Ranking Cross-Venue Funding Carries under an Estimation-Error Budget

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

humarb Research 11 · Preprint v1 · 3 November 2026
DOI: 10.5281/zenodo.23938430

Abstract

Cross-venue funding-carry rankings compare noisy current scenarios, and selecting the maximum can change the relevant error distribution. We derive an accounting error budget: for current scenario value S_0, realized net accounting Π and overestimation $D = S_0 - \Pi$, $\Pi > 0$ if and only if $D < S_0$. We separate this baseline from a preference- and distribution-dependent decision objective that remains unavailable when joint errors or capital inputs are missing. Using 9,192 public hourly funding records, a previous-settled benchmark has holdout MAE of 0.0126–0.0298 basis points across six series. Calibration-period 90% intervals attain 68.39% coverage for Hyperliquid ETH; a one-sided rule covers 94.03% of all holdout spread counterfactuals but 88.39% of selected maxima. In a five-route public-book paper illustration, current and subsequent scenario values are all negative and reconcile with negligible numerical residuals. The results establish an exact ranking and audit framework, not positive-expectation evidence or a calibrated decision rule.

Keywords: perpetual futures; funding rates; estimation error; ranking; post-selection bias; forecast evaluation

JEL classification: G12; G13; G14; C58

DOI: 10.5281/zenodo.23938430

1 Introduction

A cross-venue funding-carry ranking is a choice among noisy estimates, not merely a sort of displayed rates. The current published funding rates can differ across venues, the quantities that can actually be traded can differ across books, and the future rates, marks, exit basis and fees need not equal their current counterparts. A ranking that ignores the first distinction is incomplete; a ranking that ignores the second confuses a scenario with a distribution; and an evaluation that reports only the winning row can mistake selection for evidence. The research question is therefore narrow: *how should current cross-venue funding scenarios be ranked, how much joint estimation error can a selected scenario tolerate, and what additional evidence is required before that ranking becomes a defensible decision rule?*

This paper develops an error-budget answer. Let $S_{0,z}$ be the net value of route–quantity–horizon scenario z under the books, rates, marks, fee basis and event schedule known at decision time. Let Π_z be the closed-position accounting under the same contract. Their difference $D_z = S_{0,z} - \Pi_z$ is net overestimation error. By construction,

$$\Pi_z > 0 \quad \Longleftrightarrow \quad D_z < S_{0,z}. \tag{1}$$

The current scenario value is thus the total adverse error that the position can absorb before its sign changes. This statement is an accounting identity, not a forecast, a confidence interval or a claim that positive S_0 is attainable in a particular market. Its practical value is that funding realization, basis, crossing, fees and other costs must reconcile to one observable boundary rather than disappear into an undifferentiated slippage residual.

The identity also exposes what a ranking does not know. A decision maker who dislikes losses and reserves capital needs a preference-weighted expectation over the joint error distribution. We write the research objective as $J_z = \mathbb{E}[\Pi_z \mid \mathcal{F}_t] - \lambda\mathbb{E}[(-\Pi_z)^+ \mid \mathcal{F}_t] - \rho A_z$. The first two terms require the joint distribution of both legs' rates, prices and costs; the last requires a declared capital-reservation plan and its price. If any of these inputs is unavailable, J_z is unavailable. It is not zero, and its absence does not turn a positive current scenario into a negative one. Conversely, replacing missing covariance or preferences with convenient zeros would manufacture an answer that the evidence does not contain.

1.1 Contributions and principal findings

The paper makes four contributions. First, it derives the error budget from the event-time, common-quantity ledger of Liu (2026a) and gives component-wise and sensitivity forms for funding, basis, execution and fee discrepancies. The accounting contract is imported rather than claimed as new. Second, it separates the descriptive S_0 ranking from the preference- and distribution-dependent J objective, proves the limited implication $J_z > 0 \Rightarrow \mathbb{E}[\Pi_z \mid \mathcal{F}_t] > 0$ under true conditional inputs, and supplies a deterministic horizon reversal showing why the final objective must be evaluated at every candidate endpoint. Third, it specifies an evaluation design with chronological isolation, label purging, shared-event resampling, post-selection metrics and explicit NA denominators. Fourth, it implements a public, reproducible component analysis and a one-hour public-book paper illustration without presenting either as proof of economic effectiveness.

The historical component sample contains 9,192 consecutive hourly funding records for BTC, ETH and SOL at Hyperliquid and Extended from 2026-09-01 through 2026-11-03. The chronological holdout contains 3,720 one-step predictions after 2026-10-09. A previous-settled-rate benchmark has holdout mean absolute errors of 0.0126–0.0298 basis points across the six venue–asset series, equivalent to 0.0126–0.0298 USD-denominated scenario units at a fixed 10,000-unit notional. This is a funding-rate error scale, not a return. Previous-settled prediction is more accurate than an expanding mean in every series; comparisons with a trailing 24-hour median are mixed after multiplicity adjustment. These results support a transparent benchmark, not an assertion of persistence in future regimes.

Coverage provides the more consequential warning. Calibration-period central 90% intervals attain only 68.39% holdout coverage for Hyperliquid ETH, and several independence or conditional-coverage tests reject even where unconditional coverage is close to nominal. In the two-venue spread exercise, a calibration one-sided rule covers 94.03% of all holdout counterfactuals but 88.39% of the selected maxima. Selecting the largest scenario spread therefore changes both the error distribution and the meaning of a coverage statement. This is exactly the post-selection problem that a full-candidate average conceals.

The forward illustration freezes public books and current-rate scenarios at 22:41 UTC on 2026-11-03 for the 23:00 UTC settlement. Five routes have complete current-book measurements and one is retained as depth-incomplete. Current costs reverse the ordering suggested by the raw funding gap. All monetary aggregates use an explicit one-for-one USDC/USD parity scenario; they are not realized USD, USDT or USDC account cash. The subsequent exit books and public settlement records close the accounting error budget, but the exercise remains a correlated public paper proxy with no orders, account funding cash or account fee tier. It is not an externally preregistered trial and not a validation of J.

1.2 Evidence boundaries and relation to earlier work

Three evidence levels remain separate throughout. Algebraic results are exact once their assumptions hold. Public-data results are reproducible measurements of the stated samples and proxies. Economic usefulness of the complete recommendation objective remains unresolved because no independently preregistered, locked forward sample of complete closed positions has been assembled. The paper does not estimate a probability of profit, does not supply values for λ or ρ, and does not recommend a venue, direction, amount or holding period.

The series division of labor is deliberate. Liu (2026c) establishes the executable-net-carry measurement problem, while Liu (2026b) frames settled history, depth walking and liquidation awareness. Liu (2026a) supplies common-quantity closure, four-fill accounting and event-time funding enumeration. This paper begins where that ledger ends: it studies the tolerable error in a current scenario and the statistical conditions for using the scenario in a selection rule. Premium-integral prediction and backtest-overfitting diagnostics are not re-estimated as a competing carry model; forecast evaluation and selection bias enter only as constraints on the proposed decision layer.

Section 2 positions the argument in the perpetual, arbitrage, execution and forecast-evaluation literatures. Section 3 states the institutional facts needed to interpret public fields. Section 4 defines the samples and provenance. Section 5 derives the scenario, error budget, decision objective and evaluation estimands. Section 6 reports the empirical component results and forward accounting illustration. Section 7 examines alternative estimators, subperiods, windows, loss functions and dependence choices. Sections 8 and 9 state limitations and conclusions. Appendices contain proofs, the prospective protocol, the complete trial ledger and the data dictionary.

2 Related literature and constraints on interpretation

The argument draws on four literatures but does not treat any of them as a source of ready-made trading thresholds. The relevant distinction is between a result about cash accounting, a result about a probability model, and a result about selecting the best among many estimated alternatives. Combining them requires explicit assumptions at each junction.

2.1 Perpetual cash flows and arbitrage frictions

He et al. (2022) model perpetual-futures cash flows and examine convergence and arbitrage in their sample. Their framework supports separating mark-to-market price com-

ponents from funding and including entry and exit costs. It does not make a particular cross-venue, single-settlement position profitable, and their longer-horizon empirical returns cannot be imported as a probability for the present setting. Ackerer et al. (2026) distinguish linear, inverse and quanto contracts and show why the timing of cash flows matters for pricing and financing. Their risk-neutral valuation is not a real-world distribution of the error in a displayed funding rate. For the present ledger, contract identity, settlement currency and event-time cash amounts therefore precede any comparison of rates.

Cross-market price gaps also need not be immediately executable. Makarov and Schoar (2020) show that cryptocurrency price discrepancies can coexist with available-quantity and capital frictions; their early spot-market evidence is not a 2026 perpetual-funding profit estimate. More generally, Shleifer and Vishny (1997) explain why arbitrage can remain limited when capital and agency constraints bind. These papers motivate complete quantity, liquidity and capital measurements rather than a count of theoretical venue pairs. They do not imply that a large candidate set contains a genuinely positive opportunity or that mutually exclusive routes can be aggregated as independent trades.

The present question differs from pricing a perpetual contract or measuring a historical spread. It asks how much error a current, fully costed scenario can tolerate and what must be learned before a noisy maximum is used for selection. The no-arbitrage and limits-to-arbitrage literatures constrain that question but do not answer its estimation problem.

2.2 Execution measurement and cost completeness

Implementation shortfall separates a reference decision from subsequent execution (Perold, 1988). Almgren and Chriss (2001) further distinguish reference prices, spread, market impact and inventory risk. We use these distinctions to require four quantity-matched fills and to prevent double counting: if depth-walked execution prices already contain spread crossing and displayed-depth impact, the same amount cannot be deducted again as a generic slippage charge. Public-book walking remains a paper measurement, not a guarantee of queue position, replenishment or actual fills.

This distinction changes the estimand. A mid-price basis movement and a movement in executable VWAP are different components of D. A fee difference caused by changed notional is different from a fee-tier difference. If an exit book or official settlement-price input is missing, the affected component is NA; it cannot be set to zero merely to complete an attractive decomposition. The execution literature supplies these accounting cautions, not impact coefficients that can be transferred across venues, assets or time.

2.3 Selection, utility and uncertainty

The optimizer's curse is directly relevant to a maximum over route estimates. Smith and Winkler (2006) show that selecting the largest of noisy valuations can produce postdecision disappointment even when individual estimates are unbiased. Their remedy requires an explicit probabilistic model; it does not justify a universal lower-confidence multiplier, a fixed cost threshold or a claim that more combinations imply more independent opportunities. Shared assets, books and funding events make the dependence structure central to this paper.

A risk measure likewise presupposes a loss distribution. Rockafellar and Uryasev (2000) provide a variational definition and optimization framework for conditional value-at-risk

(CVaR), not the tail probability or risk preference appropriate to a particular funding-carry decision. We therefore use a single piecewise-linear utility objective J rather than stacking uncalibrated CVaR, variance and confidence deductions. CVaR remains a possible alternative for future research once a joint distribution and preferences are defensible. It is not silently inserted as another penalty.

Conformal prediction offers finite-sample marginal coverage under stated exchangeability and calibration conditions (Angelopoulos and Bates, 2022). Marginal coverage over all candidates is not conditional coverage for the maximum, for an asset subset, or under arbitrary temporal drift. The distinction between exchangeability, time dependence and selected subsets limits what our chronological diagnostics can claim. Similarly, proper scoring rules evaluate probabilistic forecasts only when a probabilistic forecast exists (Gneiting and Raftery, 2007); we do not assign CRPS or a log score to the deterministic S_0 ranking as though it were a density.

