally, the ex post two-sided reference uses future information by construction. It is included only to count realized windows that could cover a scenario cost under perfect direction. It is neither a tradable oracle strategy nor a valid performance benchmark. Removing look-ahead from the forecasts while retaining it in this clearly labelled reference keeps the evidence interpretable: the sample can contain rare favourable outcomes without supplying a feasible method for selecting them.

8.4 Scope of the negative result

The supported conclusion is conditional: for these eight contracts, these three historical venues, this May–November 2026 sample, and these public-data estimators, ordinary taker-cost scenarios are not covered reliably by settled funding differentials. The paper does not establish universal market efficiency, exclude maker or rebate-based implementations, value-hedged portfolios with additional return sources, or discontinuous stress episodes. It shows what must be demonstrated before a spread is worth trading: a time-stamped forecast, persistent directional cash after all four executions, proper out-of-sample evaluation, and complete risk accounting. In the present evidence, those conditions are not jointly met.

9 Conclusion

A cross-venue funding-rate spread is worth trading only when its future, correctly directed, settled cash can cover the complete cost of entering and exiting both legs. This paper operationalizes that condition with public settlement histories, a common eight-hour grid, a chronological holdout, fixed forecasting rules, explicit cost hurdles, and separate current-book execution scenarios. The design keeps three questions distinct: how large the spread is, whether it is predictable, and whether the prediction survives cost and selection controls.

In the sample of eight major contracts on Gate, Hyperliquid, and Pacifica, the answer is predominantly no. The median absolute spread is 0.45 bp per eight hours and no aligned observation reaches 20 bp. The capped estimator delivers a genuine but small out-of-sample reduction in MAE and RMSE, supported by Diebold–Mariano tests. Yet no estimator enters in the primary 24-hour, 34 bp case, and the ex post reference finds no cost-covering window there either. Sparse entries in the relaxed 72-hour, 20 bp case all lose on the funding-cash proxy. Probability forecasts for a lower-threshold statistical event underperform a constant training-frequency baseline, and the Reality Check finds no active candidate that improves on zero cash.

The main lesson is not that forecasting is useless. It is that average point-forecast accuracy and economic tail selection are different objects. A tightly anchored rate distribution can reward conservative shrinkage while offering almost no observations large enough to pay four-fill taker costs. Extending the holding period increases potential cumulative carry but also increases exposure to sign reversal. The break-even calculation must therefore be joined to a genuinely out-of-sample directional decision, not attached to a contemporaneous scanner after the fact.

The appropriate research and operational response is disciplined abstention when these tests fail. Future work can add synchronized historical books, complete basis and margin accounting, more heterogeneous venues, and prospective stress-period samples. Any such extension should preserve the same evidentiary sequence: normalize native settlement, freeze the information set, evaluate point and probability forecasts, reconstruct all four executions, control data snooping,

and only then discuss net returns. Until those conditions are satisfied, a visible funding-rate difference remains a quotation to investigate, not an arbitrage profit to recognize.

10 Declarations

Corresponding author and research channel

Hang Liu, humarb Research, London, United Kingdom. Research channel: https://humarb.ai/research/; institutional website: https://humarb.ai. This research was completed by humarb Research, and humarb.ai is its public research channel.

Data and code availability

The replication package accompanying this paper contains the public-data acquisition script, deterministic analysis script, accounting component and unit tests, data dictionary, acquisition instructions, raw public funding histories, aligned panel, held-out decisions, generated tables, and vector figures. Public exchange data remain subject to the respective venues' terms of use. The acquisition manifest records timestamps, endpoints, contract mappings, and file hashes. No authentication, private endpoint, order submission, or user data is used.

Funding

No external funding was received for this research.

Conflicts of interest

The author reports no competing financial or non-financial interests directly related to this study. humarb Research conducts research on digital-asset market mechanisms. This affiliation does not alter the evidence boundaries or the negative empirical findings reported here.

Author contributions

Hang Liu designed the study, implemented the public-data acquisition and analysis, verified the references and accounting tests, interpreted the results, and wrote the paper.

