HUMARB RESEARCH 06

Forecasting Next-Settlement Funding from the Premium Integral:

Mechanism-Based Ornstein–Uhlenbeck Evidence across Three Venues

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

Methodological and empirical preprint ů November 2026 ů Not peer reviewed

DOI: pending Zenodo publication

Abstract

Perpetual-futures funding is a deterministic transformation of a kernel-weighted premium integral, yet the integral is rarely used as the state variable in next-settlement forecasts. We derive exact discrete moments and clamped-normal predictive distributions under an Ornstein–Uhlenbeck premium model, then evaluate genuinely out-of-sample forecasts for BTC, ETH, and SOL on Binance, Gate, and Hyperliquid. The 38-day public-data sample contains 2,781 targets over 31 evaluation days. OU reduces pooled RMSE from 0.2368 bp for the last-settlement benchmark to 0.1968 bp, but remains slightly behind EWMA at 0.1947 bp. Improvements over Last are strongest on Hyperliquid; Binance and Gate favor EWMA over OU. Minute-level reconstruction supports linear rather than equal kernels on Binance and Gate. Twenty configurations and Holm/Reality-Check procedures address selection. Gaussian intervals fail by venue: 95% coverage is 39.1%, 43.5%, and 97.7%, respectively. The evidence supports mechanism-aware point forecasting, not uniformly superior accuracy, calibrated uncertainty, or executable net returns.

Keywords: perpetual futures; funding rates; premium integral; Ornstein–Uhlenbeck process; probabilistic forecasting; forecast evaluation

JEL classification: G12; G13; G14; C58

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

Contents

1 Introduction **1**

2 Related literature **3**

2.1 Perpetual claims and funding as contract design 3

2.2 Short-horizon funding predictability 3

2.3 Forecast evaluation and data snooping 4

3 Institutional background **5**

3.1 From premium observations to a settlement 5

3.2 Venue kernels and sampling approximation 5

3.3 Caps, dynamic intervals, and contract scope 6

3.4 Implications for forecasting 6

4 Data **7**

4.1 Sources and snapshot 7

4.2 Coverage and missing observations 7

4.3 Descriptive statistics 8

4.4 Forecast-origin integrity and bias controls 8

5 Methods **9**

5.1 Discrete OU state model 9

5.2 Exact next-interval moments 9

5.3 Mean, quantiles, and mixed predictive law after clamping 10

5.4 Benchmarks and walk-forward protocol 11

5.5 Loss comparisons and calibration tests 11

5.6 Model-search control and economic scale 12

6 Results **12**

6.1 The documented kernel reconstructs settlements more closely 12

6.2 OU improves on Last but not uniformly on EWMA 14

6.3 Predictive distributions succeed hourly and fail at minute-linear venues 15

6.4 Significance is selective after multiplicity control 16

7 Robustness and economic interpretation **17**

7.1 Estimation windows 17

7.2 Shrinkage and alternative persistence estimation 18

7.3 Forced kernel misspecification 18

7.4 Venue and subperiod stability 19

7.5 Reality Check and the searched family 19
7.6 Economic scale without fictitious P&L 19

8 Limitations **20**

9 Conclusion **22**

Declarations **22**

A Notation and computational definitions **25**

B Additional derivations **26**
B.1 Exact variance recursion 26
B.2 Predictive CDF and probability masses 26
B.3 Expected clamped normal with outer caps 27
B.4 Proper-score conventions 27

C Reproducible forecast protocol **27**

D Complete empirical tables **28**
D.1 Asset-level forecast loss and distribution scores 28
D.2 Complete Diebold–Mariano family 29
D.3 Complete robustness family 30

E Public prediction-table dictionary **30**

F Reproduction checks and evidence map **31**

1 Introduction

Perpetual-futures funding is often treated as an observed scalar to be extrapolated. That treatment discards useful structure. On the venues studied here, the settled rate is computed from a sequence of premium observations, aggregated with a venue-specific kernel and passed through an interest-rate adjustment, a dead zone, and a cap. The next payment therefore depends not only on the previous settlement but also on the current premium state, the persistence of that state, and the amount of averaging that remains before settlement. A forecasting model can either ignore this accounting identity or make it the center of the statistical design. We choose the latter.

The research question is narrow: *does a mechanism-based model of the premium integral improve genuine next-settlement forecasts, and under what venue conditions are its predictive distributions reliable?* The question has two parts that must not be conflated. A model can reduce mean squared point-forecast error while producing badly calibrated intervals. It can also win against one naive benchmark and lose against another. Accordingly, our estimand is not a generic claim of predictability. It is a set of out-of-sample loss comparisons against three transparent baselines, accompanied by proper scores and interval diagnostics.

We model the premium as a discretely observed Ornstein–Uhlenbeck (OU) process. The continuous-time motivation is mean reversion; the implementation is an exact discrete autoregression of order one. At each actual settlement origin, we estimate the model using only premiums already available, shrink logit persistence across the three assets within a venue, and integrate the predictive distribution over the next non-overlapping settlement interval. The kernel enters both the conditional mean and variance. The funding transformation then maps the predictive normal integral into a mixed distribution with continuous regions and point masses at the interest rate and, when relevant, the caps. This yields point forecasts, quantiles, randomized probability integral transforms (PIT), continuous ranked probability scores (CRPS), logarithmic scores, and interval scores from a single model rather than from unrelated post hoc calculations.

The empirical design uses public Binance, Gate, and Hyperliquid data for BTC, ETH, and SOL from 24 September through 31 October 2026. The first week supports estimation; the evaluation begins on 1 October and covers 31 days. There are 2,781 next-settlement targets across nine venue–asset series. Binance and Gate provide 92 eight-hour forecasts per asset, while Hyperliquid provides 743 hourly forecasts per asset. Every origin is a real settlement time and every target is the next actual settlement. The rolling window, shrinkage rule, bias correction, and kernel are stated explicitly as research examples; they are not represented as the settings of any production system.

Four findings organize the paper. First, the mechanism is empirically meaningful before any forecasting comparison. Reconstructing settled rates from minute premiums with linearly increasing weights produces RMSE no greater than 0.0588 bp across the six Binance and Gate series. Forcing equal weights produces RMSE of at least 0.1145 bp. This is not an exact reconstruction of native five-second Binance sampling, and the premium inputs and archives remain approximate. It is nevertheless direct evidence that kernel choice changes the accounting result in the direction implied by the public rules.

Second, OU point forecasts improve on the last-settlement benchmark in pooled loss. Pooled RMSE is 0.1968 bp for OU, compared with 0.2368 bp for Last, 0.1947 bp for an exponentially

weighted moving average (EWMA), and 0.3649 bp for the interest-rate forecast. The ranking changes by venue. OU RMSE is 0.4747 bp on Binance and 0.3866 bp on Gate, versus 0.4638 and 0.3703 bp for EWMA. On Hyperliquid, OU achieves 0.0437 bp, compared with 0.0604 for EWMA and 0.0611 for Last. Thus, the premium-integral model is competitive but not uniformly superior. Its advantage over Last is clearest in the hourly, equal-weight setting; the two minute-linear venues slightly favor EWMA.

Third, statistical significance is more selective than the pooled point estimate. We conduct 27 venue–asset Diebold–Mariano comparisons with Newey–West covariance estimates and the Harvey–Leybourne–Newbold small-sample correction. Holm adjustment leaves 11 rejections at the 5% level across all baselines. Against Last, Binance SOL and Hyperliquid BTC and SOL survive the correction; the other six comparisons do not. A White Reality Check across all 20 configurations rejects the null that no OU configuration improves on Last, with a fixed-length circular-block bootstrap p-value of 0.00050. That rejection concerns the best of a searched family and does not rescue any individual venue–asset comparison that fails its own multiplicity-adjusted test.

Fourth, distributional adequacy is strongly venue dependent. The nominal 95% Gaussian interval covers only 39.1% of Binance targets and 43.5% of Gate targets, but 97.7% of Hyperliquid targets. The corresponding 50% coverage rates are 16.3%, 18.8%, and 80.6%. Pooling these regimes would obscure the diagnosis: pooled 80% coverage happens to be 80.5%, yet per-series tests reject violation independence in several cases and the 50% and 95% levels are badly misaligned. The Gaussian innovation and constant-variance assumptions are therefore unsuitable as a universal description of funding uncertainty, even where point forecasts are useful.

The contributions are consequently specific. We provide a fully discrete forecast construction that respects the premium kernel and the clamp, including exact integrated AR(1) moments and a mixed predictive law. We validate the kernel against public settlements before evaluating forecasts. We report point, interval, and distributional evidence separately, with explicit multiplicity control and all attempted configurations. Finally, we connect forecast errors to economic scale without converting them into fictitious trading profits. An illustrative complete round-trip cost of 33.01 bp, or \$33.01 per \$10,000 of notional, is orders of magnitude larger than the per-interval forecast errors reported in basis points. No order-book execution, borrowing, transfer, or position-level return backtest is claimed.

This paper extends the measurement agenda of humarb Research 01 and 02 (humarb Research, 2026; Liu, 2026). Those papers distinguish quoted carry from executable net carry and frame settled history, depth-walked costs, and liquidation constraints. Here we move one step earlier in the decision sequence: given public mechanism rules and current premium information, what can be said about the next settlement itself? The answer informs expected funding inputs, but does not replace cost, margin, or execution analysis.

The remainder proceeds as follows. Section 2 locates the study relative to perpetual-contract and forecast-evaluation research. Section 3 formalizes venue mechanisms. Section 4 describes acquisition and sample integrity. Section 5 derives the forecasts and tests. Section 6 presents the main evidence, Section 7 the robustness and economic interpretation, and Section 8 the limitations. Section 9 concludes.

2 Related literature

2.1 Perpetual claims and funding as contract design

Perpetual futures remove the fixed expiry of a conventional futures contract and therefore require a mechanism that repeatedly anchors the contract price to a reference price. The conceptual lineage goes at least to Shiller (1993), who studies cash settlement for perpetual claims. Modern cryptocurrency implementations use periodic peer-to-peer funding transfers: when funding is positive, longs pay shorts; when it is negative, the direction reverses. The payment is not an exchange fee, and its sign and magnitude are determined by contract-specific rules rather than by a universal law of one price.

Recent asset-pricing work treats perpetual futures as instruments whose valuation and arbitrage restrictions depend on these rules. Ackerer et al. (2025) develop a pricing framework for perpetual futures, while Dai et al. (2025) study arbitrage in perpetual contracts and the role of contractual bounds. This literature warns against treating the funding rate as an exogenous dividend appended to an otherwise conventional futures price. The timing, premium definition, interest component, clamps, and settlement conventions are part of the instrument. Our approach is complementary: rather than deriving a no-arbitrage price, we take the public settlement rule as the measurement equation and ask whether its state variable supports short-horizon forecasting.

This distinction also separates the paper from carry measurement. humarb Research 01 measures executable net carry using settled funding histories and cost-adjusted route comparisons (humarb Research, 2026). humarb Research 02 provides a broader accounting frame for settled history, asynchronous settlement grids, execution depth, and liquidation distance (Liu, 2026). Neither attempts an out-of-sample forecast of the next settlement. The present paper supplies a candidate forecast input, but its forecast error cannot be interpreted as net carry without those additional layers.

2.2 Short-horizon funding predictability

The closest empirical predecessor is Inan (2026), who studies one-step-ahead funding predictability for Bitcoin perpetuals on Binance and Bybit using double autoregressive models. That work reports improvements over a no-change benchmark and emphasizes time variation in predictability. We share the commitment to genuine out-of-sample evaluation and transparent benchmarks, but alter the state variable and the institutional comparison. Our model forecasts the future kernel-weighted premium integral rather than applying a general time-series model directly to the settled funding sequence. It also compares three venues and three assets, evaluates the funding clamp explicitly, and treats probabilistic calibration as a separate requirement.