2.4 Forecast comparison and data-dependent selection

Point alternatives are compared with MAE and RMSE, and Diebold–Mariano loss-differential tests use a Newey–West variance to respect serial dependence (Diebold and Mariano, 1995; Newey and West, 1987). Interval evaluation requires more than an empirical coverage count: unconditional coverage can coexist with dependent violations. We therefore report Kupiec and Christoffersen diagnostics alongside interval scores (Kupiec, 1995; Christoffersen, 1998). Rejections are evidence against the tested calibration behavior, not proof that an alternative rule will be better.

A trial ledger is necessary because model, window, loss and subgroup choices multiply the number of implicit comparisons. Harvey et al. (2016), White (2000) and Hansen (2005) provide broader multiple-testing and data-snooping frameworks, while Bailey et al. (2017) address selection among backtested strategy configurations. Their presence in the literature review does not license a PBO or SPA calculation on an unsuitable matrix of asynchronous route outcomes. We use Holm adjustment for the clearly defined families of Diebold–Mariano and coverage tests, disclose all attempted specifications, and reserve combinatorial backtest-overfitting methods for a future synchronized complete-return panel.

The resulting position is deliberately narrower than a new forecasting model. Perpetual and execution papers define what must enter the ledger; decision and uncertainty papers define what a risk statement would require; forecast-evaluation papers define how to test a frozen estimator on later data. Our contribution is to connect those constraints through S_0, D and J while leaving unsupported probability and profitability claims outside the model.

3 Institutional background

3.1 Funding cash is an event, not a displayed annualized number

A perpetual funding rate links a payment to a position and a valuation price at an eligible event. In the linear, common-base-quantity case used here, signed funding cash at event τ on leg i is written $-d_i Q_i(\tau^-) M_{i,\tau} r_{i,\tau}$, where $d_i = +1$ for a long and -1 for a short. Positive rates ordinarily mean that longs pay shorts; negative rates reverse the direction. The identity is deliberately expressed per event because the quantity, valuation price, eligibil-

ity rule and schedule can vary. A displayed annual percentage rate is not the cash flow, and a current-rate field need not be a locked future receipt.

Venue mechanisms can include an interest component, a premium index, averaging or integral windows, clamps or dead bands, caps, and separate calculation, eligibility and payment times. A researcher must therefore distinguish the current calculated rate, a predicted next rate, a previously settled rate and an account's realized cash. Only the last is authoritative for an account, and even it does not establish execution evidence. If the official event valuation price is unavailable, a public mark or oracle series can support only a labelled funding proxy. A later history timestamp does not prove that all earlier expected records are present.

3.2 Hyperliquid fields used in the public sample

Hyperliquid's official funding documentation describes hourly payment and an eight-hour rate formula whose hourly implementation uses the corresponding fraction of that rate. Premium samples are taken every five seconds and averaged over the hour, with an interest-rate component and clamp. The same documentation states a 0.5% hourly cap and, critically for cash conversion, defines the interval payment as position size multiplied by oracle price and funding rate rather than by mark price (Hyperliquid, 2026b). The public information endpoint supplies current funding, oracle, mark and other market-context fields, historical funding records, and level-two books (Hyperliquid, 2026c).

Our C_0 calculation treats the current published rate and current oracle price as an unchanged-current-state scenario. It does not assert that the next settlement is locked. The subsequent proxy uses the public historical rate nearest the nominal hourly event and the later public oracle-price snapshot. Because the latter is not an archived official account-funding record, exact account funding remains unavailable. Public base-tier taker fees are 0.045% before account-specific volume, staking, referral or builder adjustments (Hyperliquid, 2026a). The illustration uses that base rate solely as a disclosed public fee scenario.

3.3 Extended fields used in the public sample

Extended's API documentation describes current market statistics with a funding rate calculated every minute and a separate timestamp for the next funding update. Public history records contain a market identifier, a rate field and a timestamp; current statistics separately contain mark and index prices (Extended, 2026a). The historical series used here is hourly in the frozen sample, but the analysis does not turn that observation into a permanent venue-wide scheduling guarantee. Contract status, funding interval and record semantics must be rechecked in any new snapshot.

For Extended, the scenario uses the current calculated funding rate and current mark price; the subsequent proxy uses the public historical rate nearest the target event and the later public mark-price snapshot. This is not an account funding payment. The public fee schedule states a flat 0.025% taker rate for ordinary perpetual and spot markets at the access date (Extended, 2026b). The illustration uses that ordinary-market rate and does not assume maker execution, rebates or an RWA discount.

3.4 Currency, fee and execution contracts

The two public interfaces do not present an identical settlement-currency label throughout. Hyperliquid funding and base-fee accounting are conventionally USDC-denominated, while Extended uses a USD collateral label in the reviewed market metadata. Rather than silently treating the two as identical account cash, every combined monetary value in the illustration is reported in *USD-denominated scenario units under an explicit one-for-one USDC/USD parity assumption*. This is a reporting convenience only. Event-time conversion evidence is absent, so realized reporting-currency account P&L remains unavailable. Rate-only results do not depend on this parity assumption; the fixed-notional conversion is always labelled as a scenario scale.

Each route uses one common base quantity for entry and exit on both legs. Entry and hypothetical same-book exit produce current crossing and four fee notionals; actual subsequent books, when available, produce the later paper fills. Transfers, financing, borrow, liquidation, queueing and opportunity costs are outside the measured boundary. Setting other cost O to zero is therefore an isolated-accounting assumption, not a statement that those costs are economically irrelevant. Margin is reserved capital rather than a trading fee and belongs in A, not as a subtraction from Π.

Finally, source timestamps and local retrieval times answer different questions. The event timestamp determines attribution to a holding window; the public record timestamp documents when the venue recorded the event; local receipt determines when a replay could have used it. The reproduction manifest preserves all available fields and marks an absent source clock rather than replacing it with the retrieval clock. These rules are public research conventions for the reported examples, not descriptions of any production system.

4 Data and evidence inventory

4.1 Frozen hourly funding histories

The retrospective component study uses public, unauthenticated funding-history endpoints at Hyperliquid and Extended. The target frame is 2026-09-01 00:00 UTC through 2026-11-03 20:00 UTC, exclusive of the final incomplete target hour. BTC, ETH and SOL were selected for common availability across both venues. Each venue–asset series contains 1,532 consecutive hourly records with no internal nominal-hour gap, producing 9,192 records in total. Table 1 reports the inventory; there is no imputation, outlier deletion or replacement of a recorded timestamp with its nominal hour.

The chronological boundary is 2026-10-09 00:00 UTC. Calibration contains 912 records per series and the holdout 620. One-step prediction drops the first observation of each series, so calibration contains 5,466 scored predictions and the holdout 3,720. The boundary was inherited from the earlier frozen collection to align with its documentation-audit date, not selected because of a favorable forecast result. Nevertheless, the present estimator comparisons are new retrospective analyses of an already collected sample. They are not externally preregistered tests and are not described as a locked D3 experiment.

Raw responses are preserved in 36 request-level files with their original SHA-256 hashes and a manifest recording public URL, request body, retrieval clock, HTTP status and response path. These responses were collected for an earlier public measurement exercise and are reused unmodified; the prediction, selection and error-budget analyses in this pa-

Table 1: Frozen public funding-history sample. Each venue–asset series has consecutive hourly records; forecast evaluation drops the first observation in each series. Source: public APIs collected through 2026-11-03 UTC.

Venue	Asset	Records	Calibration	Holdout	Internal gaps
Hyperliquid	BTC	1,532	912	620	0
Hyperliquid	ETH	1,532	912	620	0
Hyperliquid	SOL	1,532	912	620	0
Extended	BTC	1,532	912	620	0
Extended	ETH	1,532	912	620	0
Extended	SOL	1,532	912	620	0
Total	3 assets	9,192	5,472	3,720	0

per are new. Offline scripts rebuild every normalized record from those raw responses. A new online collection would be a new snapshot and is not expected to reproduce the frozen sample if a venue extends or revises its history.

The unit for point prediction is one hourly venue–asset event. The previous-settled forecast at event t uses the rate and public timestamp at $t - 1$, so the input cutoff precedes the target event. Expanding means and trailing medians use only earlier records. The sample is deliberately narrow: two venues, three assets and one hourly mechanism. It is not a comprehensive venue universe, and delisted or suspended contracts outside the three continuing series cannot support a survivorship assessment.

4.2 Two-direction funding-spread counterfactuals

The selection analysis aligns the two venues by asset and nominal hour. At each paired hour it constructs two opposite-direction fixed-notional funding-spread scenarios: long Hyperliquid/short Extended and the reverse. There are 4,593 paired event rows after lag construction and 9,186 directional counterfactuals. The holdout contains 3,720 directional rows, of which 1,860 are the selected maxima, one per asset-hour. Both directions are retained so that selection is visible rather than baked into the sample.

The scenario spread uses previous-settled rates; the subsequent proxy uses the rates recorded at the target events. The selected subset maximizes the scenario spread without a positivity or admission filter. This design isolates the mechanical effect of taking a maximum. It is not a portfolio: opposite directions are mutually exclusive, nearby hours are dependent, and fixed notional excludes actual quantity, basis, fees and capital. No win rate, annualized return or independent-trade count is computed from these rows.

4.3 Forward public-book snapshot

A separate forward accounting illustration begins at 22:41 UTC on 2026-11-03 and targets the 23:00 UTC hourly settlement. The initial public inventory contains 111 active asset names shared by the two interfaces with usable current-rate fields. The selection rule, fixed before target outcomes were collected, takes the three largest absolute current-rate gaps, the largest signed Extended-minus-Hyperliquid gap, and BTC and ETH as liquid controls. This yields PURR, TIA, XMR, LAYER, BTC and ETH. The procedure is descriptive and does not claim a market-wide optimum or an independently timestamped preregistration.

Each leg has a 10,000-unit notional ceiling. The long venue is the one with the lower cur-

rent funding rate and the short venue the one with the higher rate; quantity is rounded down to the common public quantity grid and never resized. Entry books and same-book counterfactual exits determine current crossing and public base-taker fee scenarios. XMR remains in the inventory as `insufficient_displayed_depth`; it is not assigned a synthetic VWAP or a zero loss. The other five routes have complete current-book scenarios.

The scenario file, normalized entry books and their hashes were frozen at 22:41:06.805 UTC. The exit rule requests the first paired public books on or after 23:00:20 UTC and matches historical funding records to the nominal event within a 10-second tolerance. A missing rate, book or required price remains NA. The final package includes the initial and subsequent raw responses, request manifest and exact decimal-arithmetic reconciliation. No orders are sent and no account endpoint is queried.

All combined forward monetary amounts use the one-for-one USDC/USD parity reporting scenario described in Section 3. The parity note is a post-collection reporting clarification rather than a change to the frozen numerical inputs; the original scenario and book files remain byte-identical to their frozen hashes. This assumption does not enter the rate-only historical estimators. The 10,000-unit ceiling is an example measurement scale, not production sizing or an assertion about account collateral. Likewise, the public base taker rates are example fee inputs rather than an observed account tier.

4.4 Provenance, missingness and reproducibility

The evidence inventory distinguishes five states: a complete public-book paper proxy; a current scenario with incomplete depth; a missing official public funding record; an incomplete exit; and an unavailable decision objective. NA has no numeric value. A verified absence of an eligible event would differ from a scheduled event whose record is missing, although no such distinction is used to manufacture a zero in the reported forward rows. The complete-candidate file, stage-specific status and raw evidence accompany every result.

The reproduction package contains raw history and forward responses, normalized CSV files, data dictionary, collection and analysis scripts, generated tables and vector figures, and a machine-readable verification report. Core recomputation requires no credentials, network access or orders. Every displayed numerical result maps to a generated CSV or JSON output. Public documentation is represented by official URLs, access dates and hashes rather than runtime links from the webpage or unsupported claims of permanent API semantics.