Confidentiality review

The manuscript and replication materials were reviewed against the project's confidentiality requirements. They contain public market mechanisms, reproducible public-data methods, and fully disclosed example research parameters. They do not disclose proprietary source trees, internal field names, server or deployment details, production thresholds or ranking rules, undisclosed interfaces, credentials, or user data. Redacted internal values were not reconstructed. Historical results that could not be independently recomputed from public data were excluded.

Generative-AI disclosure

Generative AI was used to assist language editing, code debugging, and literature discovery. It was not used as a source of empirical observations. The author checked the public data, statistical outputs, reference metadata, and final text and takes responsibility for all content.

Ethics and investment-research statement

The study uses only public market data and involves no human participants or personal data. It is an empirical research preprint, not investment advice, an offer, a solicitation, or a recommendation to trade. Digital-asset perpetual futures can involve leverage, liquidation, venue, custody, and operational risks not captured by the funding-cash proxy. Readers should not infer suitability or expected profitability from any example parameter or cost scenario.

References

Damien Ackerer, Julien Hugonnier, and Urban Jermann. Perpetual futures pricing. *Mathematical Finance*, 36(3):481–499, 2026. doi: 10.1111/mafi.70018.

Carol Alexander, Xi Chen, Jun Deng, and Tianyi Wang. Arbitrage opportunities and efficiency tests in crypto derivatives. *Journal of Financial Markets*, 71:100930, 2024. doi: 10.1016/j.finmar.2024.100930.

José Almeida and Tiago Cruz Gonçalves. Cryptocurrency market microstructure: A systematic literature review. *Annals of Operations Research*, 332(1–3):1035–1068, 2024. doi: 10.1007/s10479-023-05627-5.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *The Journal of Risk*, 3(2):5–39, 2001. doi: 10.21314/jor.2001.041.

Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014. doi: 10.1093/jjfinec/nbt003.

Riccardo De Blasis and Alexander Webb. Arbitrage, contract design, and market structure in bitcoin futures markets. *Journal of Futures Markets*, 42(3):492–524, 2022. doi: 10.1002/fut.22305.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995. doi: 10.1080/07350015.1995.10524599.

Gate. Gate api v4: Perpetual futures market data, 2026a. URL https://www.gate.com/docs/developers/apiv4/. Accessed 3 November 2026.

Gate. Futures funding rate and funding fee overview, 2026b. URL https://www.gate.com/help/futures/futures-logic/27569/funding-rate-and-funding-fee. Accessed 3 November 2026.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

humarb Research. Measuring executable net carry in cross-venue perpetual funding-rate arbitrage. Technical report, humarb Research, 2026a. URL https://doi.org/10.5281/zenodo.23868361.

humarb Research. Framing net carry for cross-venue perpetual funding positions: A settled-history, depth-walked, and liquidation-aware method. Technical report, humarb Research, 2026b. URL https://doi.org/10.5281/zenodo.23897013.

Hyperliquid. Fees, 2026a. URL https://hyperliquid.gitbook.io/hyperliquid-docs/trading/fees. Accessed 3 November 2026.

Hyperliquid. Funding, 2026b. URL https://hyperliquid.gitbook.io/hyperliquid-docs/trading/funding. Accessed 3 November 2026.

Emre Inan. Predictability of funding rates. *International Review of Financial Analysis*, 119: 105464, 2026. doi: 10.1016/j.irfa.2026.105464.

Ralph S. J. Koijen, Tobias J. Moskowitz, Lasse Heje Pedersen, and Evert B. Vrugt. Carry. *Journal of Financial Economics*, 127(2):197–225, 2018. doi: 10.1016/j.jfineco.2017.11.002.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

Pacifica. Trading fees and get historical funding, 2026a. URL https://docs.pacifica.fi/api-documentation/api/rest-api/markets/get-historical-funding. Accessed 3 November 2026.

Pacifica. Funding rates, 2026b. URL https://docs.pacifica.fi/trading-on-pacifica/funding-rates. Accessed 3 November 2026.