The difference matters because previous-settlement funding is a compressed statistic. It mixes the premium path, the kernel, and the transformation into one number. Two settlement observations can be identical while current premium conditions differ sharply; conversely, a large change in the current premium may have little effect after eight hours of averaging. A model based only on past settlements cannot distinguish these states. The premium integral provides a structural explanation for why persistence at the settlement frequency should vary across venues and through time.

At the same time, structural motivation does not guarantee statistical dominance. EWMA is a strong short-horizon competitor because settled funding itself is persistent and because the clamp creates many observations near the interest rate. OU imposes a single innovation variance and a linear conditional mean. If actual premiums exhibit stochastic volatility, jumps, asymmetric mean reversion, or regime shifts, a more flexible reduced-form benchmark can win on point loss even with less institutional detail. We therefore treat Inan's predictability question as open at the level of model class and venue, not as a hypothesis that must be confirmed for our mechanism to be useful.

2.3 Forecast evaluation and data snooping

Point forecasts are evaluated using MAE and RMSE, with loss differences tested by the Diebold–Mariano procedure (Diebold and Mariano, 1995). Because forecast errors are serially dependent and hourly or eight-hour samples can be short, we use Newey–West covariance estimates (Newey and West, 1987) and the finite-sample modification of Harvey et al. (1997). The testing unit remains an individual venue–asset series; pooling all targets into one very large sample would overstate precision by hiding cross-sectional and institutional heterogeneity.

Proper scoring rules provide a broader assessment. CRPS evaluates the entire predictive distribution, the logarithmic score penalizes density or probability mass assigned to the realized outcome, and the interval score combines width with penalties for misses (Gneiting and Raftery, 2007). Coverage alone is insufficient: a wide interval can achieve nominal coverage without being informative, and a sequence can have correct unconditional coverage while violations cluster. We consequently report Kupiec's unconditional-coverage test and Christoffersen's independence test (Kupiec, 1995; Christoffersen, 1998), together with interval scores. Randomized PIT values accommodate the point masses created by the clamp and provide a distributional diagnostic in the tradition of density-forecast evaluation (Berkowitz, 2001).

Multiple model configurations create a selection problem even when every individual forecast is out of sample. We report all 20 configurations, adjust the family of Diebold–Mariano p-values using Holm (1979), and implement White's Reality Check for the best configuration relative to Last (White, 2000). The bootstrap uses fixed-length circular blocks of three days. These procedures respond to different questions. Holm controls family-wise error across the reported pairwise tests; the Reality Check asks whether the best searched configuration beats its benchmark after accounting for the search. Neither establishes stability outside the sample, and neither transforms a backtest into a trading result.

Finally, small-sample persistence estimation introduces its own bias. We include the Kendall correction as an alternative estimator (Kendall, 1954), not as a device for selecting the most favorable result. The central methodological stance is that institutional structure should make a forecast more interpretable and testable, while evaluation standards determine whether it is more accurate.

3 Institutional background

3.1 From premium observations to a settlement

Let $p_{t,j}$ denote the premium index observed at step j of the settlement interval beginning after the settlement at t. An interval contains n equally spaced modeling steps, and the normalized weights satisfy $w_j \geq 0$ and $\sum_{j=1}^{n} w_j = 1$. The premium integral is

$$A_{t,n} = \sum_{j=1}^{n} w_j p_{t,j}. \tag{1}$$

The term "integral" refers to this weighted sum on the discrete observation grid. It does not imply that the exchange publishes or stores the same grid used by the researcher.

The funding transformation on an eight-hour reference frame is

$$q(a) = \mathrm{clip}[a + \mathrm{clip}(I_8 - a, -\delta, \delta), -C, C]\,, \tag{2}$$

where I_8 is the reference interest component, δ is the adjustment band, and C is a cap. For the crypto contracts and sample considered here, the working values are $I_8 = 0.0001$ and $\delta = 0.0005$. We use $C = 0.02$ as a nonbinding modeling cap for the main sample and discuss the distinction from contract-specific limits below. The inner clamp has a simple consequence: if $a \in [I_8 - \delta, I_8 + \delta]$, then $q(a) = I_8$. Outside that dead zone, $q(a) = a - \delta$ below it and $q(a) = a + \delta$ above it, before the outer cap. The mapping is monotone, continuous, and flat over the dead zone. It therefore transforms a continuous premium distribution into a distribution with a point mass at I_8.

For a settlement interval of N hours on Binance, Gate, or Hyperliquid, the relevant scaling in our sample is $s = N/8$, applied after the reference-frame transformation. The settled rate is modeled as

$$F_{t+1} = s\, q(A_{t,n}). \tag{3}$$

Hyperliquid settles hourly, so $s = 1/8$. The public documentation states that the formula is computed on an eight-hour basis and paid every hour at one eighth of the computed rate (Hyperliquid, 2026). Applying s to the interest term before evaluating q would scale it twice; our implementation applies the scaling once to the whole transformed quantity.

3.2 Venue kernels and sampling approximation

Binance's public funding documentation describes linearly increasing weights for intervals longer than one hour and native premium sampling at five-second frequency (Binance, 2026). Gate calculates once per minute and documents weights $1, 2, \ldots, n$ over the interval (Gate, 2026). For the eight-hour minute approximation used in this paper,

$$w_j^{\mathrm{lin}} = \frac{j}{n(n+1)/2}, \qquad n = 480. \tag{4}$$

Later observations consequently receive more weight. This is not equivalent to an arithmetic mean: a premium shock near the end of the interval has a larger effect on the imminent settlement than the same shock near its beginning.

Hyperliquid instead documents five-second premium sampling averaged over the hour (Hyperliquid, 2026). Our hourly normalized series represents that hour with one equal-weight observation, so $w_j^{\mathrm{eq}} = 1/n$ and $n = 1$ for each next-hour forecast. In robustness exercises, deliberately imposing linear or equal weights across all venues serves as a misspecification check, not as an alternative claim about official rules.

The minute approximation to Binance is coarser than the native five-second grid. If premiums were constant within each minute, the exact sum of native weights within minute j would be quadratic in j, not exactly proportional to j. Our reconstruction errors therefore measure the combined effect of archive premiums, missing minutes, sampling alignment, and the kernel approximation. They should not be described as a bit-for-bit audit of exchange settlement engines. Gate's native minute grid is closer to the model grid, and its reconstruction errors are correspondingly smaller.

3.3 Caps, dynamic intervals, and contract scope

Actual funding caps vary by venue and contract and can be linked to margin parameters or explicit funding limits. Gate's documentation, for example, exposes contract-specific limits and describes conditions under which settlement frequency can change (Gate, 2026). Hyperliquid documents a comparatively wide hourly cap (Hyperliquid, 2026). A universal $C = 0.02$ is suitable for the sample only because the outer cap does not bind in the observations used for the reported forecasts. It is not proposed as a universal institutional constant.

Dynamic settlement intervals create another boundary. The forecasting algorithm uses the next observed interval length and requires the next settlement to occur at the expected time. This prevents accidental alignment to a later settlement when a missing observation appears. It does not forecast a future regime change in interval length. If a venue changes the interval after the origin, a real-time forecaster would need contract metadata available at that origin and a rule for handling the change. Our sample does not identify such a rule.

The analysis is limited to ordinary USDT- or USD-linked crypto perpetuals for BTC, ETH, and SOL. It does not cover traditional-finance contracts with different interest components, special contracts with distinct premium formulas, or deployer-defined perpetuals with alternative multipliers. Contract eligibility must therefore be checked before extending the model. Public mechanism similarity is not enough to establish numerical equivalence.

3.4 Implications for forecasting

Three institutional features shape the statistical problem. First, averaging attenuates the effect of the current premium. In a persistent process, the conditional mean of $A_{t,n}$ is a weighted sum of discounted future states, not simply the current premium. Second, the kernel changes the variance as well as the mean: innovations at different times enter the integral with different cumulative weights. Third, the dead zone compresses a range of premium integrals into one funding value. This induces mass at the interest rate and makes ordinary continuous-density diagnostics invalid unless randomized at that mass.

These features also explain why venue comparisons should not be reduced to one pooled error number. Hourly equal-weight aggregation and eight-hour linear aggregation produce different forecast horizons, different amounts of averaging, and different masses at the interest

rate. A model can be well suited to one and poorly suited to another. The empirical section therefore treats venue as a primary dimension rather than as a robustness afterthought.

4 Data

4.1 Sources and snapshot

The empirical sample uses only public market data available without credentials. Binance minute premium-index series come from the public USD–M futures data archive, and settled funding rates come from the corresponding public funding-rate archive. Gate minute premium indexes and funding settlements are obtained from its public futures endpoints. Hyperliquid hourly premium and funding histories are obtained from its public information endpoint. Acquisition was completed on 3 November 2026. The reproduction package records every request URL or payload, access timestamp, response size, and content hash in a source manifest; unavailable responses are recorded rather than replaced.

The sample begins at 00:00 UTC on 24 September 2026 and ends at 00:00 UTC on 1 November 2026, with the latter timestamp exclusive. It therefore covers 38 calendar days and the full calendar month of October. Assets are BTC, ETH, and SOL, and venues are Binance, Gate, and Hyperliquid. The nine series are selected before forecast evaluation to represent large and smaller crypto assets on two centralized venues with eight-hour settlements and one venue with hourly settlement. No series is removed because its forecast result is unfavorable.

The normalized premium files contain a UTC timestamp in milliseconds and a premium expressed as a decimal fraction. Funding files contain the actual settlement timestamp, interval length in hours, and settled funding rate as a decimal fraction. One basis point (bp) equals 0.0001 in rate units, so a funding rate of 0.0001 is 1 bp. Duplicate timestamps are removed before analysis, timestamps are sorted, and numeric fields are parsed explicitly. No premium or funding observation is winsorized, smoothed, or trimmed on the basis of its magnitude.

4.2 Coverage and missing observations

Table 1 reports the data audit. Each Binance series contains 51,840 minute premiums out of 54,720 expected grid points, or 94.74%. This corresponds to two unavailable daily archives per asset. The missing days remain missing; no synthetic premiums are inserted. Each Gate series contains 54,666 of 54,720 minutes, or 99.90%, a shortfall of 54 minutes. Hyperliquid contains all 912 expected hourly observations per asset. Funding histories contain 114 settlements per Binance or Gate asset and 912 per Hyperliquid asset.

The estimation sample prior to evaluation supplies at least seven calendar days of history. Out-of-sample origins begin on 1 October, and the last targets settle on 31 October. There are 92 targets per asset on each eight-hour venue and 743 per asset on Hyperliquid, for 2,781 targets in total. The difference from a naive count of 93 eight-hour targets reflects the requirement that an origin have both a current settlement and a valid next non-overlapping settlement within the sample. Hourly targets similarly exclude origins without a complete forecast–target pair.

Table 1. Sample coverage and out-of-sample target counts

Venue	Asset	Premium rows	Expected	Coverage (%)	Funding rows	OOS targets
Binance	BTC	51,840	54,720	94.74	114	92
Binance	ETH	51,840	54,720	94.74	114	92
Binance	SOL	51,840	54,720	94.74	114	92
Gate	BTC	54,666	54,720	99.90	114	92
Gate	ETH	54,666	54,720	99.90	114	92
Gate	SOL	54,666	54,720	99.90	114	92
Hyperliquid	BTC	912	912	100.00	912	743
Hyperliquid	ETH	912	912	100.00	912	743
Hyperliquid	SOL	912	912	100.00	912	743

Source: authors' calculations from public venue data, 24 September–31 October 2026. OOS denotes out of sample. Expected premiums use a full minute or hourly grid.

4.3 Descriptive statistics

The sample is characterized by negative average premiums and frequent funding near the interest component. Across assets, mean minute premiums are −4.591 bp on Binance and −4.691 bp on Gate; their standard deviations are 1.487 and 1.742 bp. Hyperliquid's hourly premium mean is −3.130 bp with a standard deviation of 1.834 bp. Extreme minute premiums are much larger than the average, especially on the centralized venues, indicating non-Gaussian tail behavior even before any forecast model is fitted.