5 Methods

5.1 Scenario keys and the current net scenario value

A candidate is not a venue pair alone. It is a complete action and information contract $z = (r, Q, T, b, v)$, where r identifies the two instruments and directions, Q is their common base quantity, T is a predeclared exit endpoint, b is the fee and execution basis, and v identifies the input version and decision cutoff. All terms entering one sum or comparison must share this key. Amount, quote and reporting basis, holding view, schedule and fee mode are fixed before ranking. Unsupported instruments, infeasible quantity grids and unmeasured books remain separate states rather than silently reduced quantities.

For each leg i and eligible event τ through T, the current scenario uses the rate $r_{i,0}$, valuation price $M_{i,0}$ and schedule known at t. Signed scenario funding is

$$C_{0,z}(T) = -\sum_{i} \sum_{\tau \in E_{i,z}(t,T)} d_i Q M_{i,0} r_{i,0}. \tag{2}$$

Added or cancelled events are event-set discrepancies, not rate changes. The same-book counterfactual uses four depth-walked executions at Q to construct current crossing $X_{0,z}$, basis $B_{0,z}$ and fee total $\Phi_{0,z}$. Other costs $O_{0,z}$ include only costs not already represented in fills or fees. The current net scenario value is

$$\boxed{S_{0,z} = C_{0,z} + B_{0,z} - X_{0,z} - \Phi_{0,z} - O_{0,z}.} \tag{3}$$

An immediate round trip on one unchanged pair of books has $B_{0,z} = 0$ by definition. Retaining B in the identity prevents a later forecast of exit basis from being mislabeled as the same-book baseline. Equation (3) is not $\mathbb{E}[\Pi_z \mid \mathcal{F}_t]$: future rates, exit books and actual costs can differ, and the displayed current rate need not be contractually locked.

The baseline ranking orders complete scenarios by descending $S_{0,z}$ at the same amount, reporting basis, view and information version. Ties are resolved by earlier T and then a fixed route key. The quantity is not reduced merely because a smaller amount is easier to make positive, and a longer diagnostic window cannot be inserted into a single-settlement comparison. Every candidate T must be evaluated using the complete Equation (3). Table 2 collects the notation.

Table 2: Notation used in the error-budget and decision layers. Monetary terms in one comparison share currency, quantity, horizon and information version.

Symbol	Meaning
z	A complete route–quantity–horizon–information scenario.
Q	Common base quantity on both legs; never reset at exit.
T	Predeclared holding endpoint within the current view.
C_0, C	Scenario and subsequently measured signed funding cash.
B_0, B	Scenario and measured mid-price basis terms.
X_0, X	Scenario and measured four-fill crossing terms.
Φ_0, Φ	Scenario and measured trading-fee totals.
O_0, O	Other measured costs not already inside fills or fees.
S_0	Current net scenario value $C_0 + B_0 - X_0 - \Phi_0 - O_0$.
Π	Closed-position net accounting under the same scenario key.
D	Net overestimation error $S_0 - \Pi$; negative values are underestimation.
J	Research decision objective requiring joint distributions and capital preferences.
λ, ρ	Nonnegative loss weight and capital-hour price; not calibrated here.
A	Reserved capital integrated over time, in currency-hours.
$\mathcal{F}_t$	Information available at the decision cutoff.
NA	Unavailable evidence; not zero, a loss, or a complete observation.

5.2 Realization error and the break-even boundary

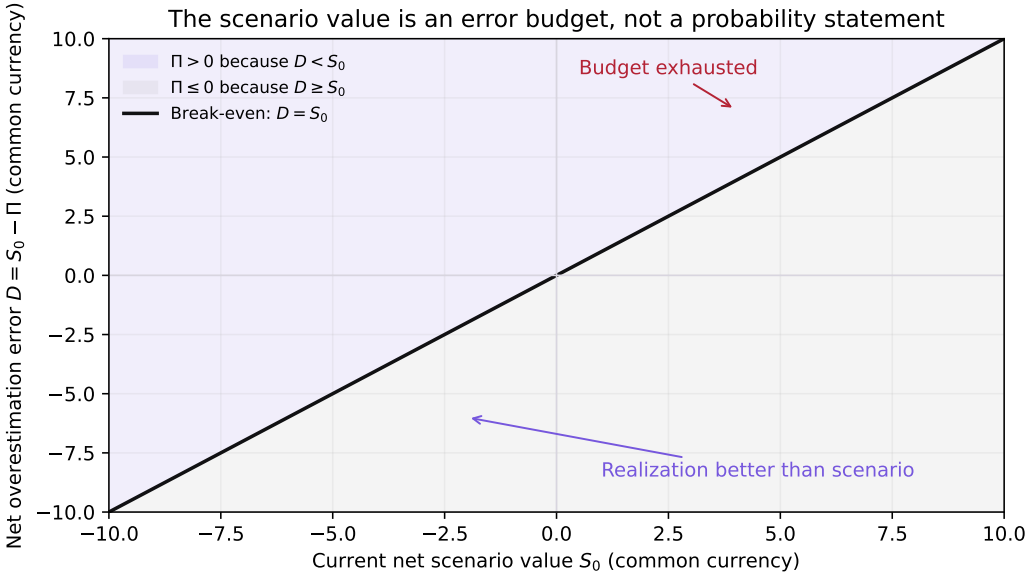
Once subsequent evidence is complete, let Π_z be the closed-position accounting under the same key. Net overestimation error is

$$\begin{aligned} D_z &= S_{0,z} - \Pi_z \\ &= (C_0 - C) + (B_0 - B) + (X - X_0) + (\Phi - \Phi_0) + (O - O_0), \end{aligned} \tag{4}$$

where the route subscript z is suppressed on the right for readability. Positive D means that the scenario was too favorable; negative D means that the subsequent accounting was better than the scenario. Direct substitution gives

$$\boxed{\Pi_z > 0 \quad \Longleftrightarrow \quad D_z < S_{0,z}.} \tag{5}$$

Thus $S_{0,z}$ is the position's error budget. The identity does not reveal D_z in advance and does not associate a probability with the budget. It does provide an exact ex post attribution of sign changes and a precise quantity for forward validation.



Source: algebraic identity Pi = S0 - D. Schematic regions; no market observations or coverage claim.

Figure 1: The break-even boundary in scenario–error space. The positive-outcome region is defined by $D < S_0$; the diagram is algebraic and carries no probability mass or coverage claim.

Funding discrepancies can be decomposed without hiding the rate–price interaction. With fixed event sets and Q, write $\Delta r_{i,\tau} = r_{i,\tau} - r_{i,0}$ and $\Delta M_{i,\tau} = M_{i,\tau} - M_{i,0}$. Then

$$C - C_0 = -\sum_{i,\tau} d_i Q \left(M_{i,0} \Delta r_{i,\tau} + r_{i,0} \Delta M_{i,\tau} + \Delta M_{i,\tau} \Delta r_{i,\tau} \right). \tag{6}$$

If event eligibility changes, it is recorded separately. If rate, price and event sets all change, an implementation either fixes a replacement order in advance or uses a symmetric attribution; it cannot assign the interaction to whichever component is most convenient.

Two sensitivity statements follow only under ceteris-paribus restrictions. If one adverse rate change affects one event on one leg and everything else is fixed, the break-even absolute rate change is $S_0/(QM)$ in signed exposure-adjusted form. With several changed events, the denominator is their signed aggregate exposure, not a single notional. If the only change is an adverse exit-basis movement, the break-even movement is S_0/Q in quote currency per base unit. These are deterministic sensitivities, not confidence intervals. Missing an exit book, official event record or required valuation price makes the corresponding attribution partial or unavailable.

5.3 The decision objective and what it does not imply

Current-scenario ranking is descriptive. A decision rule additionally needs preferences over losses and committed capital. Let $K_z(u) \geq 0$ be a capital-reservation plan declared at

decision time and let

$$A_z = \int_t^{T_z} K_z(u)\,du \tag{7}$$

be measured in currency-hours. If capital use depends on the future path, the objective instead uses $\mathbb{E}[A_z \mid \mathcal{F}_t]$. Notional is not automatically collateral. Let $\lambda \geq 0$ weight losses beyond their expected value and $\rho \geq 0$ price a currency-hour. Using $u(x) = x - \lambda(-x)^+$, which is increasing and concave, gives

$$\boxed{\begin{aligned} J_z &= S_{0,z} - \mathbb{E}[D_z \mid \mathcal{F}_t] - \lambda\mathbb{E}[(D_z - S_{0,z})^+ \mid \mathcal{F}_t] - \rho A_z \\ &= \mathbb{E}[\Pi_z \mid \mathcal{F}_t] - \lambda\mathbb{E}[(-\Pi_z)^+ \mid \mathcal{F}_t] - \rho A_z. \end{aligned}} \tag{8}$$

The first adjustment corrects mean joint error; the second penalizes realizations in the loss region rather than subtracting an arbitrary standard-deviation multiple; the third expresses the capital objective. All terms share one reporting basis. A probability, annualized rate and dollar amount cannot be added directly.

Let $\mathcal{Z}_t^J$ contain only scenarios whose required distributions, costs and capital contract are available. Within a fixed holding view, selection is

$$z^* = \arg\max_{z \in \mathcal{Z}_t^J \cup \{\varnothing\}} J_z, \qquad J_\varnothing = 0. \tag{9}$$

The no-position alternative is available only inside this fully specified comparison. If every J_z is missing, the output is "decision model unavailable," not "all opportunities have zero value." The S_0 ranking can remain visible with its own label. Partially measured scenarios form an unknown set, so the selected action cannot be called a market-wide optimum.

Proposition 1. *If Equation* (8) *uses the true conditional distribution and capital inputs, the relevant expectations are finite, and* $\lambda, \rho \geq 0$*, then* $J_z > 0$ *implies* $\mathbb{E}[\Pi_z \mid \mathcal{F}_t] > 0$*.*

Proof. Both $\lambda\mathbb{E}[(-\Pi_z)^+ \mid \mathcal{F}_t]$ and ρA_z are nonnegative. Positivity of J_z therefore requires the unconditional first term $\mathbb{E}[\Pi_z \mid \mathcal{F}_t]$ to exceed their sum and hence to be positive. □

The proposition does not imply $\Pi_z > 0$, a high win rate, or an accurate estimate $\widehat{J}_z$. Maximizing noisy estimates can itself produce upward selection bias. Nor does it show that the true conditional inputs are available in this study.

5.4 Horizon selection requires the final objective

A two-stage shortcut that first maximizes cash less cost over T and then adds a horizon-dependent basis or risk term does not solve Equation (9). Consider a deterministic example with cash/cost pairs $8/4$ at 10 minutes and $12/4$ at 60 minutes. Let

$$B(T) = -20\left(1 - e^{-T/60}\right). \tag{10}$$

The surrogate $C - \Phi$ scores 4 at 10 minutes and 8 at 60 minutes, selecting the latter. The complete means are approximately

$$S_0(10) = 8 - 4 - 3.0704 = 0.9296, \qquad S_0(60) = 12 - 4 - 12.6424 = -4.6424. \tag{11}$$

The complete objective selects 10 minutes. This constructed example is not a market observation. It demonstrates that horizon-dependent basis, risk and capital terms must enter every candidate before the maximum is taken.

5.5 Joint error estimation and coverage restrictions

A usable estimate of J requires a joint error law for both legs and all cost components. If cash-flow components form a vector Z and a contains their signed ledger coefficients,

$$\mathrm{Var}(a^\top Z \mid \mathcal{F}_t) = a^\top \mathrm{Cov}(Z \mid \mathcal{F}_t) a. \tag{12}$$

Unknown covariance is not zero covariance. Adding marginal lower confidence bounds does not create the same joint coverage level, and setting correlations to zero can be either conservative or anticonservative depending on the signed exposures.

