

Kalshi Copper 15-Minute — The Last 5 Minutes

Local Mathematics: every model tested on all 3,656 settled windows since launch

Chronological 70/30 holdout throughout — nothing in this volume is fitted on the windows it is graded on.

Moves are self-consistent (spot move since window open) — the copper method, validated 21,936 cells / 0 mismatches against the uploaded sheet.

Sources: Kalshi KXCOPPER15M markets + 1-minute candlesticks, Binance/Yahoo copper spot (local data only)

Kalshi Copper 15-Minute — The Last 5 Minutes: Complete Local Mathematics Report

Scope: minutes left **5, 4, 3, 2, 1 only**. Local files only — no web research, no trades, no live jobs touched. All probability scores are judged on a **chronological 70/30 holdout**: models and sigma are fitted/estimated on the first 70% of windows and scored, untouched, on the last 30%.

1. Dataset — verified before any modelling

Item	Value
Settled windows (rows)	3,656
Up / Down	1,812 / 1,844
First window	2026-08-27T20:45:00Z UTC (Aug 27, 01:45 PM PT) — first usable window of the series
Last window	2026-10-21T07:30:00Z UTC (Oct 21, 12:30 AM PT)
Train (first 70%)	2,559 windows, 2026-08-27T20:45:00Z → 2026-10-05T01:45:00Z UTC
Holdout (last 30%)	1,097 windows, 2026-10-05T02:00:00Z → 2026-10-21T07:30:00Z UTC
Validation vs uploaded sheet	Rebuilt self-consistent moves vs the sheet_cps values in data_copper.json at all six sheet checkpoints (12/10/8/6/4/2): 21,936 cells checked, 0 mismatches (tolerance 0.0011), 0 missing. This extends the earlier 12/10/8/6 validation (14,624 cells, 0 mismatches) to 4 and 2 minutes left with the same result.
Missing data at 5–1 min	Checkpoint moves missing: 0 of 3,656 at every minute 5–1. Kalshi % missing: 0 of 3,656 at every minute 5–1. Copper trades weekdays only, with daily breaks and weekends off — the gaps between windows are the market schedule, not missing data, and every settled window in the span is present.
Sources	data_copper.json (3,656 settled KXCOPPER15M rows: ticker, strike, result, expiration_value, sheet_cps) + kalshi_candles_copper.json (Kalshi 1-min candles for all 3,656 tickers). Series settles on Pyth COPPER .
Spot (proxy)	Merged series built exactly as in build_copper_conf_14_6.py: Binance COPPERUSDT perp 1-minute opens, tail minutes filled from Yahoo HG=F (only where Binance has no minute), then Yahoo forward-fill of holes within 60 minutes. 81,107 merged minutes.
Kalshi %	kalshi_candles_copper.json — 1-min candle Yes open at the checkpoint minute, with the same fallback chain as the BTC report (that candle's close, previous candle's close, last earlier close). Flipped to the move side where a "move-side %" is shown. As a probability, $P(\text{up}) = \text{the raw Yes price}$.

Definitions: at minutes-left m , offset = $(15-m) \times 60$ s after window open; **move \$** = spot at open+offset – spot at window open (**self-consistent**, in copper price units, tick 0.001). The move is deliberately NOT spot – strike: the spot proxy (Binance perp / Yahoo HG=F) carries a level basis vs Pyth, the settlement source, and that basis drifts — it cancels in a move measured from the window open, exactly as with Gold. **Result** = Kalshi official result (yes→up); final \$ = expiration_value – strike.

2. Method

- **Accuracy** = fraction of windows where the predicted side (move sign, or $P \geq 0.5$) equals the result. **Brier score** = mean $(P(\text{up}) - \text{outcome})^2$, lower is better; a coin-flip 0.50 forecast scores 0.2500, and a hard 0/1 forecast scores exactly 1 – accuracy.
- **Probit / Brownian bridge**: $P(\text{up}) = \Phi(\text{move} / (\sigma \sqrt{m}))$, m = minutes left. σ = per-minute volatility in copper price units, estimated three ways: (i) **global** — standard deviation of all consecutive 1-minute spot changes inside the train span = **0.0020/min**; (ii) **rolling** — realized vol of the 60 minutes *before* the window opened (per window, no lookahead; requires ≥ 30 consecutive 1-min changes, else the global σ is used); (iii) **in-window** — realized vol of the 1-minute changes from window open up to the checkpoint.
- **Settlement-averaging variant**: modelled as if the settlement value is based on the final minute rather than a single print. For a Brownian path, if the settlement value is the average of the last 60 s, the variance of (settlement average – current price) is $\sigma^2 \cdot (m - 2/3)$ instead of $\sigma^2 \cdot m$ (the last minute contributes $\int_0^1 (1-t)^2 dt = 1/3$, not 1). The variant therefore replaces $\sqrt{m}$ with $\sqrt{(m-2/3)}$: at 1 minute left the effective remaining variance falls to **one third**. This is a modelling

assumption about the averaging window (60 s), stated openly — and for copper it is tested, not assumed (see §9: it fails at 1 minute here).