Mean settled funding is 0.366 bp on Binance, 0.272 bp on Gate, and 0.090 bp on Hyperliquid; the respective standard deviations are 0.456, 0.398, and 0.068 bp. The funding rate equals the applicable interest component in 10.82% of Binance settlements, 4.97% of Gate settlements, and 61.99% of Hyperliquid settlements. These frequencies are descriptive of the full sample, not forecast probabilities. The large Hyperliquid mass explains why the dead zone is central to both point and distributional evaluation.

4.4 Forecast-origin integrity and bias controls

At each origin, model fitting uses premiums strictly before that origin. The current premium state uses the last observation available at the origin; Hyperliquid's current hourly observation is aligned to the public settlement timestamp. The target is the first subsequent actual funding settlement, and its timestamp must equal the origin plus the stated interval within a two-minute tolerance for settlement processing. If this alignment fails, the pair is skipped and the exclusion remains visible in target counts. There is no use of the target premium path, target funding rate, future interval metadata beyond the next observed settlement alignment, or future model residuals.

The primary rolling estimation window is 168 hours. Shorter and longer windows are examined as robustness cases. Shrinkage at an origin is computed across assets within the same venue using only their current rolling fits. EWMA and Last baselines use previous funding settlements only. The Interest baseline uses the applicable reference interest component and no estimated market state.

The three assets remained available throughout the sample, so there are no delisted contracts to remove or retain. This avoids outcome-based deletion but does not establish freedom

from survivorship bias in a broader contract universe. The results are conditional on nine liquid, continuously observable series and a short period. Likewise, public archives may later be revised; the source manifest identifies the snapshot used here, but cannot prove that it equals every historical real-time display. These qualifications motivate the reproduction files and limit cross-period interpretation.

5 Methods

5.1 Discrete OU state model

The premium state is motivated by the Ornstein–Uhlenbeck stochastic differential equation

$$dp_t = \kappa(\theta - p_t)\,dt + \sigma\,dW_t, \qquad \kappa > 0, \tag{5}$$

where θ is the long-run premium mean, κ is the mean-reversion rate, and W_t is Brownian motion. On a unit observation grid, exact discretization gives

$$p_{j+1} = \theta + \phi(p_j - \theta) + \varepsilon_{j+1}, \qquad \phi = e^{-\kappa}, \qquad \varepsilon_{j+1} \sim N(0, \sigma_\varepsilon^2), \tag{6}$$

with $\sigma_\varepsilon^2 = \sigma^2(1-\phi^2)/(2\kappa)$ under the continuous-time model. The empirical implementation uses equation (6) directly. No continuous-time parameter is needed to compute the forecasts.

At origin t, ordinary least squares regresses p_{j+1} on an intercept and p_j over the trailing window. Writing the fitted intercept as $\widehat{a}$ and slope as $\widehat{\phi}_0$, the long-run mean is $\widehat{\theta} = \widehat{a}/(1-\widehat{\phi}_0)$ when well defined and the residual variance is estimated with denominator $n-2$. For numerical stability, the persistence used in forecasting is restricted to $(0, 0.999999)$, and an implausibly large intercept ratio is replaced by the window mean. The primary specification uses the raw slope. The alternative Kendall correction is $\widehat{\phi}_K = \widehat{\phi}_0 + (1+3\widehat{\phi}_0)/n$ (Kendall, 1954). These are estimation choices for a stationary short-horizon approximation, not assertions that premiums are stationary at all horizons.

Within each venue and origin, persistence can be shrunk across BTC, ETH, and SOL on the logit scale. Let $\eta_i = \log\{\phi_i/(1-\phi_i)\}$ and let s_i be its approximate standard error from the AR regression. The inverse-variance target is $\bar{\eta} = \sum_i s_i^{-2}\eta_i / \sum_i s_i^{-2}$. The empirical-Bayes between-asset variance is

$$\widehat{\tau}^2 = \max\{\mathrm{Var}(\eta_1, \ldots, \eta_m) - m^{-1}\textstyle\sum_i s_i^2, 0\}. \tag{7}$$

The shrunk value is $\widetilde{\eta}_i = \omega_i\eta_i + (1-\omega_i)\bar{\eta}$, with $\omega_i = \widehat{\tau}^2/(\widehat{\tau}^2 + s_i^2)$, and $\widetilde{\phi}_i = \mathrm{logit}^{-1}(\widetilde{\eta}_i)$. If $\widehat{\tau}^2 = 0$, persistence is fully pooled. The alternatives are no shrinkage and complete pooling. Only persistence is shrunk; each asset retains its own mean and innovation variance. The primary model uses the 168-hour window, empirical-Bayes shrinkage, raw persistence, and native kernels. All numerical settings are disclosed research examples, not representations of any production configuration.

5.2 Exact next-interval moments

At the origin, the current premium p_0 is known. The future state at step j satisfies $E(p_j \mid p_0) = \theta + (p_0 - \theta)\phi^j$. Therefore the conditional mean of the premium integral in equation (1)

is

$$\mu_A = \theta + (p_0 - \theta) \sum_{j=1}^{n} w_j \phi^j. \tag{8}$$

This expression shows why the kernel and persistence must be considered jointly. High persistence gives the current state more influence; heavy late weights reduce the influence of observations that will soon be far in the past.

For the variance, define cumulative future coefficients by the backward recursion

$$H_n = w_n, \qquad H_j = w_j + \phi H_{j+1}, \quad j = n-1, \ldots, 1. \tag{9}$$

The innovation at step j affects the current state and all later states, entering the integral with coefficient H_j. Independence of AR innovations then gives

$$\sigma_A^2 = \sigma_\varepsilon^2 \sum_{j=1}^{n} H_j^2. \tag{10}$$

Equations (8)–(10) are exact under the fitted discrete AR(1), including for the one-step hourly case. They avoid a simulation grid and make every forecast deterministic given the data and seed-independent except for randomized PIT evaluation.

The conditional law before applying the funding transformation is $A \mid p_0 \sim N(\mu_A, \sigma_A^2)$. This law is deliberately simple. It does not include parameter uncertainty, stochastic volatility, jumps, missing-data uncertainty, or model-selection uncertainty. The calibration tests below are therefore tests of the complete approximate predictive distribution, not only of the kernel algebra.

5.3 Mean, quantiles, and mixed predictive law after clamping

The point forecast is $E[sq(A) \mid p_0]$, not $sq(\mu_A)$. The distinction matters because the dead zone and caps make q nonlinear. Let Φ and φ be the standard normal distribution and density, and define $z(x) = (x - \mu_A)/\sigma_A$. For a threshold x, the lower and upper first-loss terms are

$$L_-(x) = (\mu_A - x)\Phi(-z(x)) + \sigma_A \varphi(z(x)), \tag{11}$$

$$L_+(x) = (x - \mu_A)\Phi(z(x)) + \sigma_A \varphi(z(x)). \tag{12}$$

Ignoring a binding outer cap for display, the normal hinge identity gives

$$E[q(A) \mid p_0] = I_8 + L_-(I_8 + \delta) - L_+(I_8 - \delta). \tag{13}$$

The implemented expression additionally subtracts the two tail terms associated with $\pm C$. When σ_A is numerically zero, the expectation reduces to $q(\mu_A)$.

Quantile forecasts use monotonicity. For $0 < u < 1$,

$$Q_{sq(A)}(u) = s\, q\{\mu_A + \sigma_A \Phi^{-1}(u)\}. \tag{14}$$

This produces valid central intervals and preserves the point masses. CRPS is approximated

from 199 equally spaced probability levels from 0.005 to 0.995 using the quantile-loss identity

$$\mathrm{CRPS}(F, y) \approx \frac{2}{J} \sum_{\ell=1}^{J} (y - Q(u_\ell))\{u_\ell - \mathbf{1}(y < Q(u_\ell))\}. \tag{15}$$

We report CRPS in basis points. The approximation truncates the extreme half-percent tails; it is a numerical quadrature choice, not a claim that those tails cannot occur.

The logarithmic score uses the actual mixed law. At $F = sI_8$, predictive probability mass equals $\Phi(z(I_8 + \delta)) - \Phi(z(I_8 - \delta))$. At the caps, it equals the corresponding normal tail probability. Away from masses, the density is the normal density evaluated at the inverse shifted premium, divided by σ_A and by the settlement scale s. The score is $-\log f(F_{t+1})$, with a numerical probability floor of 10^{-300}. It is reported for diagnosis, not compared with the point-only baselines.

For PIT evaluation, let G be the predictive CDF after clamping. Away from point masses, the PIT is $G(y)$. At a mass, it is drawn uniformly between $G(y-)$ and $G(y)$. This randomization is necessary because a constant PIT at the interest rate would mechanically create a histogram spike and invalidate uniform diagnostics. The random seed is fixed for reproduction. Interval inclusion and point-mass comparisons use an absolute tolerance of 10^{-12} in rate units, equal to 10^{-8} bp, solely to prevent floating-point boundary artifacts.

5.4 Benchmarks and walk-forward protocol

Three baselines anchor the comparison. *Last* predicts the next funding rate by the settlement immediately before the origin. *EWMA* applies an exponentially weighted average to all previous settlements with a three-settlement half-life; its smoothing coefficient is $\alpha = 1 - 2^{-1/3}$. *Interest* predicts the applicable interest component, I_8 for eight-hour settlements and $I_8/8$ for hourly Hyperliquid settlements. Last and Interest require no fitted parameter. EWMA has one explicitly stated smoothing rule and no future tuning.

The protocol is rolling-origin and genuinely out of sample. At every actual settlement origin from 1 October onward, the model is re-estimated from available premiums, the next non-overlapping settlement is forecast, and only then is the realized target attached for evaluation. Forecasts are stored with origin, target, interval length, model configuration, point forecast, predictive mean and standard deviation, quantile intervals, PIT, and scores. A target is never used to fit its own predictor. Robustness configurations repeat the complete walk-forward rather than reusing residuals from the primary model.

5.5 Loss comparisons and calibration tests

For model m, error is $e_{m,t} = \widehat{F}_{m,t} - F_t$ and is reported in basis points. MAE and RMSE are $n^{-1} \sum_t |e_{m,t}|$ and $\{n^{-1} \sum_t e_{m,t}^2\}^{1/2}$. Directional accuracy is measured relative to the interest component, excluding realized rates exactly at that component. It is secondary because sign accuracy can be high even when magnitude forecasts are poor.

For each venue–asset series and baseline b, the loss differential is $d_t = e_{b,t}^2 - e_{\mathrm{OU},t}^2$, so a positive mean favors OU. The Diebold–Mariano statistic uses a Newey–West estimate of the long-run variance with Bartlett weights (Diebold and Mariano, 1995; Newey and West, 1987). The lag is three for eight-hour series and 23 for hourly series, approximately one calendar

day in each case. We apply the Harvey et al. (1997) correction and a Student-t reference distribution with $n-1$ degrees of freedom. The 27 raw p-values from nine series times three baselines are adjusted together using Holm's step-down procedure (Holm, 1979). Both raw and adjusted values are supplied.

Central interval coverage is $\widehat{c}_\alpha = n^{-1}\sum_t \mathbf{1}(L_{\alpha,t} \le F_t \le U_{\alpha,t})$ for nominal levels $\alpha \in \{0.50, 0.80, 0.95\}$. Kupiec's likelihood-ratio test compares the observed violation rate with $1-\alpha$ (Kupiec, 1995); Christoffersen's independence test compares transition-dependent and independent violation likelihoods (Christoffersen, 1998). The interval score is

$$IS_\alpha = (U-L) + \frac{2}{1-\alpha}(L-y)\mathbf{1}(y<L) + \frac{2}{1-\alpha}(y-U)\mathbf{1}(y>U), \tag{16}$$

reported in basis points. Christoffersen transitions are computed only within each venue–asset series. Venue and pooled rows report descriptive coverage and Kupiec values but no cross-asset independence statistic, because concatenating assets would create an artificial transition path. Because independence has power against clustered violations even when unconditional coverage is close to nominal, all three pieces—coverage, interval score, and tests—are necessary.