A one-sided calibration rule illustrates both the possible construction and its limits. Given completed, contract-consistent calibration errors $D_1, \ldots, D_n$, define

$$k = \lceil (n+1)(1-\alpha) \rceil, \qquad q_\alpha = D_{(k)}, \qquad L_\alpha(z;t) = S_{0,z} - q_\alpha. \tag{13}$$

If $k > n$, the conservative quantile convention is $+\infty$, not the largest observed error. Under exchangeability, a frozen scoring rule and appropriate calibration/test isolation, a marginal one-sided coverage statement can be discussed for a future random point. The present routes exhibit temporal dependence, distribution change, shared legs and selection after viewing many candidates. Equation (13) therefore does not establish 90% or 95% coverage for a selected route.

The empirical analysis deliberately stops at component-level evidence. A full calibration target requires a closed four-fill position, matched event-time funding, complete costs and the same scenario key. A funding-rate forecast error, a fixed-notional spread proxy or an incomplete exit is not a substitute for D in the full sense. Alternative tail objectives such as CVaR would still require an explicit tail level, joint distribution and preference; they are not layered onto J as unvalidated extra deductions.

5.6 Estimators for the public component analysis

For model m and hourly actual rate r_t, point errors are $e_{m,t} = \widehat{r}_{m,t} - r_t$. I report

$$\mathrm{MAE}_m = n^{-1} \sum_t |e_{m,t}|, \qquad \mathrm{RMSE}_m = \sqrt{n^{-1} \sum_t e_{m,t}^2}. \tag{14}$$

The alternatives are the previous settled rate, an expanding mean and trailing medians. Diebold–Mariano statistics compare mean loss differentials $\bar{\ell}$ using

$$\mathrm{DM} = \frac{\bar{\ell}}{\sqrt{\widehat{\mathrm{Var}}_{\mathrm{NW}}(\bar{\ell})}}, \qquad \ell_t = |e_{1,t}| - |e_{2,t}|, \tag{15}$$

with a 23-lag Newey–West variance in the main specification. The lag covers one calendar day of hourly dependence; alternatives of 0, 6 and 47 lags and squared loss are retained as robustness checks. Holm adjustment covers all 12 absolute-loss comparisons and, separately, all 12 squared-loss comparisons.

For a central interval, calibration residuals $r_t - \widehat{r}_t$ supply empirical 5th and 95th percentiles. Later intervals are scored by coverage and the interval score

$$\begin{aligned} IS_\alpha(l,u;y) = (u-l) \\ + \frac{2}{\alpha}(l-y)\mathbf{1}(y<l) \\ + \frac{2}{\alpha}(y-u)\mathbf{1}(y>u), \end{aligned} \tag{16}$$

with $\alpha = 0.10$. I report Kupiec's unconditional-coverage test and Christoffersen's independence and conditional-coverage diagnostics with Holm adjustment across six series. These are retrospective diagnostics, not distribution-free guarantees under temporal drift. No probabilistic density is estimated, so CRPS and log score are not applicable.

The spread analysis defines a funding-only scenario S_0^C and subsequent proxy Π^C at fixed notional. For each asset-hour it reports all candidates and the maximum separately, including mean error, MAE, RMSE and coverage of the calibration one-sided q90. The fixed-notional conversion provides a readable scale only; it is not complete Π and does not include costs, quantity closure or capital.

5.7 Prospective evaluation and NA denominators

A valid economic evaluation must separate diagnostic data D0, calibration D1, rule-selection D2 and locked forward data D3. Previously inspected observations belong to D0 even if later subdivided. Before the first D3 label can be seen, the candidate population, amount, endpoints, fees, delays, missing-data rules, comparison rule, trial ledger, analysis code hash and stopping rule must be fixed. Repeated inspection converts a holdout into development data.

Each observation carries a label interval from the first entry fill to the last exit or required cash event. Training or selection observations whose label intervals overlap a validation or test interval are purged. Feature lookbacks and normalization statistics are cut off at the decision time. If non-forward folds are used, an embargo reflects maximum holding and label-delay requirements; a chronological D3 does not claim that an embargo eliminates every form of market dependence.

Uncertainty is computed on complete capital paths over predeclared day or time blocks, with asset-family and shared funding-event groups preserved. Paired strategies move together in each resample. Repeated directions, amount tiers and publications of one event are not independent replications. A separate unseen-asset test, if conducted, keeps every route of an asset family on one side of the split and is reported separately from the future-time test.

The reporting funnel retains the complete candidate index and mutually exclusive primary states: comparable identity, schedule available, payable event, requested depth, successful depth, selected, execution attempted, both legs filled, fully exited and complete official label. Unmeasured, unsupported, stale, insufficient depth, unfilled, pending settlement, missing official event and missing exit remain distinct. A confirmed no-fill can have zero position P&L under a registered strategy while retaining its opportunity status; an open position with a missing label cannot be zero-filled.

Finally, shared books and capital are consumed once. Simultaneous routes using the same venue–instrument–side are aggregated before depth walking, and insufficient capital is a rejection reason rather than permission to open every route and sum the outcomes. Opposite directions and mutually exclusive horizons are counterfactuals, not portfolio positions. These restrictions define the evidence needed to test J; the present public component analyses do not claim to have completed that test.

6 Results

The results answer four distinct questions. How accurately does a transparent benchmark predict the next funding rate? Does calibration-period uncertainty survive chronological evaluation? What changes when the largest scenario is selected? And how does a complete public-book scenario reconcile with its subsequent public paper outcome? Keeping these questions separate prevents a funding-rate result from being mistaken for a complete strategy result.

6.1 Previous-settled errors are small but heterogeneous

Table 3 reports holdout point errors. Previous-settled MAE ranges from 0.0126 basis points for Extended SOL to 0.0298 for Hyperliquid ETH; RMSE ranges from 0.0312 to 0.0543 basis points. At a fixed 10,000-unit scenario notional, the MAE range is 0.0126–0.0298 currency units per event. This conversion communicates scale only. It excludes the event valuation price, both legs, execution costs and capital and therefore is not economic P&L.

Table 3: Holdout point-error metrics for funding-rate components. MAE/RMSE are in basis points of rate. All series have $n = 620$. In the text, previous-settled MAE is also converted at fixed 10,000 USD-denominated scenario notional under a one-for-one USDC/USD parity assumption; that conversion is not account P&L.

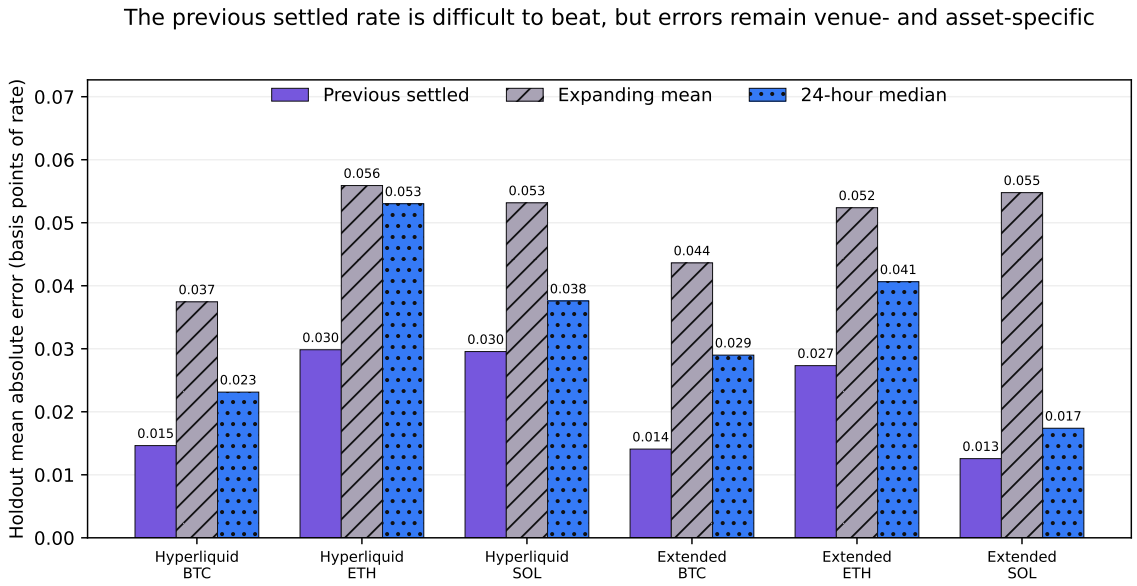
Venue	Asset	Previous MAE	Previous RMSE	Mean MAE	24h median MAE
Hyperliquid	BTC	0.0146	0.0331	0.0374	0.0231
Hyperliquid	ETH	0.0298	0.0531	0.0559	0.0530
Hyperliquid	SOL	0.0296	0.0543	0.0532	0.0376
Extended	BTC	0.0141	0.0368	0.0436	0.0290
Extended	ETH	0.0273	0.0522	0.0524	0.0406
Extended	SOL	0.0126	0.0312	0.0548	0.0174

The expanding mean has higher MAE than the previous-settled rate in every series. The trailing 24-hour median lies between the two in some series and above both in others. Figure 2 shows that these differences are not dominated by one venue or asset. The expanding mean's disadvantage is consistent with regime variation: older observations can be less representative of the next event than the most recent settlement. That interpretation is a mechanism hypothesis, not proof of a particular funding-rule change.

Table 4 formalizes the absolute-loss comparisons. Previous-settled prediction has significantly lower loss than the expanding mean in all six series after Holm adjustment. It has significantly lower loss than the 24-hour median in the three Hyperliquid series, while none of the Extended comparisons clears 5% after adjustment. These statements concern predictive loss, not a tradable advantage. Section 7 shows that squared loss yields fewer significant differences.

6.2 Coverage failure is more informative than average error

Average error does not establish a reliable error allowance. Table 5 and Figure 3 show central-interval coverage for intervals whose residual quantiles were fixed on calibration data. Extended BTC attains 92.74%, Extended ETH 86.29%, Extended SOL 97.90%,



Source: public hourly funding histories; holdout 2026-10-09 to 2026-11-03, 620 predictions per series. Retrospective component analysis.

Figure 2: Holdout MAE for three transparent funding-rate estimators. All 620 later predictions in each venue–asset series are included. Source: frozen public histories; retrospective component analysis.

Table 4: Diebold–Mariano comparisons on holdout absolute errors. A negative statistic favors the previous-settled benchmark. Newey–West uses 23 hourly lags; Holm adjustment covers all 12 absolute-loss comparisons. Squared-loss and lag alternatives are reported in the reproduction outputs.

Venue	Asset	Alternative	Loss difference (bp)	DM	Raw p	Holm p
Extended	BTC	expanding mean	-0.0295	-6.353	2.1e-10	1.7e-09
Extended	BTC	24h median	-0.0149	-1.914	0.056	0.111
Extended	ETH	expanding mean	-0.0251	-6.632	3.3e-11	3.0e-10
Extended	ETH	24h median	-0.0133	-2.384	0.017	0.051
Extended	SOL	expanding mean	-0.0422	-13.829	1.7e-43	2.0e-42
Extended	SOL	24h median	-0.0048	-1.809	0.070	0.111
Hyperliquid	BTC	expanding mean	-0.0228	-8.335	7.7e-17	8.5e-16
Hyperliquid	BTC	24h median	-0.0085	-3.337	8.5e-04	0.004
Hyperliquid	ETH	expanding mean	-0.0261	-4.225	2.4e-05	1.7e-04
Hyperliquid	ETH	24h median	-0.0232	-3.433	6.0e-04	0.004
Hyperliquid	SOL	expanding mean	-0.0236	-7.410	1.3e-13	1.3e-12
Hyperliquid	SOL	24h median	-0.0080	-3.130	0.002	0.007

Hyperliquid BTC 91.13%, Hyperliquid ETH 68.39%, and Hyperliquid SOL 91.13%, against a nominal 90%.