- Logistic regression:** sklearn is not installed in this environment, so logistic regression was implemented directly (same code as the BTC report): features standardized on train (mean/SD), intercept plus L2 penalty ($\lambda=1.0$ on standardized coefficients), fitted by L-BFGS (scipy). Feature sets are **cumulative** — each set adds one block to the previous one, so every step's delta is that block's marginal value. NaN features are mean-imputed from train.
- "Helps" standard used in the verdicts:** a block HELPS only if it improves holdout Brier by ≥ 0.002 **and** does not reduce holdout accuracy. Anything smaller is called NULL (noise-level), with the exact delta printed so the reader can judge.

3. Sigma (volatility) diagnostics

Global train $\sigma = \mathbf{0.0020}$ per minute ($\sqrt{m}$ scaling: $\sigma\sqrt{5} = 0.0045$ at 5 min left, $\sigma\sqrt{1} = 0.0020$ at 1 min left). Per-window sigma distributions (price units/min):

Sigma estimate	Sample	Median	25th pct	75th pct
Rolling pre-window 60m	all	0.0018	0.0014	0.0023
Rolling pre-window 60m	train	0.0018	0.0013	0.0023
Rolling pre-window 60m	holdout	0.0018	0.0014	0.0023
In-window realized vol (to 5 min left)	all	0.0016	0.0010	0.0022
In-window realized vol (to 4 min left)	all	0.0016	0.0011	0.0022
In-window realized vol (to 3 min left)	all	0.0016	0.0011	0.0022
In-window realized vol (to 2 min left)	all	0.0016	0.0011	0.0022
In-window realized vol (to 1 min left)	all	0.0016	0.0011	0.0022

Reading: per-window sigma varies around the global 0.0020 (rolling pre-60m median ≈ 0.0018), but the spread is proportionally narrower than BTC's, and — as §9 shows — swapping in a per-window σ buys copper far less than it bought BTC.

4. Baseline A — move sign by signed move bucket (FULL sample, all 3,656 windows)

Move-sign accuracy overall by minute (full sample): 5 min: 80.0% · 4 min: 81.2% · 3 min: 82.9% · 2 min: 84.7% · 1 min: 87.5%.

4.1 — 5 minute(s) left, full sample

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate (result opposite move)
≤ -0.015	76	1.3%	98.7%	1.3%
$-0.015..-0.010$	136	2.2%	97.8%	2.2%
$-0.010..-0.0075$	190	3.2%	96.8%	3.2%
$-0.0075..-0.005$	230	7.8%	92.2%	7.8%
$-0.005..-0.002$	576	18.4%	81.6%	18.4%
$-0.002..0$	496	32.5%	67.5%	32.5%
$0..+0.002$	766	61.2%	61.2%	38.8%
$+0.002..+0.005$	570	80.4%	80.4%	19.6%
$+0.005..+0.0075$	228	93.0%	93.0%	7.0%
$+0.0075..+0.010$	174	97.1%	97.1%	2.9%
$+0.010..+0.015$	140	96.4%	96.4%	3.6%
$>+0.015$	74	100.0%	100.0%	0.0%

4.2 — 4 minute(s) left, full sample

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate (result opposite move)
<=-0.015	84	1.2%	98.8%	1.2%
-0.015..-0.010	158	1.9%	98.1%	1.9%
-0.010..-0.0075	201	5.0%	95.0%	5.0%
-0.0075..-0.005	228	6.1%	93.9%	6.1%
-0.005..-0.002	577	15.1%	84.9%	15.1%
-0.002..0	465	35.1%	64.9%	35.1%
0..+0.002	754	61.1%	61.1%	38.9%
+0.002..+0.005	542	83.2%	83.2%	16.8%
+0.005..+0.0075	238	94.5%	94.5%	5.5%
+0.0075..+0.010	167	97.0%	97.0%	3.0%
+0.010..+0.015	147	95.9%	95.9%	4.1%
>+0.015	95	98.9%	98.9%	1.1%