5.6 Model-search control and economic scale

The complete family comprises 20 configurations: the primary specification; 17 alternative combinations of 72-, 168-, or 336-hour windows, none/empirical-Bayes/pooled shrinkage, and raw/Kendall persistence; and two forced-kernel misspecification checks. White's Reality Check compares every configuration with Last using daily cross-sectional mean squared losses (White, 2000). The fixed-length circular-block bootstrap uses 2,000 replications and a block length of three days. The test statistic is the maximum mean daily loss reduction over configurations. This preserves dependence within daily blocks and accounts for the search at the family level.

Finally, economic interpretation is separated from statistical accuracy. We compute an illustrative complete round-trip cost allowance containing four taker fees, four spread allowances, four impact allowances, a transfer/time allowance, and a per-interval borrow allowance. The assumptions sum to 0.003301, or 33.01 bp and \$33.01 per \$10,000 notional. These values are examples, not measured order-book costs. No trading signal, leverage, position sizing, liquidation path, or realized P&L is simulated. Consequently, the exercise can reject an inference from small forecast-error improvements to executable profit, but cannot estimate a net trading return.

6 Results

6.1 The documented kernel reconstructs settlements more closely

Before assessing forecasts, we ask whether the premium accounting in Section 3 is consistent with realized settlements. For each completed eight-hour interval, we compute the premium integral from available minute observations, apply the funding transformation, and compare the result with the public settled rate. Figure 1 displays the transformation and the two candidate kernels; Figure 2 reports reconstruction RMSE on a logarithmic scale.

Linear weights outperform equal weights in all six Binance and Gate series. Binance linear-

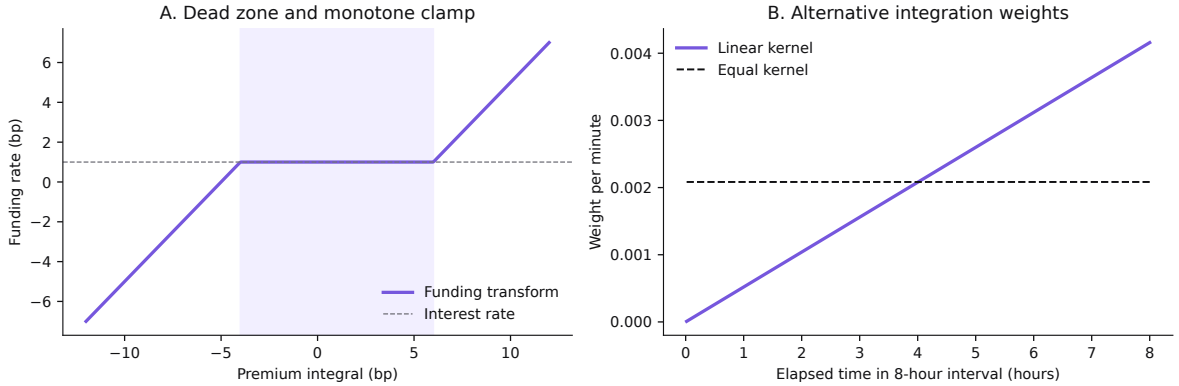


Figure 1. Funding transformation and integration kernels. Panel A shows the interest-rate dead zone and monotone clamp; Panel B compares linear and equal weights. Source: public venue rules cited in Section 3.

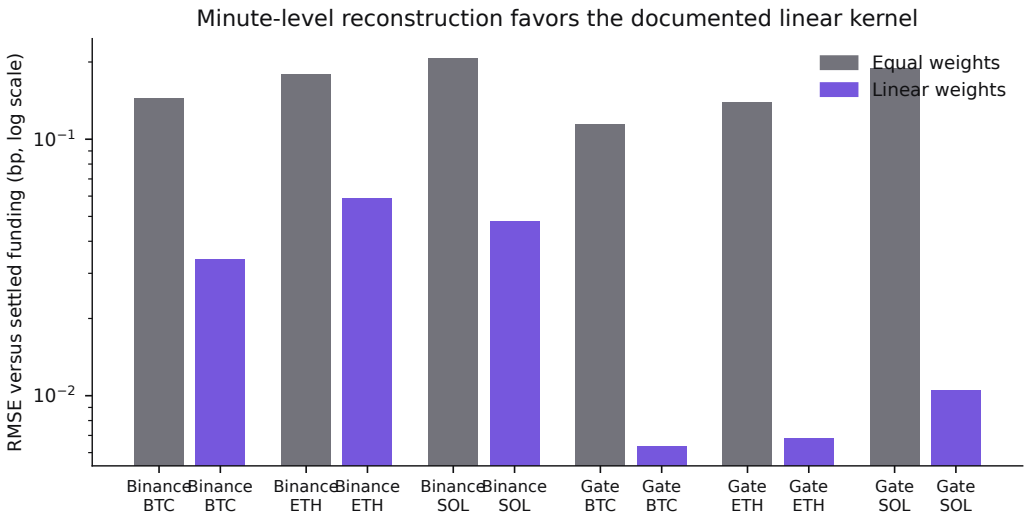


Figure 2. Reconstruction error under equal and linear kernels. Bars show RMSE against public settled funding in basis points on a log scale. Source: public minute premiums and settlements, 24 September–31 October 2026.

kernel RMSE is 0.0341 bp for BTC, 0.0587 bp for ETH, and 0.0477 bp for SOL; the corresponding equal-kernel errors are 0.1437, 0.1787, and 0.2071 bp. Gate's linear-kernel errors are smaller—0.0063, 0.0068, and 0.0105 bp—while equal-kernel errors are 0.1145, 0.1391, and 0.1894 bp. The largest linear-kernel RMSE across the six series, 0.0587 bp, is below the smallest equal-kernel RMSE, 0.1145 bp.

This result validates an important modeling choice but has a limited interpretation. It does not say that the public premium archives exactly reproduce the exchanges' native impact-price inputs or five-second sampling. Nor does it establish that a linear kernel is universally correct for other contracts or interval lengths. It shows that, within the stated minute-data approximation, the documented linear shape is materially more consistent with observed settlements than an equal-weight alternative.

6.2 OU improves on Last but not uniformly on EWMA

Table 2 and Figure 3 summarize out-of-sample point loss. The pooled OU RMSE of 0.1968 bp is 16.9% below Last's 0.2368 bp. OU also has the lowest pooled MAE, 0.0836 bp, compared with 0.0898 for EWMA and 0.1027 for Last. However, pooled OU RMSE is 1.1% above EWMA's 0.1947 bp. The contrast between MAE and RMSE indicates that OU improves typical absolute errors but does not dominate the larger squared errors.

Table 2. Out-of-sample point-forecast loss by venue

Venue	Model	n	MAE (bp)	RMSE (bp)	RMSE relative to OU
Binance	OU	276	0.3591	0.4747	1.000
	EWMA	276	0.3504	0.4638	0.977
	Last	276	0.4272	0.5722	1.205
	Interest	276	0.6262	0.7842	1.652
Gate	OU	276	0.2894	0.3866	1.000
	EWMA	276	0.2840	0.3703	0.958
	Last	276	0.3511	0.4554	1.178
	Interest	276	0.7149	0.8206	2.122
Hyperliquid	OU	2,229	0.0240	0.0437	1.000
	EWMA	2,229	0.0334	0.0604	1.381
	Last	2,229	0.0317	0.0611	1.395
	Interest	2,229	0.0387	0.0813	1.858
Pooled	OU	2,781	0.0836	0.1968	1.000
	EWMA	2,781	0.0898	0.1947	0.989
	Last	2,781	0.1027	0.2368	1.203
	Interest	2,781	0.1641	0.3649	1.854

Source: walk-forward forecasts, 1–31 October 2026. Venue rows compute losses from all targets in the venue; pooled rows weight each target equally. Lower MAE and RMSE are better.

Venue rankings explain the pooled result. On Binance, OU reduces RMSE relative to Last from 0.5722 to 0.4747 bp but is slightly behind EWMA at 0.4638 bp. On Gate, OU improves on Last from 0.4554 to 0.3866 bp but is again behind EWMA at 0.3703 bp. On Hyperliquid, OU's 0.0437 bp improves on both EWMA's 0.0604 and Last's 0.0611 bp. Thus the model's clearest advantage occurs under hourly equal-weight aggregation, not under the two minute-linear mechanisms.

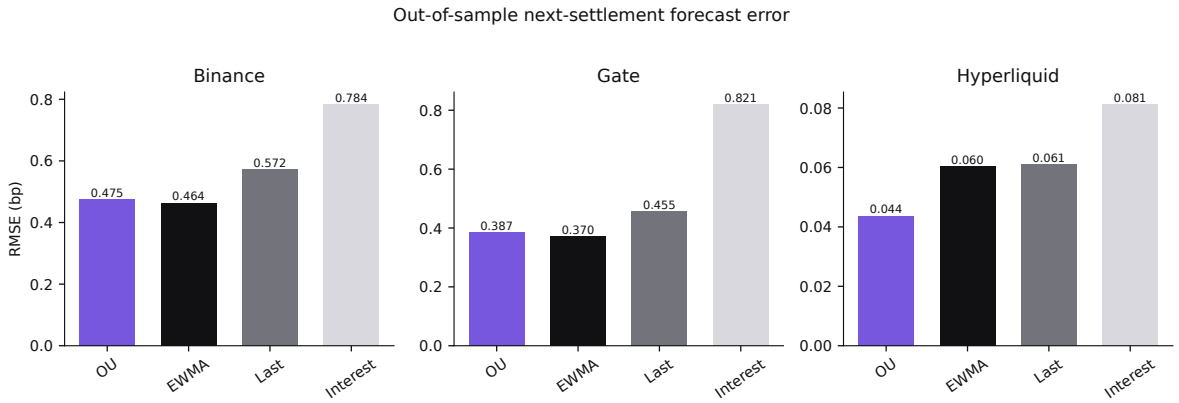


Figure 3. Out-of-sample RMSE by venue and model. Values above bars are in basis points. Source: 2,781 next-settlement forecasts over 31 evaluation days.

The asset-level results in Appendix Table 7 show that the venue pattern is not driven by one asset. OU RMSE is below Last for every venue–asset series. EWMA remains lower than OU for Binance BTC, ETH, and SOL and for Gate ETH and SOL; OU is marginally lower for Gate BTC. OU is below EWMA for all three Hyperliquid assets. These are loss estimates, not nine independent proofs: assets share market conditions and the hourly venue contributes many more targets.

Directional accuracy relative to the interest component is high for OU in this sample, but it adds little to the loss evidence. OU classifies every non-interest realization on the correct side of the interest component. This result is partly mechanical because its forecasts respect the same monotone transformation and because premiums are persistently negative in much of the sample. We do not interpret it as evidence of a profitable trading signal.

6.3 Predictive distributions succeed hourly and fail at minute-linear venues

The distributional results are more heterogeneous than the point results. Table 3 reports CRPS, logarithmic score, and empirical coverage. Hyperliquid's CRPS is 0.0166 bp and its 95% interval covers 97.7% of targets. Its 50% and 80% intervals overcover at 80.6% and 93.2%, respectively, consistent with wide intervals relative to the many interest-rate outcomes. The distribution is not perfectly calibrated, but its errors are qualitatively different from those on the other venues.

Table 3. Probabilistic forecasts and central interval coverage

Venue	CRPS (bp)	Log score	50% cov.	80% cov.	95% cov.	IS_{80}	IS_{95}
Binance	0.3054	−1.551	0.163	0.290	0.391	2.579	7.381
Gate	0.2483	0.079	0.188	0.297	0.435	2.112	6.141
Hyperliquid	0.0166	−5.367	0.806	0.932	0.977	0.109	0.170
Pooled	0.0682	−4.448	0.681	0.805	0.865	0.553	1.478

Source: primary OU walk-forward predictive distributions. Coverage is the fraction of actual rates inside the closed central interval. IS_α is the interval score in bp; lower scores are better. Log scores use natural logarithms and rate units.