Table 5: Holdout interval and one-sided error-budget coverage. Central intervals target 90%; p-values use Holm adjustment across six series. The one-sided calibration order statistic targets 90% overestimate coverage and is not a selection-conditional guarantee.

Venue	Asset	Central cov.	Kupiec p	Cond. p	One-sided cov.	Exc.
Extended	BTC	92.74%	0.052	6.1e-09	93.23%	42
Extended	ETH	86.29%	0.013	6.6e-16	88.39%	72
Extended	SOL	97.90%	2.1e-14	8.1e-16	96.94%	19
Hyperliquid	BTC	91.13%	0.681	0.003	90.16%	61
Hyperliquid	ETH	68.39%	1.3e-48	1.2e-52	70.32%	184
Hyperliquid	SOL	91.13%	0.681	0.005	90.65%	58

For the two-venue funding-spread proxy, the same calibration one-sided rule covers 94.03% of all holdout counterfactuals but 88.39% of the selected maxima. These are fixed-notional funding components, not complete trades.

Hyperliquid ETH is the strongest counterexample. Its calibration residual distribution is too narrow for the later sample, leaving 38.71% of observations outside the central interval in one or both tails. Extended SOL's 97.90% is not correspondingly "better" in a decision sense: overly wide intervals can achieve high coverage while providing little useful discrimination. Conditional-coverage diagnostics reject in every series after Holm adjustment, indicating that unconditional percentages alone conceal dependence in hits and misses. A coverage level therefore cannot be read as a per-route promise.

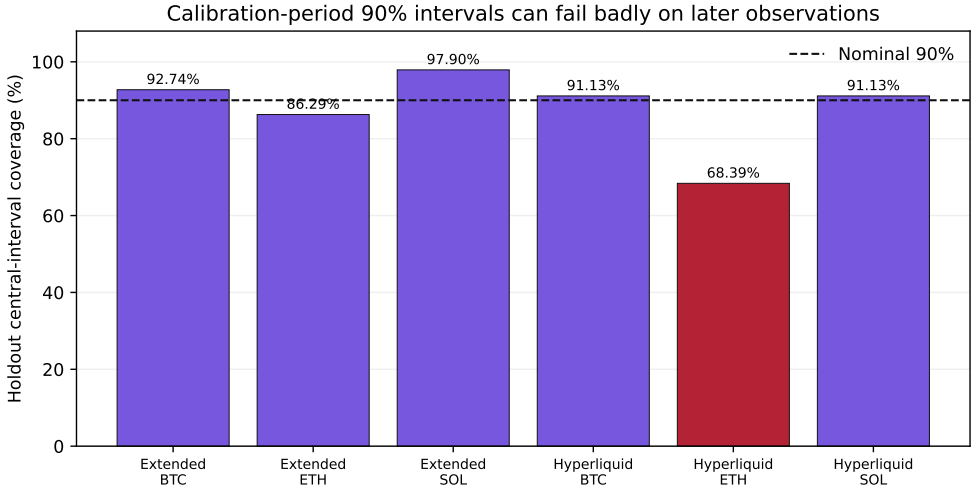


Figure 3: Holdout coverage of calibration-period central 90% intervals. The dashed line is nominal rather than guaranteed coverage. Source: 620 later observations per series.

The one-sided results lead to the same conclusion. At the 0.90 level, holdout coverage ranges from 70.32% for Hyperliquid ETH to 96.94% for Extended SOL. The calibration order statistic is mechanically correct, but the assumed stability of the error distribution is not. This is why the error budget is an exact accounting quantity while a calibrated allowance for that budget remains a separate empirical problem.

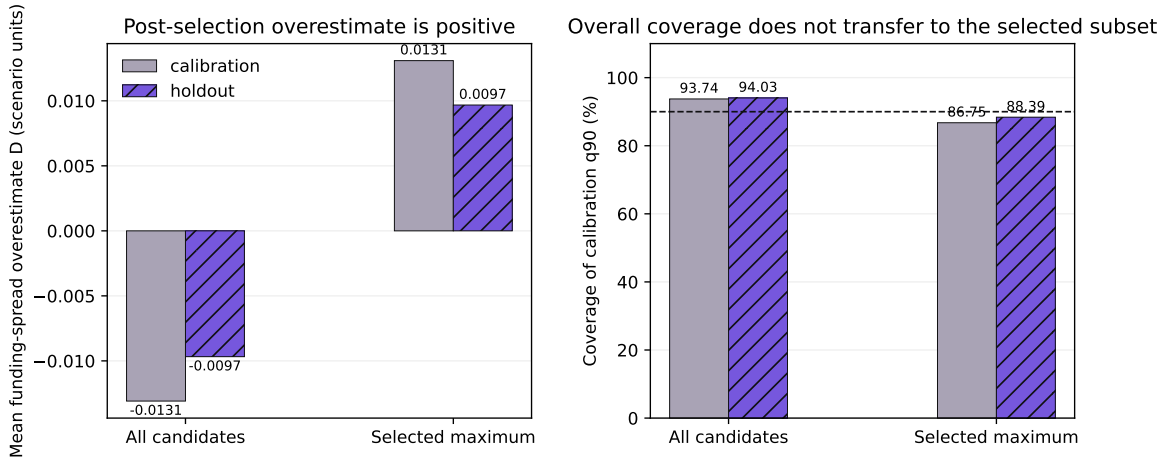
6.3 Selection changes the error distribution

Table 6 reports the two-direction funding-spread exercise. In the holdout, selected maxima have mean scenario spread 0.0523 currency units and mean subsequent spread proxy 0.0427, producing mean overestimation 0.0097. All scorable candidates have the opposite signed means because both directions are present; this symmetry is why pooling them would conceal the selection issue. Selected MAE is 0.0349 and RMSE 0.0612, but these are fixed-notional funding components rather than profits or losses after costs.

Table 6: Post-selection funding-spread errors at fixed 10,000 USD-denominated scenario notional per leg under a one-for-one USDC/USD parity assumption. The selected row is the larger of two opposite-direction scenario spreads in each asset-hour. No fees, basis, execution, capital or full-profit interpretation is attached.

Period	Subset	n	Mean scen.	Mean proxy	Mean D^C	MAE	q90 cov.
Calibration	All candidates	5,466	-0.0522	-0.0391	-0.0131	0.0368	93.74%
Calibration	Selected maximum	2,733	0.0522	0.0391	0.0131	0.0368	86.75%
Holdout	All candidates	3,720	-0.0523	-0.0427	-0.0097	0.0349	94.03%
Holdout	Selected maximum	1,860	0.0523	0.0427	0.0097	0.0349	88.39%

The calibration q90 is 0.07 currency units. It covers 94.03% of all holdout counterfactuals but only 88.39% of selected maxima. The calibration period shows the same direction: 93.74% overall and 86.75% after selection. This persistent gap is direct evidence that an overall calibration statement need not transfer to the subset used for decisions. Figure 4 displays both the positive post-selection overestimate and the coverage shortfall.



Source: two opposite-direction counterfactuals per asset-hour, fixed 10,000 USD-denominated scenario notional per leg; USDC/USD parity assumed. Funding-only proxy, not complete P&L.

Figure 4: Post-selection mean error and calibration-q90 coverage. The selected maximum is evaluated separately from all counterfactuals in each asset-hour. Source: fixed-notional funding-spread proxies, not complete trades.

The exercise does not estimate the optimizer's curse in a complete trading model, because full costs and joint errors are absent. It demonstrates the narrower mechanical point: taking a maximum selects a different error population. A valid recommendation evaluation must preserve that selected population instead of reporting only the more favorable full-candidate average.

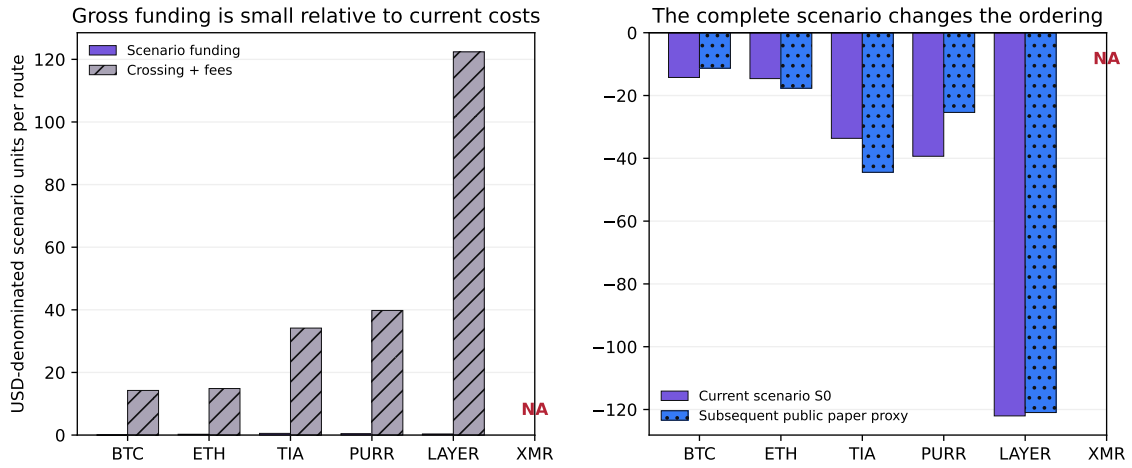
6.4 The forward ledger turns the identity into an audit

Table 7 reports the frozen one-hour illustration. Scenario funding is positive for all five complete routes, ranging from 0.005 currency units for BTC to 0.534 for TIA. Current crossing plus public base-taker fees is much larger, from 14.25 for BTC to 122.39 for LAYER. The resulting S_0 values are all negative and range from -14.25 to -122.04. The complete-scenario order is BTC, ETH, TIA, PURR and LAYER, rather than an order based on the largest raw funding gap. XMR has no complete scenario because displayed depth is insufficient.

Table 7: Forward public-book paper illustration for the 23:00 UTC settlement on 2026-11-03. Cost includes current crossing and public base-taker fee scenarios. Subsequent values use public books and settlement-price proxies, not orders or account cash. Source snapshots frozen at 22:41 UTC. All amounts are USD-denominated scenario units under the stated USDC/USD parity assumption.

Asset	Long / short venue	Funding C_0	Current cost	S_0	Subsequent proxy	D
BTC	Hyperliquid / Extended	0.005	14.25	-14.25	-11.30	-2.95
ETH	Extended / Hyperliquid	0.255	14.86	-14.61	-17.68	3.07
TIA	Extended / Hyperliquid	0.534	34.17	-33.64	-44.45	10.81
PURR	Extended / Hyperliquid	0.452	39.80	-39.35	-25.38	-13.97
LAYER	Hyperliquid / Extended	0.350	122.39	-122.04	-120.97	-1.07
XMR	Extended / Hyperliquid	Depth incomplete at t0; no forward label				

The subsequent public paper proxies are also all negative, ranging from -11.30 for BTC to -120.97 for LAYER. Their order is BTC, ETH, PURR, TIA and LAYER, so PURR and TIA exchange positions. This is a descriptive ordering of five correlated route scenarios, not a ranking of realized investments. Figure 5 contrasts funding with current costs and then S_0 with the subsequent proxy.



Source: frozen public books and current rates at 2026-11-03 22:41 UTC; settled records and exit books after 23:00 UTC. No orders or account cash.