Paradex. Trading fees, 2026a. URL https://docs.paradex.trade/trading/trading-fees. Accessed 3 November 2026.

Paradex. Funding mechanism, 2026b. URL https://docs.paradex.trade/risk/funding-mechanism. Accessed 3 November 2026.

Edson Pindza. Centralized–decentralized exchange funding rate arbitrage as a basis trade: Risk decomposition, stress testing, and portfolio construction under venue uncertainty. *Digital Finance*, 8(3), 2026. doi: 10.1007/s42521-026-00213-3.

Maik Schmeling, Andreas Schrimpf, and Karamfil Todorov. Crypto carry. *Management Science*, 2026. doi: 10.1287/mnsc.2024.05069.

Robert J. Shiller. Measuring asset values for cash settlement in derivative markets: Hedonic repeated measures indices and perpetual futures. *The Journal of Finance*, 48(3):911–931, 1993. doi: 10.1111/j.1540-6261.1993.tb04024.x.

Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000. doi: 10.1111/1468-0262.00152.

A Replication and supplementary results

A.1 Deterministic reconstruction sequence

The replication workflow is intentionally linear. First, the acquisition script downloads native public funding records for the eight pre-specified assets and saves contract metadata, endpoint responses, acquisition timestamps, and hashes. Second, native timestamps are deduplicated. Gate contributes one settled eight-hour record per admitted block; Hyperliquid and Pacifica require eight distinct hourly observations. Third, rates are expressed as eight-hour fixed-notional equivalents and paired by asset and UTC block. No interpolation, forward fill, or zero fill is applied to incomplete blocks.

Fourth, each pair–asset series receives a nine-block warm-up. Predictors at block t use only observations through $t-1$. The chronological split retains the first 312 blocks for training and the final 209 blocks for testing. Fifth, the analysis computes point losses, non-overlapping decision outcomes, probability scores, and robustness slices. Finally, the same script writes all tables, `results.json`, and figures used in the manuscript. The fixed random seed is 20261103 and is relevant only to bootstrap procedures. The supplied current-book file is analyzed as a snapshot and is never joined to historical decisions as though it were contemporaneous.

The scripts are research examples, not production software. All windows, hurdles, and estimators are disclosed to make the statistical experiment reproducible; none is represented as a live-system configuration. Acquisition requires network access to public APIs, while analysis can run entirely from the archived files. Running the unit tests before analysis checks common-quantity execution, actual exit-notional fees, depth failures, duplicate and missing events, and settlement-mark handling.

A.2 Supplementary forecast comparisons

Table 5 reports all Diebold–Mariano comparisons against persistence. Loss differentials are averaged over the 24 pair–asset panels at each test timestamp. Newey–West standard errors use 18 lags. The table makes clear that the capped estimator is the only candidate with significant improvements under both absolute and squared loss.

Table 5: Diebold–Mariano tests against last-value persistence. Negative statistics favour the alternative estimator. Each test uses 209 cross-sectionally averaged test timestamps.

Alternative	Loss	Mean differential	t	p-value
Three-day mean	Absolute	2.933×10^{-6}	2.521	0.0117
Three-day mean	Squared	1.481×10^{-10}	0.627	0.5308
Three-day median	Absolute	1.977×10^{-6}	1.993	0.0463
Three-day median	Squared	4.161×10^{-10}	1.564	0.1177
Capped	Absolute	-3.079×10^{-6}	−4.510	0.0000065
Capped	Squared	-4.131×10^{-10}	−2.561	0.0104

The signs of the mean and median comparisons reflect their slightly worse aggregate MAE and RMSE despite occasional timestamp-level differences. This is another reason to report both aggregate losses and formal comparisons rather than relying on a single test statistic. Inference concerns forecast accuracy only. No DM statistic enters the cash proxy, and no significance level is interpreted as an economic return.

A.3 Complete decision grid

Table 6 condenses all 36 estimator–horizon–cost cases. "Entries" is the number of active forecast decisions; "mean net" is calculated only across those entries. The two-sided ex post reference is common across estimators within each horizon–cost case and is therefore not repeated as a purported strategy result.