Binance and Gate exhibit severe undercoverage. Their 95% intervals cover only 39.1% and 43.5% of targets; even their 50% intervals cover just 16.3% and 18.8%. Standardized

error diagnostics show why. The standard deviation of errors divided by predictive standard deviations ranges from 3.38 to 4.55 across the six series. A single normal innovation variance fitted to high-frequency premiums therefore understates next-settlement uncertainty, whether because of time-varying volatility, regime changes, parameter error, nonnormal innovations, or archive approximation.

Pooling hides this failure. Pooled 80% coverage is 80.5%, numerically close to nominal, and Kupiec's unconditional test does not reject at that level ($p = 0.470$). Per-series Christoffersen tests reject independence for all three Hyperliquid assets at 50% and for BTC and ETH at 80%; pooled 50% and 95% coverage are 68.1% and 86.5%, with unconditional-coverage rejections. No independence statistic is computed by concatenating assets into an artificial pooled sequence. The apparently satisfactory 80% number is an averaging artifact across under- and overcovering regimes, not evidence of calibration.

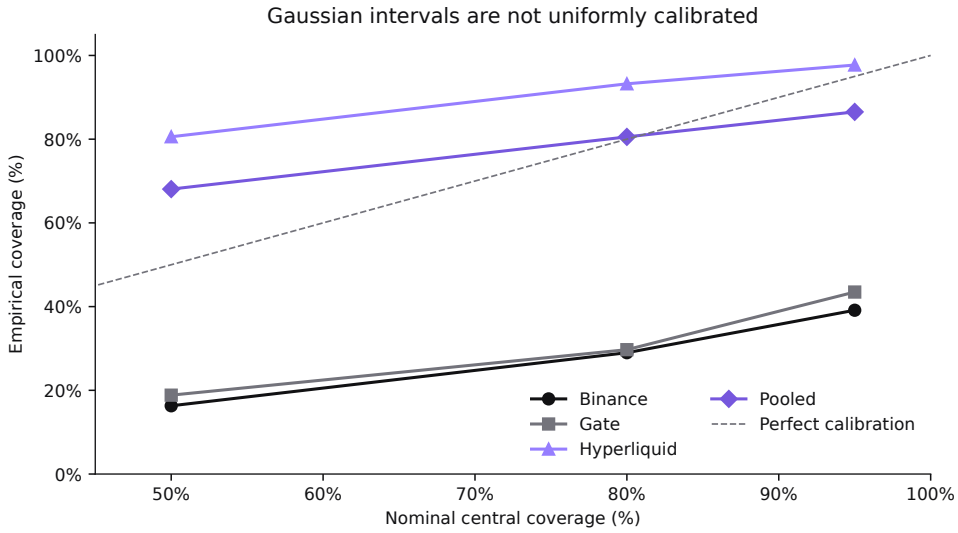


Figure 4. Nominal versus empirical central coverage by venue. The dashed line is perfect calibration. Source: primary OU predictive intervals over 2,781 out-of-sample targets.

Randomized PIT histograms in Figure 5 reinforce the diagnosis. Binance and Gate have U-shaped distributions with standard deviations well above the uniform value of approximately 0.289, indicating too many extreme realizations relative to the predictive law. Hyperliquid's PIT distribution is closer to uniform, although its interval overcoverage shows that uniformity and sharpness are not interchangeable goals. Across all targets, the PIT mean is 0.503 and standard deviation is 0.315.

6.4 Significance is selective after multiplicity control

Table 4 reports the comparisons against Last and EWMA that remain significant at 5% after Holm adjustment. Binance SOL improves on Last with a mean squared-loss reduction of 0.2192 bp^2 and adjusted $p = 0.0158$. Hyperliquid BTC and SOL also improve on Last, with adjusted $p = 0.00133$ and $p = 0.000038$. Hyperliquid SOL improves on EWMA with adjusted $p = 0.00515$. No Binance or Gate comparison against EWMA survives the correction, and no other OU–Last comparison among those six series survives either.

Across all three baselines, 11 of 27 comparisons remain significant. Several additional rejections involve the Interest baseline, which is weak by construction in this sample. They should

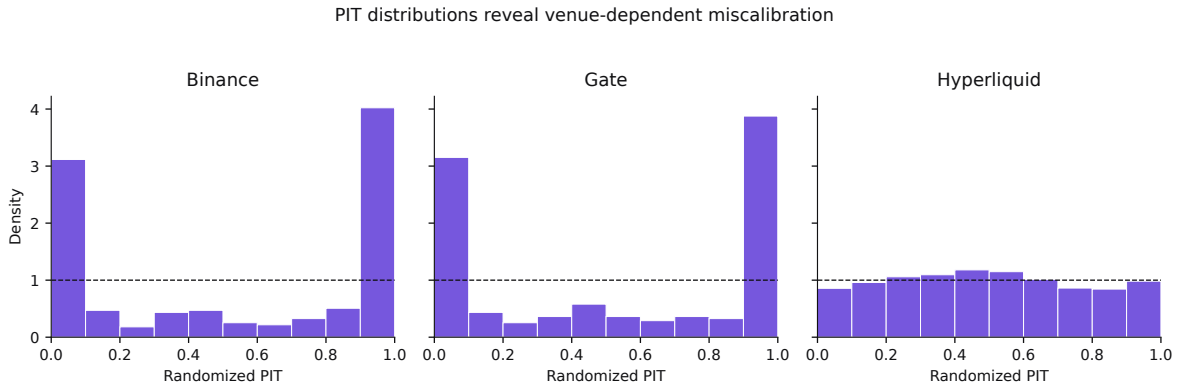


Figure 5. Randomized PIT histograms by venue. Dashed lines show the density of a uniform distribution. Source: primary OU predictive distributions; randomization is used only at point masses.

Table 4. Holm-significant OU comparisons against Last or EWMA

Venue	Asset	Baseline	Loss reduction (bp^2)	HLN t	Raw p	Holm p
Binance	SOL	Last	0.219229	3.423	0.000929	0.015788
Hyperliquid	BTC	Last	0.000780	4.011	0.000067	0.001332
Hyperliquid	SOL	Last	0.002433	4.840	0.000002	0.000038
Hyperliquid	SOL	EWMA	0.001446	3.659	0.000271	0.005150

Source: Diebold–Mariano tests with Newey–West lags of three (eight-hour series) or 23 (hourly series), HLN finite-sample correction, and Holm adjustment across all 27 comparisons. Positive loss reduction favors OU. Only adjusted $p < 0.05$ rows against Last/EWMA are shown.

not be used to inflate the evidence against the more competitive EWMA benchmark. The complete table in the reproduction package includes negative loss differences and nonsignificant results, not only the selections shown here.

The overall evidence is therefore balanced. The premium-integral model has a statistically defensible advantage in particular series, most clearly on Hyperliquid, and a broad descriptive improvement over Last. It does not establish uniform superiority over EWMA, and its Gaussian uncertainty estimates are rejected for the minute-linear venues.

7 Robustness and economic interpretation

7.1 Estimation windows

The first robustness group replaces the primary 168-hour estimation window with 72 or 336 hours while repeating the complete walk-forward. As Figure 6 shows, shorter windows produce the lowest pooled RMSE. Across all 72-hour configurations, RMSE ranges from 0.19319 to 0.19599 bp; the corresponding range is 0.19666–0.19683 bp at 168 hours and 0.19856–0.19859 bp at 336 hours. MAE increases modestly with window length as well.

This ordering suggests adaptation to a changing premium state, but the differences are small and the 72-hour winner is selected from the same family whose multiplicity we explicitly address. The best configuration is the 72-hour, unshrunk, Kendall-corrected, native-kernel model with RMSE 0.19319 bp. It should be read as the upper edge of in-sample search performance, not as a newly confirmed model for future deployment. The primary 168-hour

specification remains the basis of the main tables so that the narrative is not rewritten around the minimum.

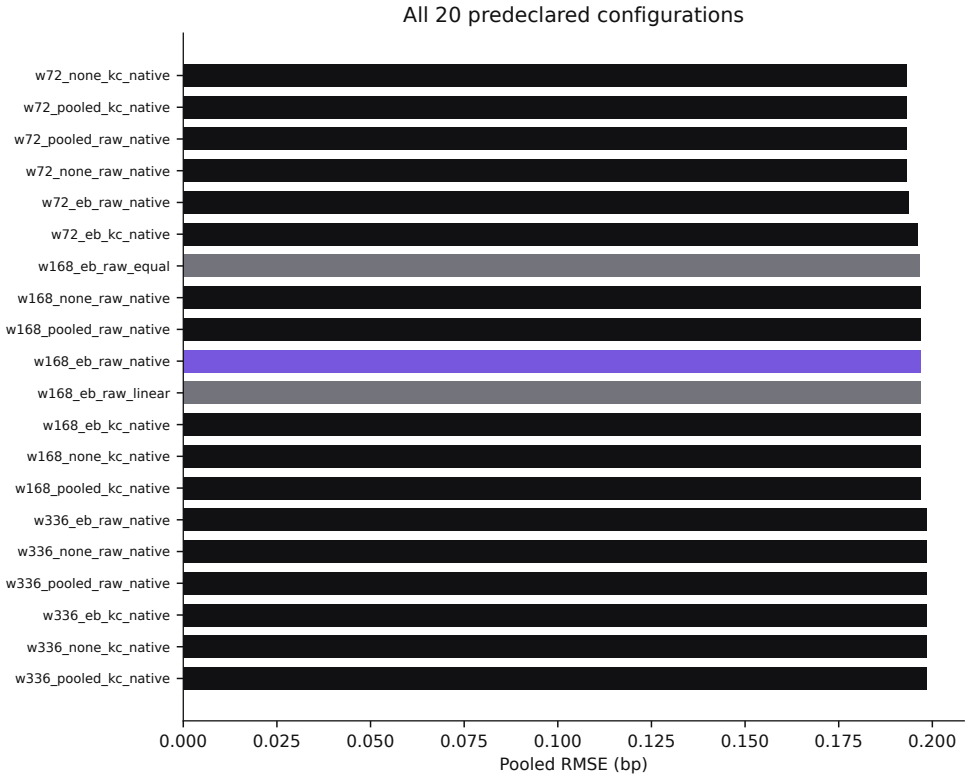


Figure 6. Pooled RMSE for all 20 configurations. The primary specification is shown in purple; forced-kernel checks are grey. Source: complete walk-forward runs for each configuration.

7.2 Shrinkage and alternative persistence estimation

The second group compares no shrinkage, empirical-Bayes shrinkage, and complete pooling of logit persistence. Across the family, minimum and maximum RMSE by shrinkage category span only about 0.0054 bp. The third group replaces raw AR persistence with the Kendall correction. Its effect is also small: raw configurations span 0.19319–0.19856 bp and corrected configurations span 0.19319–0.19859 bp. The Kendall correction can matter near a unit root, but these forecasts are dominated more by the kernel and current premium state than by a large change in estimated persistence.

Some none/pooled and raw/corrected rows have nearly identical pooled losses. This is not evidence that the estimation details are universally irrelevant. It indicates that, after averaging over 2,781 targets and applying the clamp, the remaining differences are small in this particular sample. Per-series files preserve differences that a pooled table can conceal.

7.3 Forced kernel misspecification

The fourth group holds the primary estimation design fixed and forces either equal or linear weights across venues. The forced-equal configuration has pooled RMSE 0.19666 bp, forced-linear has 0.19680 bp, and native has 0.19680 bp. These pooled differences are negligible. That result does not contradict the reconstruction evidence in Figure 2. Reconstruction applies

alternative kernels retrospectively to the entire realized interval, where wrong weights directly alter the accounting integral. Forecasting integrates an estimated mean-reverting path, and the dead zone compresses many integral values into the same funding rate. Kernel misspecification can therefore have little effect on pooled point loss while still being wrong as a settlement approximation.