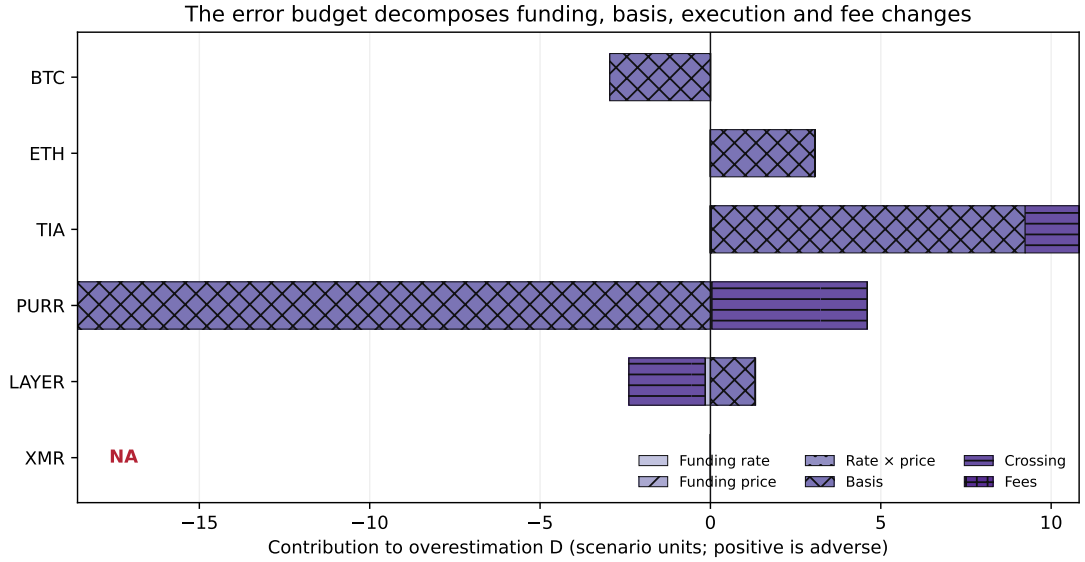
Figure 5: Current funding and costs on the left; current scenario and subsequent public paper proxy on the right. All values are USD-denominated scenario units under the stated parity assumption. No orders or account cash are used.

The error decomposition in Table 8 shows why a single slippage label is inadequate. PURR has $D = -13.97$: an 18.57-unit basis improvement more than offsets 4.55 units of increased crossing and small funding and fee changes. TIA has $D = 10.81$, driven by 9.21 units of basis deterioration and 1.58 units of increased crossing. BTC and ETH have almost no cross-

ing change but opposite basis contributions. LAYER's 2.23-unit crossing improvement offsets part of its basis deterioration and funding-rate error.

Table 8: Forward error attribution. Positive components increase overestimation. Funding price includes the rate–price cross term; the total also includes the displayed funding-rate component. All amounts are USD-denominated scenario units under the stated USDC/USD parity assumption, not account cash.

Asset	Rate	Funding price	Basis	Crossing	Fee	Total D	Identity residual
BTC	0.000	-0.000	-2.958	0.000	0.012	-2.947	4.00e-16
ETH	0.000	-0.000	3.057	0.000	0.012	3.069	1.00e-15
TIA	0.030	0.001	9.207	1.583	-0.010	10.811	5.80e-15
PURR	0.050	-0.001	-18.575	4.551	0.003	-13.971	9.10e-15
LAYER	-0.162	-0.001	1.311	-2.233	0.017	-1.068	0.00e+00



Source: equation-level reconciliation of the same forward public paper scenarios. Components sum to D up to displayed rounding; missing rows are not zero-filled.

Figure 6: Forward error attribution by component. Positive values increase overestimation; negative values indicate a more favorable subsequent proxy. Source: the same frozen public paper scenarios and subsequent evidence.

Every completed row satisfies the sign identity, and the largest absolute numerical reconciliation residual is 9.1×10^{-15} scenario units. This verifies the arithmetic implementation, not the economic model. The five negative subsequent proxies do not show that positive routes do not exist, and PURR's improvement does not show that negative scenarios are attractive. The useful result is that each discrepancy can be traced to a declared assumption and recomputed independently.

6.5 What the combined evidence establishes

The algebra and data jointly support three restrained conclusions. First, S_0 is a well-defined error budget only when scenario and outcome share the full accounting key. Second, simple funding benchmarks can expose later error, but calibrated uncertainty and selected-subset behavior can fail even when average errors look small. Third, complete costs and basis movements can dominate raw funding gaps and change the ordering of

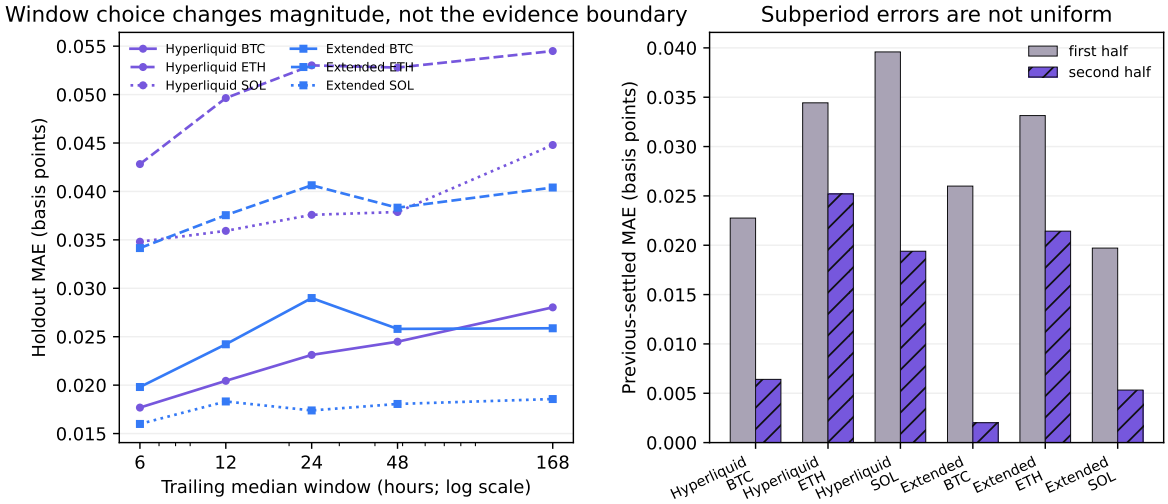
candidates. The evidence does not establish a positive-expectation strategy, a calibrated J, a market-wide optimum or a universal threshold for admission.

7 Robustness and sensitivity

The robustness question is not whether one flattering specification survives. It is whether the interpretation remains appropriately limited when the sample, estimator, loss and dependence choices change. Table 9 summarizes the full grid; every run is retained in the reproduction outputs, including unfavorable comparisons.

Table 9: Robustness grid summarized across all six venue–asset series. Values are mean, minimum and maximum series-level holdout MAE in basis points. The complete unaggregated grid, squared-loss tests and Newey–West lag sensitivity accompany the paper.

Dimension	Specification	Mean MAE	Minimum	Maximum
Trailing median window	6h	0.0275	0.0160	0.0428
Trailing median window	12h	0.0310	0.0183	0.0496
Trailing median window	24h	0.0335	0.0174	0.0530
Trailing median window	48h	0.0329	0.0181	0.0528
Trailing median window	168h	0.0354	0.0186	0.0545
Previous-settled subperiod	first half	0.0293	0.0197	0.0396
Previous-settled subperiod	second half	0.0133	0.0020	0.0252



Source: all five rolling windows and two predetermined-for-analysis holdout halves. Every specification is reported; no best-only selection.

Figure 7: Rolling-window and holdout-subperiod sensitivity. All specifications and all six series are retained; the figure reports component error rather than strategy performance.

7.1 Subperiods and venue–asset groups

Splitting the holdout at 2026-10-22 changes the scale of previous-settled errors. Mean series-level MAE is 0.0293 basis points in the first half and 0.0133 in the second. The corresponding ranges are 0.0197–0.0396 and 0.0020–0.0252 basis points. A single full-holdout average therefore conceals substantial time variation. We do not infer a permanent improvement from the quieter second half or select it as a preferred sample.

The main tables already report all six venue–asset series rather than pooling them into one apparently precise estimate. Hyperliquid ETH is the clearest coverage failure, while Extended SOL has unusually high central coverage. These differences prevent a venue-independent calibration claim. The two-venue spread selection result is also reported by period rather than treated as independent evidence across assets that share hourly market shocks.

7.2 Alternative estimators and window sensitivity

The trailing median window varies over 6, 12, 24, 48 and 168 hours. Mean series-level MAE rises from 0.0275 basis points at 6 hours to 0.0354 at 168 hours, but the ranking is not uniform in every series and the shortest window is not declared optimal. The expanding mean is less accurate than the previous-settled benchmark on absolute error in every main-series comparison. These results favor simple, local benchmarks in this sample; they do not justify restoring a long historical average as an admission condition.

The alternative-estimator grid also demonstrates why model counts matter. Three point estimators, five median windows, four one-sided levels, two loss functions, four Newey–West lag choices and two holdout subperiods are all disclosed. Selecting only the 24-hour median or only the most favorable subperiod would change the apparent conclusion without changing the underlying evidence.

7.3 Loss function and dependence choices

Absolute-loss Diebold–Mariano tests favor the previous-settled rate over the expanding mean in every series after Holm adjustment. It also favors the previous-settled rate over the 24-hour median in three series at the 5% level after adjustment. Squared loss materially weakens that conclusion: only Extended SOL still favors previous-settled over expanding mean after adjustment, and no series favors previous-settled over the 24-hour median after adjustment. Large errors therefore carry information that an MAE-only summary would suppress.

Dependence choices matter as well. With zero or six Newey–West lags, absolute-loss comparisons against the 24-hour median are uniformly significant before adjustment; at 23 lags, three raw p-values exceed 0.05, and at 47 lags the largest raw p-value is 0.1002. The main 23-lag specification is deliberately more conservative than an independence assumption, but no finite lag proves that all shared-event dependence has been removed. Complete strategy-level inference still requires the grouped capital-path design in Section 5.

7.4 Coverage-level sensitivity and residual fragility

One-sided levels of 0.80, 0.90, 0.95 and 0.99 are evaluated without retuning the quantiles to the holdout. The minimum series-level holdout coverage rises from 68.39% at the 0.80 level to 97.42% at 0.99, while the maximum rises from 92.58% to 100%. Higher nominal levels naturally produce wider or more conservative allowances; they do not repair a shifted distribution, and their greater coverage can be economically uninformative.

Taken together, the checks support a *conditional* conclusion: the error-budget identity is exact, simple funding benchmarks are useful for exposing measurement error, and coverage or ranking behavior is fragile across time, series and loss choices. They do not support

a robust profitability claim, a universal confidence multiplier or a validated J rule. The most damaging perturbation for the calibration argument is the Hyperliquid ETH regime shift, because it turns nominal 90% central coverage into 68.39% and also produces the weakest one-sided coverage.

8 Limitations

8.1 The complete decision rule remains unvalidated

The central limitation is not a missing refinement but missing evidence for the complete decision problem. The public histories measure funding rates, not the joint distribution of closed-position errors. The fixed-notional spread exercise excludes basis, depth, fees and capital. The forward illustration closes a public paper ledger for one event, but it is too small and too dependent to estimate a distribution, calibrate λ or ρ, or demonstrate positive expectation. No independently preregistered D3 sample of complete positions has been collected. Accordingly, J remains a research object rather than a validated selection rule.

The proposition about J is conditional on true conditional distributions and capital inputs. Estimated substitutes can behave differently, especially after maximization. A positive current S_0 is neither a positive expectation nor a positive realized outcome, and an unavailable J is not evidence that the market contains no opportunity. These limitations cannot be repaired by choosing a confidence multiplier, a CVaR level or a loss aversion parameter after inspecting the results.

8.2 Retrospective design and statistical limits

The chronological component analysis is retrospective and was not externally preregistered. Its boundary avoids direct look-ahead but does not recreate the evidentiary status of a locked forward protocol. Diebold–Mariano tests use asymptotic normal approximations and a chosen Newey–West lag; coverage tests use binary-hit approximations even though violations can be clustered and rates highly discrete. Holm adjustment addresses only the stated test families, not every possible analysis that a researcher could imagine. PBO, White's reality check and Hansen's SPA are not computed because the asynchronous funding-component rows do not form the required synchronized complete-strategy return matrix.