Table 6: Complete out-of-sample decision grid. All cases have 5,016 decisions at eight hours, 1,656 at 24 hours, and 552 at 72 hours. A dash means that no active trade exists and the per-trade statistic is undefined.

Horizon	Cost bp	Estimator	Entries	Win rate	Mean net bp
8h	20	All four	0	–	–
8h	34	All four	0	–	–
8h	60	All four	0	–	–
24h	20	All four	0	–	–
24h	34	All four	0	–	–
24h	60	All four	0	–	–
72h	20	Last value	7	0.000	−8.246
72h	20	Three-day mean	6	0.000	−14.674
72h	20	Three-day median	3	0.000	−14.671
72h	20	Capped	0	–	–
72h	34	Last value	0	–	–
72h	34	Three-day mean	2	0.000	−28.534
72h	34	Three-day median	0	–	–
72h	34	Capped	0	–	–
72h	60	All four	0	–	–

At eight hours, the ex post reference has zero entries at every hurdle. At 24 hours, it likewise has zero entries at 20, 34, and 60 bp. At 72 hours, it has two entries at 20 bp and none at 34 or 60 bp. Its 20 bp proxy sum of \$4.277 is computed with perfect future direction and is not attainable performance. These counts reinforce that zero forecast entries in many cases are not a coding failure or a consequence of withholding favourable outcomes.

A.4 Probability-score definitions

For probabilities p_i and realized binary events y_i, the Brier score and log loss are

$$BS = n^{-1} \sum_i (p_i - y_i)^2, \qquad LL = -n^{-1} \sum_i \left[y_i \log p_i + (1 - y_i) \log(1 - p_i) \right]. \tag{13}$$

The baseline sets every p_i to the training event frequency, 0.028846; the test event frequency is 0.022343. The logistic coefficient on the log forecast-to-threshold ratio is 1.0814 and the intercept is −2.1422. These fitted values are included for exact replication, not as a recommendation for deployment. The out-of-sample scores, rather than coefficient significance, determine the paper's negative assessment.

The Brier-skill interval is obtained by resampling moving blocks of 21 eight-hour timestamps 2,000 times. Pair–asset observations within a timestamp remain together. The interval $[-0.02688, 0.00044]$ includes zero. The negative point estimate therefore does not establish worse population skill, although it provides no evidence of improvement over the baseline. Because the primary 34 bp training sample has zero positive events, no two-class model is estimated for that case. Substituting an arbitrary smoothing constant to manufacture

probabilities would conceal, rather than solve, the identification problem.

A.5 Data dictionary summary

The raw funding-history file records venue, asset, native event timestamp, native funding rate, and native interval metadata. The aligned panel records UTC block end, asset, venue pair, signed paired differential, and the information-set forecasts. The held-out decision file adds horizon, cost scenario, estimator, entry indicator, realized gross proxy, cost deduction, and net funding-cash proxy. The current-book file preserves local quote timestamps and price–quantity levels by venue and contract. The acquisition manifest records endpoint families, coverage, errors, and SHA-256 hashes. Full machine-readable descriptions accompany the replication package.

Rates in analysis tables are fractions unless a column name explicitly ends in `_bp`; one basis point equals 10^{-4}. Dollar proxy values use \$10,000 fixed notional per leg and do not include a capital denominator. Empty numeric cells indicate undefined values, most often because a case has no entries. They are not zeros. Current-book slippage and fees are snapshot estimates based on actual displayed depth and fee schedules; they are not historical realized transaction costs.

A.6 Checklist for interpreting a new spread

The workflow implies six research checks. First, distinguish indicated from settled rates and normalize accrual intervals. Second, reconstruct a common base quantity and four executions. Third, forecast only from pre-entry information and evaluate on a later period. Fourth, compare signed cumulative cash with the full round-trip hurdle, not with zero gross funding. Fifth, use proper probability scores and dependence-aware inference. Sixth, state what remains unidentified, especially historical depth, price/basis P&L, liquidation, and margin effects. These checks make the negative result informative without yielding a trading rule.