The kernel checks are included in the 20-configuration multiplicity family. They are not treated as independent confirmatory tests. Their main value is to show which conclusion is sensitive to the kernel—mechanical reconstruction—and which is comparatively insensitive—short-horizon point forecasting under a clamp.

7.4 Venue and subperiod stability

The fifth group is the venue decomposition already reported in Table 2. OU's improvement over Last appears in every venue, while its ranking against EWMA changes sign between Hyperliquid and the two minute-linear venues. This is the most economically important robustness split because it corresponds to actual mechanism variation rather than a researcher-controlled parameter.

The sixth group divides October into two halves. From 1–15 October, pooled OU RMSE is 0.19592 bp, compared with 0.19582 for EWMA and 0.25492 for Last. From 16–31 October, OU RMSE is 0.19761 bp, compared with 0.19356 for EWMA and 0.21854 for Last. OU's MAE is 0.08312 and 0.08408 bp in the two periods, versus 0.09049 and 0.08909 for EWMA. The point-loss pattern is therefore stable in direction: OU beats Last, has lower MAE than EWMA, and does not clearly beat EWMA on pooled RMSE. The later-period widening of OU's RMSE disadvantage also cautions against extrapolating the best Hyperliquid comparisons to all regimes.

7.5 Reality Check and the searched family

White's Reality Check uses daily cross-sectional losses for all 20 configurations and Last. The best configuration has mean daily loss reduction 0.01872 bp^2. With 2,000 fixed-length circular-block bootstrap replications and a three-day block length, the Reality-Check p-value is 0.00050, the smallest attainable value with the plus-one correction. The null that no candidate OU configuration improves on Last is rejected.

This is stronger than an unadjusted claim based on the best row but narrower than a claim of universal superiority. The test's benchmark is Last, not EWMA. Its cross-sectional daily weighting also answers a different question from the per-series Diebold–Mariano tests. A configuration family can contain a useful alternative to Last while individual series remain uncertain and while EWMA retains a slight pooled RMSE advantage over the primary OU model.

7.6 Economic scale without fictitious P&L

The forecast errors are small in rate units. One basis point on $10,000 notional corresponds to $1 of funding, so the primary pooled OU RMSE of 0.1968 bp has the scale of $0.197 per target interval. This is an error magnitude, not an expected gain or loss from a position. The

RMSE improvement over Last, approximately 0.0400 bp, has an even smaller mechanical scale of $0.040 per interval. Neither number measures a trading opportunity because a forecast must be mapped into a position, held through entry and exit, and charged all costs.

Table 5 states the illustrative complete-cost scenario. Four taker fees contribute 20 bp, spread allowances 8 bp, impact allowances 4 bp, transfer/time 1 bp, and borrowing 0.01 bp per interval, totaling 33.01 bp. The dollar scale is $33.01 per $10,000 notional per round trip. Repeating such a round trip every eight hours corresponds to a simple annualized cost rate of 361.46%; repeating hourly corresponds to 2,891.68%. These arithmetic annualizations illustrate turnover sensitivity, not expected performance.

Table 5. Illustrative round-trip cost allowances

Component	Rate (bp)	USD per $10,000	Interpretation
Four taker fees	20.00	20.00	Entry and exit on two legs
Spread allowance	8.00	8.00	Four execution allowances
Market-impact allowance	4.00	4.00	Not depth-estimated
Transfer/time allowance	1.00	1.00	Scenario allowance
Borrow allowance per interval	0.01	0.01	Scenario allowance
Total	33.01	33.01	Not an executable P&L estimate

Source: explicit illustrative assumptions. No order-book walk, realized borrow quote, transfer observation, or position-level backtest is used. Actual costs vary by venue, asset, size, time, and account.

The economic conclusion is necessarily negative in scope, not in estimated return: the statistical results do not demonstrate a net benefit after costs. A proper economic evaluation would require timestamped order books, executable depth, fee tier, borrow availability, transfer constraints, margin and liquidation rules, and a predeclared position rule. Those inputs are outside this paper. The forecast can inform an expected-funding model, but humarb Research 01 and 02's cost and risk layers remain indispensable (humarb Research, 2026; Liu, 2026).

8 Limitations

The most important limitation is time coverage. Thirty-eight calendar days, of which 31 enter evaluation, cannot establish stability across volatility regimes, prolonged positive-premium periods, exchange-rule changes, or market stress. Negative average premiums and frequent interest-rate settlements characterize this sample. A model that performs well under these conditions may not do so when premiums reverse or when outer caps bind. The hourly Hyperliquid sample has 2,229 targets, but these targets are not 2,229 independent regimes; their effective information is limited by serial and cross-asset dependence.

The cross-section is also narrow. BTC, ETH, and SOL remained listed and observable at all three venues. No delistings, prolonged suspensions, contract migrations, or new listings enter the test. Although this avoids deleting failures after seeing their outcomes, it does not provide a delisting-aware universe for a general forecasting claim. Extending the method requires point-in-time contract eligibility, explicit suspension handling, and a larger set of assets, including less liquid contracts whose premium behavior and data quality may differ.

Public data are approximations to native settlement inputs. Two Binance daily premium archives per asset are unavailable, Gate has a small number of missing minutes, and the Binance

model grid is coarser than native five-second sampling. Public archives can differ from real-time displays or later revisions. Reconstruction errors quantify the combined discrepancy but cannot separate every source. A missing interval is not treated as if it were observed, yet missingness can still affect fitted persistence and predictive variance.

The OU model itself is intentionally restrictive. It assumes linear mean reversion, constant innovation variance, and Gaussian innovations within each rolling window. It omits stochastic volatility, jumps, asymmetric adjustment, intraday seasonality, common market factors, parameter uncertainty, and uncertainty about the kernel or current premium measurement. The Binance and Gate coverage failures show that these omissions are consequential. A post hoc widening of intervals until pooled coverage looks attractive would hide rather than solve the problem; a defensible extension requires new out-of-sample evaluation of a predeclared volatility or calibration model.

Shrinkage is limited to three assets within a venue and uses approximate sampling variances. It does not borrow information across venues, estimate a hierarchical volatility model, or account for correlation between assets. The empirical-Bayes variance estimate can be zero, producing complete pooling. The near-identical losses under some shrinkage alternatives should not be read as validation of the approximation.

The walk-forward design eliminates direct target leakage, but it does not eliminate researcher selection. Window lengths, shrinkage modes, bias corrections, kernels, baselines, and evaluation choices form a modeling family. I report 20 configurations and apply Holm and Reality-Check procedures, but these methods cannot include every choice made during the broader research process. The best 72-hour row remains a selected result. External validation on a later, untouched period is the appropriate next test.

The tests have additional sampling limitations. Newey–West lag choices of three and 23 are transparent daily approximations, not data-optimal guarantees. The nine Diebold–Mariano series share market shocks, and Holm adjustment controls a specified family rather than all conceivable analyses. The Reality Check uses only 31 daily observations; its block bootstrap approximates, but cannot fully recover, the time-series dependence of model comparison. Log scores are sensitive to the numerical probability floor, and CRPS uses finite quantile quadrature. These numerical choices are disclosed to make the calculations reproducible.

Interval tests diagnose calibration but do not identify its remedy. Hyperliquid's overcoverage can coexist with useful PIT behavior, while Binance and Gate's undercoverage produces large interval-score penalties. Pooled scores are dominated in number by hourly targets and should not be used to choose a venue-specific risk model. Equal target weighting is an evaluation convention, not an economic portfolio weight.

No executable economic conclusion is tested. The cost scenario is illustrative rather than measured, and the paper contains no order-book replay, position sizing, leverage, borrowing history, transfer latency, liquidation simulation, or realized return. Forecast error in basis points cannot be subtracted from cost to create a profit estimate. The appropriate conclusion is that economic value remains unproven, not that it is zero under every possible strategy.

Finally, this is a methodological and empirical preprint, not a peer-reviewed article. Similarity to prior text was reviewed through source-aware drafting and bibliographic checks, but no external iThenticate or Turnitin report was available. That administrative limitation does not alter the empirical results and is stated separately in the delivery self-check.

9 Conclusion

This paper turns the public funding mechanism into a next-settlement forecasting model. The premium is a mean-reverting state, the venue kernel determines its next-interval integral, and the interest adjustment and clamp determine the settled funding distribution. Exact discrete AR(1) moments make the construction computationally simple and fully reproducible.

The mechanism is empirically informative. Linear kernels reconstruct Binance and Gate settlements more closely than equal kernels, and OU point forecasts improve on Last across the reported venues. In the pooled sample, OU lowers RMSE from 0.2368 to 0.1968 bp. The improvement is strongest on Hyperliquid, where OU also beats EWMA. On Binance and Gate, however, EWMA retains slightly lower RMSE. After multiplicity control, only selected OU comparisons against Last and one against EWMA remain significant.

The distributional answer is more cautious. Gaussian intervals severely undercover on Binance and Gate and overcover on Hyperliquid. A pooled 80% coverage rate close to nominal is an averaging artifact, not evidence of reliability. Mechanism-aware point forecasting and calibrated probabilistic forecasting are therefore separate achievements; this paper supports the former in selected settings and rejects the latter as a universal claim.

The results do not establish executable net returns. Forecast improvements are small relative to an illustrative complete round-trip cost, and no execution or position-level backtest is performed. Future work should preserve the walk-forward discipline while testing later periods, richer volatility models, point-in-time contract universes, and depth-based economic evaluation.

Declarations

Authorship and corresponding author

Hang Liu is solely responsible for the research design, analysis, and manuscript. Correspondence: humarb Research, London, United Kingdom; research channel https://humarb.ai/research/; ORCID 0009-0008-1059-3926. This paper was completed by humarb Research, and humarb.ai is its public research channel.

Conflicts of interest

The author declares no external financial conflicts of interest related to this study. The research concerns public market mechanisms and does not constitute a solicitation for any venue, asset, or trading product.

Funding

No external grant or sponsor funded this work. humarb Research supported the research and publication preparation.

Data availability

The empirical data are derived from public Binance, Gate, and Hyperliquid archives or public endpoints, accessed on 3 November 2026. The reproduction package contains normalized public data, a source manifest with URLs or request payloads and content hashes, and instructions for independently acquiring the same sample. Venue documentation is cited in the references. No private user data, credentials, or undisclosed interfaces are used.

Code availability

The reproduction package includes the public-data acquisition script, deterministic analysis script, numerical tests, figure-generation script, data dictionary, and run instructions. Forecast-level outputs and all 20 robustness configurations are included so that the reported tables and figures can be checked independently.

Generative AI use

Generative AI was used to assist with literature retrieval, code drafting and debugging, numerical-result checking, language drafting, and layout preparation. All quantitative outputs were generated by deterministic scripts from the stated public data. The author reviewed the methods, references, calculations, and final text and accepts responsibility for the content. No generative system was used to invent observations, citations, or publication identifiers.

Research status and license

This is humarb Research 06, a methodological and empirical preprint. It has not been peer reviewed. The DOI, Zenodo record, version DOI, and formal publication date are pending at the time of this delivery. The intended license is Creative Commons Attribution 4.0 International (CC BY 4.0).

Not investment advice

This paper is for research and methodological information only. It does not provide investment, legal, accounting, or tax advice; does not recommend a trade or position; and does not promise accuracy, profit, or risk-free returns. Forecasts are historical estimates under stated assumptions. Actual execution, funding, borrowing, margin, and liquidation outcomes can differ materially.

References

Damien Ackerer, Julien Hugonnier, and Urban Jermann. Perpetual futures pricing. *Mathematical Finance*, 36(3):481–499, 2025. doi: 10.1111/mafi.70018.