The sample covers only two venues, three continuing assets and approximately two months. It cannot establish cross-venue generality, survival-adjusted coverage, long-run regime stability or future exchange API semantics. Asset-family and time shocks remain dependent despite hourly alignment. The selected-spread analysis is useful for exposing mechanical post-selection differences, but its two directions per asset-hour are counterfactuals rather than independent opportunities.

8.3 Public-proxy and execution limitations

The historical endpoints supply rates and timestamps but not account cash. The forward proxy uses later public oracle or mark snapshots rather than an official account-funding record. Public base taker rates are not an observed account fee tier, and public-book

VWAP does not include queue position, order reaction, replenishment, hidden liquidity, partial fills or latency after an order is sent. Missing source-clock fields are not recovered by substituting local receipt time. A one-hour target and a 10-second record-matching tolerance are example protocols, not universal exchange rules.

The mixed-currency reporting assumption is also restrictive. One-for-one USDC/USD parity makes a common arithmetic illustration possible but is not event-time conversion evidence and is not realized USD, USDT or USDC P&L. Transfer, borrow, financing, liquidation, collateral and capital-network constraints remain outside the measured boundary. No margin or liquidation model is inferred from public books.

The forward snapshot freezes only a local file and hash before the outcome. It lacks an independent public timestamp before all results and therefore is not represented as full external preregistration. The six assets were chosen by a disclosed current-rate rule, not sampled randomly. One asset is depth-incomplete and five are complete, so any complete-case economic statement would additionally be selected on observability; we avoid such a statement.

8.4 Review and reproduction boundaries

Reproduction means recomputation from the included public raw responses and frozen scenario files. It does not mean that a third party has independently rerun the collection, verified every venue's account-level semantics or reproduced real orders. Official documentation and APIs can change after the access dates. The included hashes identify the evidence used here rather than guaranteeing identical future responses.

The manuscript receives internal multi-role technical review, not external peer review. Local text-similarity screening is not a commercial plagiarism certification. These are process limitations rather than reasons to weaken the algebra: the identity remains exactly testable, and the public measurements remain reproducible. They are reasons to keep the economic conclusion explicitly open until a preregistered, complete, forward evaluation exists.

9 Conclusion

A current funding-carry ranking should answer a modest question precisely before it attempts a larger one. The modest question is what a complete scenario is worth under a fixed quantity, event schedule, book and fee basis. The error-budget identity then states exactly how much joint overestimation the subsequent position can tolerate: $\Pi > 0$ if and only if $D < S_0$. This makes funding, basis, crossing, fees and other costs reconcilable components of one break-even condition rather than independent adjustments to a displayed rate.

The larger question requires more evidence. The objective J combines expected closed-position value, loss aversion and the opportunity cost of reserved capital. Without a joint error distribution and a declared capital contract, it is unavailable rather than zero. Its conditional proposition does not supply a probability of profit, and choosing a horizon by a surrogate that omits horizon-dependent terms can reverse the optimum. Descriptive S_0 ranking and decision-ready J selection should therefore remain separate outputs.

The public measurements reinforce that distinction. Previous-settled funding rates provide a useful transparent benchmark, but later errors and coverage vary across series and

time. Calibration coverage does not automatically transfer to the selected maximum, and a complete current-cost calculation can reorder candidates that looked attractive on funding gap alone. The appropriate response is not another universal threshold. It is a frozen evaluation contract with complete candidates, chronological isolation, shared-event dependence handling, post-selection metrics and explicit NA states.

The paper consequently ends with a testable research agenda rather than a trading claim. First validate the scenario and error ledger on complete public or account-consistent outcomes. Then estimate joint error on data isolated from rule selection, evaluate the selected subset as well as all candidates, and only afterward consider whether J improves a frozen baseline on locked forward capital paths. Until that evidence exists, the contribution is an exact error budget and an evaluation discipline, not a validated arbitrage recommendation.

Declarations

Corresponding author and research channel

Correspondence: humarb Research, London, United Kingdom · `https://humarb.ai` · admin@humarb.ai. This paper was completed by humarb Research; humarb.ai is its research channel. The named author is Hang Liu, ORCID 0009-0008-1059-3926.

Competing interests

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.

Funding

This research received no external funding; it was conducted by humarb Research.

Data availability

The institutional description uses official Hyperliquid and Extended documentation accessed on 3–4 November 2026. The historical analysis uses frozen public hourly funding responses from 2026-09-01 through 2026-11-03. The forward illustration uses public current statistics, level-two books and history responses frozen around the 23:00 UTC settlement on 2026-11-03. Raw responses, normalized data, source hashes, a data dictionary and acquisition notes accompany the paper. No credentials, user data, private endpoints or account records are included.

Code availability

A clean-room public-research implementation accompanies the paper. It rebuilds the history, computes point and interval diagnostics, reconciles the forward error budget, and

generates every table and vector figure. It contains no order submission, credentials, private infrastructure, production rules or production parameters. Core analysis runs offline from the included evidence.

Generative AI statement

Generative AI tools assisted with literature discovery, language drafting and editing, code drafting, and multi-role technical review. Source selection, numerical provenance, confidentiality boundaries, corrections and final responsibility rest with Hang Liu and humarb Research. No AI system is listed as an author. The multi-role review included in the handoff is an internal AI-assisted quality check, not external peer review.

Originality and review status

The manuscript was newly structured around the error-budget and decision-layer contribution and checked locally for textual overlap with the supplied companion materials. The local screen is not iThenticate, Turnitin or a comprehensive internet plagiarism certification. The paper is an author-handoff preprint and has not undergone external peer review. The DOI and Zenodo record are given on the title page.

Research and investment disclaimer

This paper is for research and methodological information only. It does not provide investment, legal, accounting or tax advice; does not recommend a trade, venue, leverage level or position size; and does not promise income, execution, capital protection or protection from liquidation. Public-book paper values and fixed-notional proxies are not live-trading outcomes. Digital-asset perpetual positions can involve substantial market, liquidity, operational, counterparty and legal risks.

Acknowledgements

The research and manuscript were completed within humarb Research's collaborative research process. humarb.ai is the research channel used to document and distribute the work. The author retains responsibility for the final manuscript and its evidence boundaries.

References

References

Ackerer, D., Hugonnier, J., and Jermann, U. (2026). Perpetual futures pricing. *Mathematical Finance*, 36(3):481–499. Journal metadata verified; arXiv:2310.11771v2 consulted.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *The Journal of Risk*, 3(2):5–39. December 2000 author manuscript consulted.

Angelopoulos, A. N. and Bates, S. (2022). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv:2107.07511v6, 7 December 2022; fixed version consulted.

Bailey, D. H., Borwein, J. M., López de Prado, M., and Zhu, Q. J. (2017). The probability of backtest overfitting. *The Journal of Computational Finance*, 20(4):39–69. 27 February 2015 author manuscript consulted.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Extended (2026a). Extended api documentation. Official API documentation; accessed 3–4 November 2026.

Extended (2026b). Trading fees and rebates. Official fee documentation; accessed 3–4 November 2026.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). …and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

He, S., Manela, A., Ross, O., and von Wachter, V. (2022). Fundamentals of perpetual futures. arXiv:2212.06888v7, 17 September 2026; fixed version consulted.

Hyperliquid (2026a). Fees. Official fee documentation; accessed 3–4 November 2026.

Hyperliquid (2026b). Funding. Official protocol documentation; accessed 3–4 November 2026.

Hyperliquid (2026c). Perpetuals: Info endpoint. Official API documentation; accessed 3–4 November 2026.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Liu, H. (2026a). Event-time accounting for two-leg perpetual positions: Quantity closure, asynchronous settlement and evidence levels. humarb Research 10; Zenodo preprint.

Liu, H. (2026b). Framing net carry for cross-venue perpetual funding positions: A settled-history, depth-walked, and liquidation-aware method. humarb Research 02; Zenodo preprint.

Liu, H. (2026c). Measuring executable net carry in cross-venue perpetual funding-rate arbitrage. humarb Research 01; Zenodo preprint.

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319.

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

Perold, A. F. (1988). The implementation shortfall. *The Journal of Portfolio Management*, 14(3):4–9.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41. 5 September 1999 author manuscript consulted.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1):35–55.

Smith, J. E. and Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

A Additional derivations

A.1 Closed-position and crossing identities

For completeness, this appendix records the imported ledger identity without reasserting it as the paper's novel contribution. Let P_i^e and P_i^x be entry and exit VWAPs on leg i, with attributed mid-prices m_i^e and m_i^x. Under common Q and full closure,

$$\Pi = Q(P_L^x - P_L^e) + Q(P_S^e - P_S^x) + C - \Phi - O. \tag{17}$$

Define

$$X = Q\left[(P_L^e - m_L^e) + (m_L^x - P_L^x) + (m_S^e - P_S^e) + (P_S^x - m_S^x)\right], \tag{18}$$

$$B = Q\left[(m_S^e - m_L^e) - (m_S^x - m_L^x)\right]. \tag{19}$$

Substitution gives $Q(P_L^x - P_L^e) + Q(P_S^e - P_S^x) = B - X$. Therefore $\Pi = B - X + C - \Phi - O$. Crossing is already embedded in the four execution prices; subtracting it again would violate the identity.

A.2 Equivalence of the two decision-objective forms

The loss-region identity is $(-\Pi)^+ = (D - S_0)^+$ because $\Pi = S_0 - D$. Since S_0 is known at t,

$$J = \mathbb{E}[\Pi \mid \mathcal{F}_t] - \lambda \mathbb{E}[(-\Pi)^+ \mid \mathcal{F}_t] - \rho A \tag{20}$$

$$= S_0 - \mathbb{E}[D \mid \mathcal{F}_t] - \lambda \mathbb{E}[(D - S_0)^+ \mid \mathcal{F}_t] - \rho A. \tag{21}$$

For $x \geq 0$, $u(x) = x$; for $x < 0$, $u(x) = (1 + \lambda)x$. The left derivative at zero is $1 + \lambda$ and the right derivative is 1, so u is concave when $\lambda \geq 0$. Increasingness follows from positive one-sided slopes. This utility is a transparent research device, not an elicitation of any user's actual preferences.

A.3 Exact values in the horizon reversal

At $T = 10$, $B(10) = -20(1 - e^{-1/6}) \approx -3.070356$; at $T = 60$, $B(60) = -20(1 - e^{-1}) \approx -12.642412$. Subtracting the same cost of 4 from cash 8 and 12 and then adding the corresponding basis yields 0.929644 and -4.642412 after rounding. The displayed values in the main text use four decimal places. Changing only the order in which terms are introduced cannot change the true maximum; the reversal identifies an inconsistent objective, not a parameter to be tuned.

A.4 Why marginal coverage need not cover a selected subset

Suppose a population has 100 equally weighted candidates. Ninety have valid one-sided coverage and ten have none. Overall coverage is 90%. If the selection rule always chooses from the ten uncovered candidates, selected-subset coverage is zero. No contradiction arises because the 90% statement is marginal over a different distribution from the selected one. The public spread analysis is an empirical analogue of this distinction, with coverage of 94.03% for all holdout counterfactuals and 88.39% for selected maxima. It does not assert that the selected conditional coverage must equal 90%.

A.5 CVaR as an alternative, not an additional free penalty

For loss $L = -\Pi = D - S_0$ and a specified level β, one can define

$$\mathrm{CVaR}_\beta(L) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{\mathbb{E}[(L-\eta)^+]}{1-\beta} \right\}. \tag{22}$$

A possible alternative ranking value is $R_\beta = S_0 - \mathrm{CVaR}_\beta(D) = -\mathrm{CVaR}_\beta(-\Pi)$. This is a different objective from J, not a penalty to be stacked on top without a preference rationale. With $D = 0$ having probability 0.99 and $D = 100$ probability 0.01, $\beta = 0.95$ gives $\mathrm{CVaR}_\beta(D) = 20$ by Equation (22); averaging only losses strictly above a naive VaR threshold would give 100 and mishandle the probability atom. Neither calculation supplies β or a future distribution for the present market.