Jeremy Berkowitz. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474, 2001. doi: 10.1198/07350010152596718.

Binance. Introduction to binance futures funding rates, 2026. URL https://www.binance.com/en/support/faq/introduction-to-binance-futures-funding-rates-360033525031. Accessed October 15, 2026.

Peter F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4): 841–862, 1998. doi: 10.2307/2527341.

Min Dai, Linfeng Li, and Chen Yang. Arbitrage in perpetual contracts. SSRN Working Paper 5262988, 2025.

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

Gate. Futures funding rate and funding fee overview, 2026. URL https://www.gate.com/help/futures/futures-logic/27569/funding-rate-and-funding-fee. Accessed October 15, 2026.

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

David Harvey, Stephen Leybourne, and Paul Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291, 1997. doi: 10.1016/S0169-2070(96)00719-4.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

humarb Research. Measuring executable net carry in cross-venue perpetual funding-rate arbitrage, 2026.

Hyperliquid. Funding, 2026. URL https://hyperliquid.gitbook.io/hyperliquid-docs/trading/funding. Accessed October 15, 2026.

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

M. G. Kendall. Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3/4):403–404, 1954. doi: 10.1093/biomet/41.3-4.403.

Paul H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84, 1995. doi: 10.3905/jod.1995.407942.

Hang Liu. Framing net carry for cross-venue perpetual funding positions: A settled-history, depth-walked, and liquidation-aware method, 2026.

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

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

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

A Notation and computational definitions

Table 6 collects the principal symbols. Rates are decimal fractions unless a table explicitly reports basis points. The observation grid is one minute for Binance and Gate and one hour for Hyperliquid; the settlement scale s converts the common eight-hour premium frame to the paid settlement rate.

Table 6. Principal notation

Symbol	Domain	Definition
p_j	$\mathbb{R}$	Premium state at grid step j
θ	$\mathbb{R}$	Long-run premium mean in the discrete OU approximation
ϕ	$(0,1)$	One-step AR persistence, equal to $e^{-\kappa}$ on a unit grid
σ_ε^2	$\mathbb{R}_+$	Conditional variance of the one-step premium innovation
n	$\mathbb{N}$	Number of grid steps in the next settlement interval
w_j	$\mathbb{R}_+$	Normalized integration-kernel weight, with $\sum_j w_j = 1$
A	$\mathbb{R}$	Kernel-weighted future premium integral, $\sum_{j=1}^{n} w_j p_j$
μ_A, σ_A^2	$\mathbb{R}, \mathbb{R}_+$	Conditional mean and variance of A before the funding transformation
H_j	$\mathbb{R}$	Cumulative coefficient mapping innovation j into A
I_8	$\mathbb{R}$	Eight-hour reference interest component, fixed here at 0.0001
δ	$\mathbb{R}_+$	Half-width of the interest-rate dead zone, fixed here at 0.0005
C	$\mathbb{R}_+$	Working eight-hour outer cap, fixed here at 0.02 and nonbinding in the sample
$q(a)$	$\mathbb{R}$	Eight-hour funding transformation, $\mathrm{clip}[a + \mathrm{clamp}(I_8 - a, -\delta, \delta), -C, C]$
s	$\mathbb{R}_+$	Settlement scale: one for eight-hour venues and 1/8 for Hyperliquid
F	$\mathbb{R}$	Realized next-settlement funding rate, equal to $sq(A)$ under the mechanism approximation
$\widehat{F}$	$\mathbb{R}$	Out-of-sample point forecast of F
G	$[0,1]$	Predictive cumulative distribution after the monotone funding transformation
$Q(u)$	$\mathbb{R}$	Predictive quantile at probability level u
L_α, U_α	$\mathbb{R}$	Lower and upper endpoints of the central $100\alpha\%$ predictive interval
d_t	$\mathbb{R}$	Squared-loss differential, baseline loss minus OU loss; positive favors OU

B Additional derivations

B.1 Exact variance recursion

Conditional on the current premium, write $p_j - \theta = \phi^j(p_0 - \theta) + \sum_{\ell=1}^{j} \phi^{j-\ell}\varepsilon_\ell$. Substitution into $A = \sum_j w_j p_j$ shows that innovation ε_ℓ has coefficient

$$\sum_{j=\ell}^{n} w_j \phi^{j-\ell} = w_\ell + \phi w_{\ell+1} + \cdots + \phi^{n-\ell} w_n = H_\ell. \tag{17}$$

The backward recursion in equation (9) computes these coefficients in $O(n)$ time. Independence of the innovations yields $\mathrm{Var}(A \mid p_0) = \sigma_\varepsilon^2 \sum_j H_j^2$ without a simulation approximation. The conditional mean follows in the same way because the innovation terms vanish in expectation.

For a one-step interval, $n = 1$, $w_1 = 1$, and $H_1 = 1$, so the expressions reduce to $\mu_A = \theta + \phi(p_0 - \theta)$ and $\sigma_A^2 = \sigma_\varepsilon^2$. This special case is useful for checking the hourly implementation. It also explains why the hourly model remains a genuine next-settlement forecast even though no within-interval path is simulated.

B.2 Predictive CDF and probability masses

Away from point masses, monotonicity allows direct inversion of q. For $-C < y < I_8$, the corresponding premium boundary is $y - \delta$; for $I_8 < y < C$, it is $y + \delta$. Consequently,

$$G(y) = \begin{cases} 0, & y < -C, \\ \Phi\{z(-C - \delta)\}, & y = -C, \\ \Phi\{z(y - \delta)\}, & -C < y < I_8, \\ \Phi\{z(y + \delta)\}, & I_8 < y < C, \\ \Phi\{z(C + \delta)\}, & y = C, \\ 1, & y > C. \end{cases} \tag{18}$$

The interest-rate point mass is the jump across the dead zone,

$$\Pr(q(A) = I_8 \mid p_0) = \Phi\{z(I_8 + \delta)\} - \Phi\{z(I_8 - \delta)\}. \tag{19}$$

At $-C$ and C, the masses are $\Phi\{z(-C - \delta)\}$ and $1 - \Phi\{z(C + \delta)\}$, respectively. Equation (18) is expressed on the eight-hour frame; for Hyperliquid, outcomes and boundaries are scaled once by $s = 1/8$ after evaluating the common transformation.

The randomized PIT is $U \sim \mathrm{Unif}\{G(y-), G(y)\}$ at a mass and $G(y)$ elsewhere. Randomness is used only for this diagnostic and is seeded. Point forecasts, quantiles, CRPS quadrature, and all loss comparisons are deterministic given the public-data snapshot.

B.3 Expected clamped normal with outer caps

Define $a_- = I_8 - \delta$ and $a_+ = I_8 + \delta$. The uncapped dead-zone transformation equals $A + \delta$ when $A < a_-$, I_8 when $a_- \le A \le a_+$, and $A - \delta$ when $A > a_+$. Its expectation is therefore

$$I_8 - E[(a_- - A)\mathbf{1}(A < a_-)] + E[(A - a_+)\mathbf{1}(A > a_+)], \tag{20}$$

which gives equation (13). The implemented capped expectation additionally replaces values below $-C$ by $-C$ and values above C by C. In loss-function notation, the cap correction is

$$-L_-(C + \delta) + L_+(-C - \delta), \tag{21}$$

added to the uncapped expression with the sign conventions used in Section 5. The numerical test suite compares the result with simulation and checks the zero-variance limit.

B.4 Proper-score conventions

The quantile approximation to CRPS uses the identity $\mathrm{CRPS}(F, y) = 2\int_0^1 \rho_u(y - Q(u))\, du$, where $\rho_u(x) = x\{u - \mathbf{1}(x < 0)\}$. The reported sum in equation (15) uses 199 interior levels, avoiding undefined normal quantiles at zero and one. The logarithmic score is evaluated on the mixed law rather than on a Gaussian density at the already-clamped outcome. This distinction is essential at the interest component, where a continuous density would assign zero probability to an outcome that occurs frequently.

For central intervals, the score in equation (16) rewards narrow intervals but penalizes misses in proportion to their distance and the tail probability $1 - \alpha$. Lower CRPS, logarithmic score, and interval score are better. These scores have different units and are not compared numerically with one another.

C Reproducible forecast protocol

The following sequence is repeated independently at every forecast origin and for every model configuration:

1. Sort and deduplicate the public premium and settlement histories; retain only observations available at the origin.
2. Select the trailing estimation window and fit each asset's discrete AR(1) by least squares. Apply the declared persistence correction and within-venue shrinkage rule.
3. Identify the last available premium state and the next actual non-overlapping settlement. Do not use any observation from the target interval to estimate the state or parameters.
4. Construct the native or deliberately forced kernel, calculate μ_A and σ_A^2 from equations (8)–(10), and apply the funding transformation once at the applicable settlement scale.
5. Store the point forecast, predictive quantiles, mixed-law score components, and baseline forecasts before joining the realized target.

6. Compute errors, interval violations, randomized PIT, and proper scores. Aggregate only after every origin has been processed.

7. Repeat the complete walk-forward for all 20 configurations. Apply Holm correction across the 27 Diebold–Mariano comparisons and run the family-level Reality Check against Last.

The primary example uses a 168-hour window, empirical-Bayes logit-persistence shrinkage, raw AR persistence, and native kernels. These values are disclosed research examples and do not represent any production setting. The bootstrap uses 2,000 fixed-length circular-block replications with a three-day block length; randomized PIT evaluation uses a fixed seed. The public-data acquisition snapshot is 3 November 2026. Numeric results in the paper are generated from the stored forecast outputs rather than transcribed from a separate trading system.

D Complete empirical tables

D.1 Asset-level forecast loss and distribution scores

Table 7. Complete out-of-sample point-forecast metrics by venue and asset

Venue	Asset	Model	n	MAE (bp)	RMSE (bp)	Bias (bp)	Direction hit
Binance	BTC	OU	92	0.2832	0.3479	-0.0193	100.00%
Binance	BTC	EWMA	92	0.2816	0.3416	-0.0251	100.00%
Binance	BTC	Last	92	0.3070	0.3890	-0.0059	91.25%
Binance	BTC	Interest	92	0.5683	0.6878	0.5683	0.00%
Binance	ETH	OU	92	0.3544	0.4837	0.0520	100.00%
Binance	ETH	EWMA	92	0.3368	0.4722	0.0553	100.00%
Binance	ETH	Last	92	0.4111	0.5392	0.0254	98.90%
Binance	ETH	Interest	92	0.7017	0.8503	0.7017	0.00%
Binance	SOL	OU	92	0.4396	0.5665	0.0309	100.00%
Binance	SOL	EWMA	92	0.4327	0.5529	-0.0291	100.00%
Binance	SOL	Last	92	0.5637	0.7350	-0.0165	80.56%
Binance	SOL	Interest	92	0.6086	0.8053	0.6086	0.00%
Gate	BTC	OU	92	0.2141	0.2731	-0.0418	100.00%
Gate	BTC	EWMA	92	0.2251	0.2772	-0.0314	100.00%
Gate	BTC	Last	92	0.2772	0.3658	-0.0078	96.39%
Gate	BTC	Interest	92	0.6114	0.7064	0.6114	0.00%
Gate	ETH	OU	92	0.2818	0.3724	0.0486	100.00%
Gate	ETH	EWMA	92	0.2647	0.3542	0.0327	100.00%
Gate	ETH	Last	92	0.3259	0.4109	0.0065	98.90%
Gate	ETH	Interest	92	0.7098	0.8011	0.7098	0.00%
Gate	SOL	OU	92	0.3722	0.4849	0.0115	100.00%
Gate	SOL	EWMA	92	0.3623	0.4573	-0.0047	100.00%
Gate	SOL	Last	92	0.4502	0.5653	-0.0059	94.25%
Gate	SOL	Interest	92	0.8236	0.9379	0.8236	0.00%
Hyperliquid	BTC	OU	743	0.0150	0.0299	-0.0003	100.00%
Hyperliquid	BTC	EWMA	743	0.0204	0.0376	-0.0000	100.00%
Hyperliquid	BTC	Last	743	0.0196	0.0409	0.0000	77.64%
Hyperliquid	BTC	Interest	743	0.0247	0.0545	0.0247	0.00%
Hyperliquid	ETH	OU	743	0.0262	0.0473	-0.0001	100.00%
Hyperliquid	ETH	EWMA	743	0.0398	0.0741	0.0001	99.41%
Hyperliquid	ETH	Last	743	0.0350	0.0668	0.0000	77.81%
Hyperliquid	ETH	Interest	743	0.0473	0.1000	0.0473	0.00%
Hyperliquid	SOL	OU	743	0.0309	0.0511	-0.0025	100.00%
Hyperliquid	SOL	EWMA	743	0.0401	0.0637	-0.0010	100.00%
Hyperliquid	SOL	Last	743	0.0406	0.0710	-0.0003	73.10%

Hyperliquid	SOL	Interest	743	0.0442	0.0827	0.0442	0.00%

Source: primary walk-forward forecasts, 1–31 October 2026. Bias is forecast minus realization. Direction hit is relative to the applicable interest component and excludes realizations exactly at that component. Lower MAE and RMSE are better; no row is deleted for an unfavorable result.