A.6 Missing-label bounds have limited scope

If the only target is a positive-result proportion and known labels truly measure that target, k known positives and u unknown labels among n candidates give the widest logical interval $[k/n, (k+u)/n]$. This interval does not bound net P&L when the missing outcomes lack a verified finite range, and it does not justify treating unknown labels as either wins or losses. A positive-result proportion and a monetary error budget are different estimands and require different evidence.

B Prospective evaluation contract

B.1 Scenario and version keys

A future D3 dataset should make each row reproducible from a key containing route, contract specification hash, common quantity, quote and settlement basis, fee basis, holding view, endpoint, input version and decision timestamp. It should also retain the code and configuration hashes, complete candidate index, stage-specific rejection reasons, original rate semantics, event identifiers, four book or fill records, event valuation prices, conversion evidence, residual quantity and evidence labels. Mutating any of these fields creates a new scenario rather than a silent repair of an old one.

The primary strategy, candidate endpoints and trial ledger are frozen before D3. Counterfactual endpoints are stored even when not selected, but they are not counted as additional filled positions. Strategy comparisons use the same outer session and capital constraints. A short-window strategy and a horizon-selection strategy can have different exit times only when the comparison follows their complete capital paths rather than pairing an ex post best window against a single fixed window.

B.2 Isolation, purging and dependence

D0 contains all observations already used to discover defects or write formulas. D1 estimates only necessary prediction, execution or missingness components; if none is fitted, it records “no trained estimator” rather than inventing a training stage. D2 selects among a finite registered set and records failures. D3 begins only after code, schema, parameters, metrics and stopping rules have an independently verifiable timestamp before the first label is observable.

Purging removes any training or selection row whose label interval intersects the evaluation interval. Publication delay and feature lookback receive their own cutoffs. Block length covers at least the longest holding interval and is justified from calibration dependence. Asset families, official event keys and shared books remain intact in resampling. A paired strategy difference is resampled as one unit so that common shocks are not artificially removed.

B.3 Required reporting funnel

The main table begins with the complete candidate denominator and reports comparable contracts, available schedules, eligible events, successful depth, selected actions, execution attempts, both-leg fills, complete exits and complete official labels. Primary states are mutually exclusive, while secondary reason flags can coexist. An unknown result is not zero and does not disappear from coverage. Comparisons additionally report the common-scorable intersection and each method's unique or missing rows, without calling the intersection the full market.

The minimum outcomes are complete-outcome/selected/all-candidate counts; gross funding, price and fees; S_0 error and components; selected-subset calibration; peak simultaneous reserved capital; and concentration. If predictive components exist, MAE, RMSE, post-selection bias, coverage and interval width accompany them. Complete economic paths report net value per reserved capital-hour, tail loss, drawdown and capital concentration only after capital constraints are measured. Latency is reported from source to receipt, computation, publication and user without rewriting source clocks.

Table 10: Analysis trial ledger. No specification was removed because its result was unfavorable. These retrospective analyses are not represented as an externally preregistered experiment.

Family	Specifications	Count
point estimator	naive last; expanding mean; rolling24 median	3
rolling window sensitivity	6;12;24;48;168 hours	5
one sided levels	0.80;0.90;0.95;0.99	4
DM absolute loss	2 comparisons x 6 series	12
DM squared loss	2 comparisons x 6 series	12
NW lag sensitivity	0;6;23;47 hours	4
subperiods	two holdout halves split 2026-10-22 UTC	2

C Additional result tables

This appendix reports the coverage-level and dependence-lag grids in compact form. The reproduction package additionally contains unrounded CSV files for every series, model, window, loss function, test and subperiod. No result is selected on the basis of statistical significance.

D Data dictionary and field-level provenance

The normalized files preserve public source fields rather than replacing them with a supposedly universal schema. Rates are fractions per event unless explicitly converted to

Table 11: One-sided holdout coverage under alternative calibration levels. Each column uses only calibration errors and its corresponding order statistic; these are not simultaneous or selection-conditional guarantees.

Venue	Asset	80% level	90% level	95% level	99% level
Hyperliquid	BTC	81.61%	90.16%	96.45%	99.52%
Hyperliquid	ETH	68.39%	70.32%	85.16%	97.42%
Hyperliquid	SOL	77.26%	90.65%	95.81%	98.87%
Extended	BTC	89.03%	93.23%	96.45%	99.35%
Extended	ETH	77.42%	88.39%	93.23%	98.23%
Extended	SOL	92.58%	96.94%	99.52%	100.00%

Table 12: Raw Diebold–Mariano p-values for previous-settled minus 24-hour-median absolute loss under alternative Newey–West lags. The final column reports the DM statistic at the main 23-lag choice; multiplicity-adjusted main results appear in Table 4.

Venue	Asset	Lag 0	Lag 6	Lag 23	Lag 47	DM at lag 23
Hyperliquid	BTC	1.5e-10	1.4e-05	8.5e-04	0.004	-3.337
Hyperliquid	ETH	2.4e-17	5.4e-06	6.0e-04	0.003	-3.433
Hyperliquid	SOL	2.2e-05	7.5e-04	0.002	0.006	-3.130
Extended	BTC	7.0e-10	0.004	0.056	0.099	-1.914
Extended	ETH	4.6e-08	0.002	0.017	0.034	-2.384
Extended	SOL	3.8e-04	0.024	0.070	0.100	-1.809

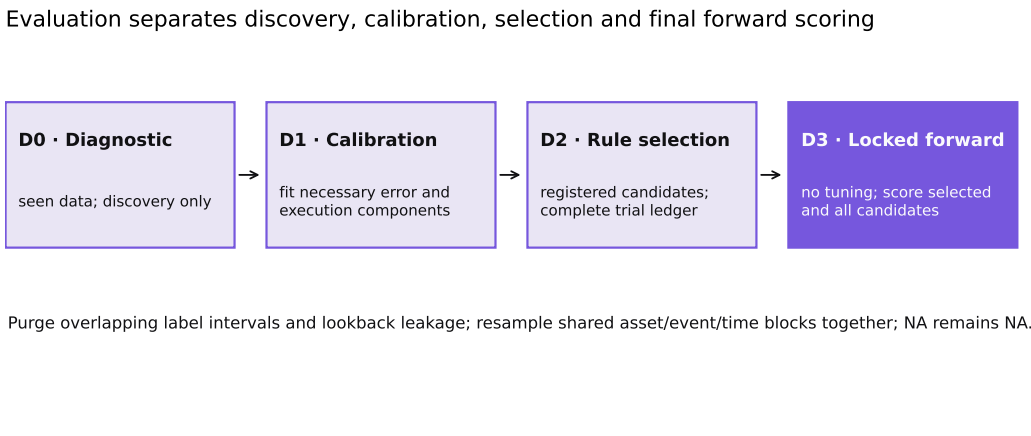


Figure 8: Prospective separation of diagnostic, calibration, rule-selection and locked-forward data. D3 is a protocol requirement rather than a claim completed by the present retrospective sample or single forward snapshot.

basis points for display. Monetary scenario values use the stated parity reporting basis and are not account cash. Blank or NA entries retain their evidence status.

Table 13: Data dictionary. The accompanying CSV additionally identifies the file and missing-data rule for every field.

Field	Unit	Meaning
venue	text	Public venue identifier Hyperliquid or Extended
asset	text	Base asset BTC ETH or SOL
timestamp_ms	milliseconds UTC	Original public record timestamp
timestamp_utc	ISO-8601 UTC	Readable form of original public timestamp
nominal_hour_utc	ISO-8601 UTC	Hour used to align events
funding_rate	fraction per hourly event	Public applied funding rate
raw_path	relative path	Raw public response supporting the row
period	text	Calibration before 2026-10-09 or holdout thereafter
decision_input_cutoff_utc	ISO-8601 UTC	Public timestamp of the previous event used as input
actual_rate	fraction	Rate recorded at target event
naive_last	fraction	Previous settled rate forecast
expanding_mean	fraction	Mean of all rates before target
rolling24_median	fraction	Median of previous 24 rates
mae_rate	fraction	Mean absolute forecast error
rmse_rate	fraction	Root mean squared forecast error
mae_currency_units_fixed_10000	USD-denominated scenario units	MAE multiplied by fixed 10000 scenario notional
scenario_funding_spread	USD-denominated scenario units	Previous-rate signed spread at fixed notional
realized_funding_spread_proxy	USD-denominated scenario units	Target-rate signed spread at fixed notional
D_funding_spread	USD-denominated scenario units	Scenario minus subsequent spread proxy
selected	boolean	Maximum scenario among two opposite directions in asset-hour
target_settlement_utc	ISO-8601 UTC	Target hourly public funding event
long_venue	text	Venue with lower current rate in forward scenario
short_venue	text	Venue with higher current rate in forward scenario
quantity_base	base asset units	Common quantity on entry and exit on both legs
common_quantity_step	base asset units	Coarse common public quantity grid
long_current_rate	fraction	Current published long-leg rate at t0
short_current_rate	fraction	Current published short-leg rate at t0
scenario_funding_price	public price units per base	Current oracle or mark used for C0
basis0	USD-denominated scenario units	Current same-book basis term
crossing0	USD-denominated scenario units	Four-fill current crossing term
funding0	USD-denominated scenario units	Current signed funding scenario C0
fees0	USD-denominated scenario units	Four-action public base-taker fee scenario
scenario_S0	USD-denominated scenario units	C0+B0-X0-fees0-O0
actual_long_rate	fraction	Public target-event long rate
actual_short_rate	fraction	Public target-event short rate
actual_funding_price_proxy	public price units per base	Later oracle or mark snapshot
basis_actual	USD-denominated scenario units	Entry/exit mid-basis term
crossing_actual	USD-denominated scenario units	Entry/exit four-fill crossing term
funding_actual_proxy	USD-denominated scenario units	Signed target funding proxy

Field	Unit	Meaning
fees_actual_scenario	USD-denominated scenario units	Entry/exit public base-taker fee scenario
profit_public_paper_proxy	USD-denominated scenario units	B-X+C-fees under parity scenario
error_D	USD-denominated scenario units	S0 minus subsequent paper proxy
D_funding_rate	USD-denominated scenario units	Funding-rate contribution to D
D_funding_mark	USD-denominated scenario units	Funding-price main-effect contribution to D
D_funding_rate_mark_cross	USD-denominated scenario units	Rate-price interaction contribution to D
D_basis	USD-denominated scenario units	B0-B contribution to D
D_crossing	USD-denominated scenario units	X-X0 contribution to D
D_fee	USD-denominated scenario units	Fee-fee0 contribution to D
identity_residual	USD-denominated scenario units	D minus displayed component sum
status	text	Evidence completeness state
reporting_currency_basis	text	One-for-one USDC/USD parity reporting scenario

Raw-evidence manifest

The raw manifest records request phase, venue, asset, public URL, request JSON, retrieval time, HTTP status, elapsed time, response path and SHA-256. Historical files retain their original collection hash; forward files receive hashes at collection. The final verification script checks every included response against its manifest and confirms that offline recomputation reproduces the generated outputs.

Minimum runtime and acquisition

Offline recomputation uses Python 3.12 with pandas, NumPy, SciPy and Matplotlib. It requires no credentials and sends no orders. Online refresh is optional, produces a different snapshot and must use the same public no-key endpoints and semantic checks. Full execution normally takes seconds for the core history analysis and under a minute for figure generation; network collection time depends on public endpoints.