Table 8. OU probabilistic scores and empirical coverage by series

Venue	Asset	CRPS	Log	Cov. 50	Cov. 80	Cov. 95	IS_{80}	IS_{95}
Binance	BTC	0.2369	-3.244	0.163	0.283	0.413	1.963	5.344
Binance	ETH	0.2976	-3.998	0.207	0.370	0.478	2.467	6.813
Binance	SOL	0.3819	2.589	0.120	0.217	0.283	3.307	9.986
Gate	BTC	0.1819	-2.376	0.174	0.272	0.402	1.537	4.397
Gate	ETH	0.2452	5.687	0.174	0.293	0.424	2.121	6.460
Gate	SOL	0.3177	-3.074	0.217	0.326	0.478	2.679	7.566
Hyperliquid	BTC	0.0102	-4.779	0.844	0.949	0.984	0.066	0.099
Hyperliquid	ETH	0.0184	-5.692	0.782	0.917	0.970	0.122	0.186
Hyperliquid	SOL	0.0211	-5.630	0.791	0.931	0.977	0.141	0.226

Source: primary OU predictive distributions. CRPS and interval scores are in bp; log scores use natural logarithms and rate units. Coverage columns are fractions. The full coverage-test output, including all violation counts and test statistics, is retained in the public results archive.

Table 9. Pooled interval-coverage tests

Nominal	Empirical	Violations	n	Kupiec p	Independence p
0.50	0.6807	888	2781	8.66×10^{-83}	–
0.80	0.8055	541	2781	0.4696	–
0.95	0.8652	375	2781	7.03×10^{-66}	–

Source: primary OU intervals pooled across nine series. Kupiec values are descriptive because pooling creates a heterogeneous, cross-sectionally dependent target set. No Christoffersen independence statistic is computed for the artificial pooled transition sequence; per-series tests remain in the reproduction results.

D.2 Complete Diebold–Mariano family

Table 10. All OU–baseline squared-loss comparisons with Holm adjustment

Venue	Asset	Baseline	n	Lag	Reduction	HLN t	Raw p	Holm p
Binance	BTC	EWMA	92	3	-0.004287	-0.455	0.6500	1.0000
Binance	BTC	Interest	92	3	0.352069	4.613	1.29×10^{-5}	2.71×10^{-4}
Binance	BTC	Last	92	3	0.030316	1.579	0.1178	1.0000
Binance	ETH	EWMA	92	3	-0.011035	-0.863	0.3902	1.0000
Binance	ETH	Interest	92	3	0.489074	4.741	7.81×10^{-6}	1.72×10^{-4}
Binance	ETH	Last	92	3	0.056763	1.030	0.3059	1.0000
Binance	SOL	EWMA	92	3	-0.015187	-0.560	0.5769	1.0000
Binance	SOL	Interest	92	3	0.327639	5.051	2.25×10^{-6}	5.18×10^{-5}
Binance	SOL	Last	92	3	0.219229	3.423	9.29×10^{-4}	0.0158
Gate	BTC	EWMA	92	3	0.002265	0.285	0.7762	1.0000
Gate	BTC	Interest	92	3	0.424393	5.678	1.61×10^{-7}	4.02×10^{-6}
Gate	BTC	Last	92	3	0.059207	2.751	0.0072	0.1075
Gate	ETH	EWMA	92	3	-0.013238	-1.112	0.2691	1.0000
Gate	ETH	Interest	92	3	0.503121	6.777	1.19×10^{-9}	3.22×10^{-8}
Gate	ETH	Last	92	3	0.030167	0.942	0.3486	1.0000
Gate	SOL	EWMA	92	3	-0.026017	-1.353	0.1795	1.0000
Gate	SOL	Interest	92	3	0.644451	6.240	1.36×10^{-8}	3.54×10^{-7}
Gate	SOL	Last	92	3	0.084382	1.732	0.0867	0.8675
Hyperliquid	BTC	EWMA	743	23	0.000518	2.867	0.0043	0.0681
Hyperliquid	BTC	Interest	743	23	0.002082	2.348	0.0191	0.2488
Hyperliquid	BTC	Last	743	23	0.000780	4.011	6.66×10^{-5}	0.0013
Hyperliquid	ETH	EWMA	743	23	0.003253	1.845	0.0654	0.7194
Hyperliquid	ETH	Interest	743	23	0.007773	1.948	0.0518	0.6217
Hyperliquid	ETH	Last	743	23	0.002229	2.648	0.0083	0.1158

Continued on next page

Venue	Asset	Baseline	n	Lag	Reduction	HLN t	Raw p	Holm p
Hyperliquid	SOL	EWMA	743	23	0.001446	3.659	2.71×10^{-4}	0.0051
Hyperliquid	SOL	Interest	743	23	0.004230	3.425	6.48×10^{-4}	0.0117
Hyperliquid	SOL	Last	743	23	0.002433	4.840	1.58×10^{-6}	3.80×10^{-5}

Source: Newey–West Diebold–Mariano tests with HLN correction. Reduction is baseline squared loss minus OU squared loss in bp^2, so positive values favor OU. Holm adjustment is across all 27 rows jointly. Lag denotes the Newey–West lag order.

D.3 Complete robustness family

Table 11. All 20 walk-forward configurations

ID	Window	Shrinkage	Persistence	Kernel	n	MAE	RMSE	CRPS
1	72	EB	Raw	Native	2781	0.082930	0.193823	0.067302
2	72	EB	KC	Native	2781	0.083304	0.195994	0.067680
3	72	None	Raw	Native	2781	0.082734	0.193191	0.067089
4	72	None	KC	Native	2781	0.082807	0.193188	0.067056
5	72	Pooled	Raw	Native	2781	0.082734	0.193191	0.067089
6	72	Pooled	KC	Native	2781	0.082807	0.193188	0.067056
7	168	EB	Raw	Equal	2781	0.083572	0.196658	0.069192
8	168	EB	Raw	Linear	2781	0.083619	0.196798	0.068227
9	168	EB	Raw	Native	2781	0.083619	0.196798	0.068227
10	168	EB	KC	Native	2781	0.083661	0.196800	0.068218
11	168	None	Raw	Native	2781	0.083633	0.196797	0.068219
12	168	None	KC	Native	2781	0.083735	0.196825	0.068246
13	168	Pooled	Raw	Native	2781	0.083633	0.196797	0.068219
14	168	Pooled	KC	Native	2781	0.083735	0.196825	0.068246
15	336	EB	Raw	Native	2781	0.086814	0.198556	0.070737
16	336	EB	KC	Native	2781	0.086888	0.198575	0.070753
17	336	None	Raw	Native	2781	0.086825	0.198558	0.070737
18	336	None	KC	Native	2781	0.086921	0.198586	0.070764
19	336	Pooled	Raw	Native	2781	0.086825	0.198558	0.070737
20	336	Pooled	KC	Native	2781	0.086921	0.198586	0.070764

Source: complete walk-forward runs. Window is in hours; MAE, RMSE, and CRPS are in bp. Raw and KC denote raw AR persistence and Kendall-corrected persistence. EB denotes empirical-Bayes logit-persistence shrinkage. Native uses each venue's documented kernel; Equal and Linear are forced misspecification checks. The primary specification is ID 9; IDs are table identifiers, not a ranking.

E Public prediction-table dictionary

Table 12 defines the public forecast-output schema. All rate-valued fields are decimal fractions unless the field name ends in `_bp`, in which case the unit is one basis point. Timestamps are UTC. Empty probabilistic fields for point-only baselines are intentional. The separate data-acquisition guide defines the normalized public input fields and the source-manifest fields.

Table 12. Data dictionary for public out-of-sample predictions

Field	Type	Definition
`venue`	string	Public venue identifier: Binance, Gate, or Hyperliquid.
`asset`	string	Contract asset: BTC, ETH, or SOL.
`origin`	timestamp	Actual settlement origin at which the forecast is issued.
`target_time`	timestamp	Timestamp of the next non-overlapping realized settlement.
`hours`	number	Settlement interval in hours; eight for Binance/Gate and one for Hyperliquid.
`model`	string	Forecast model: OU, EWMA, Last, or Interest.

Continued on next page

Field	Type	Definition
`actual`	rate	Realized settled funding rate.
`prediction`	rate	Out-of-sample point forecast.
`error_bp`	basis points	Forecast minus realization, multiplied by 10,000.
`crps_bp`	basis points	Quantile-quadrature CRPS for OU; empty for point-only baselines.
`pit`	probability	Randomized probability integral transform for OU; empty otherwise.
`mu`	premium rate	Conditional mean μ_A on the common eight-hour premium frame.
`sd`	premium rate	Conditional standard deviation σ_A on the same frame.
`interest`	rate	Applicable paid-settlement interest component after settlement scaling.
`log_score`	score	Negative natural log predictive density or point mass; empty for baselines.
`window_hours`	hours	Trailing estimation-window length for the disclosed research example.
`shrink`	string	Persistence shrinkage: none, empirical Bayes, or pooled.
`bias_correction`	Boolean	Whether the Kendall AR-persistence correction is applied.
`kernel`	string	Native, forced equal, or forced linear integration kernel.
`phi`	probability	Estimated one-step premium persistence used at the origin.
`theta`	premium rate	Estimated long-run premium mean used at the origin.
`lower_50, upper_50`	rate	Endpoints of the central 50% predictive interval.
`interval_score_50_bp`	basis points	Central 50% interval score.
`lower_80, upper_80`	rate	Endpoints of the central 80% predictive interval.
`interval_score_80_bp`	basis points	Central 80% interval score.
`lower_95, upper_95`	rate	Endpoints of the central 95% predictive interval.
`interval_score_95_bp`	basis points	Central 95% interval score.

F Reproduction checks and evidence map

The numerical tests check seven properties independently of the empirical rankings: the lower, dead-zone, upper, and capped branches of the funding transformation; monotone transformed quantiles; the clamped-normal mean against simulation; the transformed CDF against simulation; the exact discrete integral moments against simulation; one and only one Hyperliquid settlement scaling; and inclusion of realizations at point-mass interval boundaries. These are implementation checks, not statistical evidence of forecast superiority.

The empirical claim-to-evidence map is deliberately narrow. Kernel reconstruction supports the documented linear minute weights on Binance and Gate within the public-data approximation. Walk-forward losses support mechanism-aware point forecasting and a descriptive improvement over Last. Venue splits and Holm-adjusted tests restrict the strongest confirmatory comparisons to particular series. Coverage, PIT, and proper scores reject a universal calibrated Gaussian law. The Reality Check addresses the searched family only against Last. The cost scenario demonstrates economic scale but supplies no executable return estimate. No table or figure should be read beyond those boundaries.