4.3 — 3 minute(s) left, full sample

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate (result opposite move)
<=-0.015	111	1.8%	98.2%	1.8%
-0.015..-0.010	172	0.0%	100.0%	0.0%
-0.010..-0.0075	176	2.3%	97.7%	2.3%
-0.0075..-0.005	259	5.4%	94.6%	5.4%
-0.005..-0.002	554	14.8%	85.2%	14.8%
-0.002..0	443	33.0%	67.0%	33.0%
0..+0.002	711	60.3%	60.3%	39.7%
+0.002..+0.005	542	86.0%	86.0%	14.0%
+0.005..+0.0075	245	95.1%	95.1%	4.9%
+0.0075..+0.010	186	97.8%	97.8%	2.2%
+0.010..+0.015	150	98.0%	98.0%	2.0%
>+0.015	107	100.0%	100.0%	0.0%

4.4 — 2 minute(s) left, full sample

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate (result opposite move)
<=-0.015	112	0.0%	100.0%	0.0%
-0.015..-0.010	177	1.1%	98.9%	1.1%
-0.010..-0.0075	204	2.9%	97.1%	2.9%
-0.0075..-0.005	248	3.6%	96.4%	3.6%
-0.005..-0.002	544	12.5%	87.5%	12.5%
-0.002..0	439	30.5%	69.5%	30.5%
0..+0.002	702	60.4%	60.4%	39.6%
+0.002..+0.005	500	90.4%	90.4%	9.6%
+0.005..+0.0075	252	96.4%	96.4%	3.6%
+0.0075..+0.010	202	98.5%	98.5%	1.5%
+0.010..+0.015	166	99.4%	99.4%	0.6%
>+0.015	110	100.0%	100.0%	0.0%

4.5 — 1 minute(s) left, full sample

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate (result opposite move)
<=-0.015	126	0.0%	100.0%	0.0%
-0.015..-0.010	174	0.6%	99.4%	0.6%
-0.010..-0.0075	206	1.9%	98.1%	1.9%
-0.0075..-0.005	228	1.3%	98.7%	1.3%
-0.005..-0.002	545	6.6%	93.4%	6.6%
-0.002..0	445	27.9%	72.1%	27.9%
0..+0.002	677	64.3%	64.3%	35.7%
+0.002..+0.005	507	91.9%	91.9%	8.1%
+0.005..+0.0075	236	99.2%	99.2%	0.8%
+0.0075..+0.010	221	99.1%	99.1%	0.9%
+0.010..+0.015	164	99.4%	99.4%	0.6%
>+0.015	127	100.0%	100.0%	0.0%

5. Baseline A (holdout) — move buckets on the 1,097 holdout windows only

5.1 — 5 minute(s) left, holdout

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate
<=-0.015	16	0.0%	100.0%	0.0%
-0.015..-0.010	41	2.4%	97.6%	2.4%
-0.010..-0.0075	58	5.2%	94.8%	5.2%
-0.0075..-0.005	63	9.5%	90.5%	9.5%
-0.005..-0.002	192	17.7%	82.3%	17.7%
-0.002..0	154	31.2%	68.8%	31.2%
0..+0.002	226	56.2%	56.2%	43.8%
+0.002..+0.005	173	79.2%	79.2%	20.8%
+0.005..+0.0075	63	90.5%	90.5%	9.5%
+0.0075..+0.010	57	98.2%	98.2%	1.8%
+0.010..+0.015	34	97.1%	97.1%	2.9%
>+0.015	20	100.0%	100.0%	0.0%

5.2 — 4 minute(s) left, holdout

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate
<=-0.015	18	0.0%	100.0%	0.0%
-0.015..-0.010	51	0.0%	100.0%	0.0%
-0.010..-0.0075	58	3.4%	96.6%	3.4%
-0.0075..-0.005	61	9.8%	90.2%	9.8%
-0.005..-0.002	192	14.6%	85.4%	14.6%
-0.002..0	149	34.9%	65.1%	34.9%
0..+0.002	223	56.5%	56.5%	43.5%
+0.002..+0.005	143	80.4%	80.4%	19.6%
+0.005..+0.0075	84	94.0%	94.0%	6.0%
+0.0075..+0.010	53	96.2%	96.2%	3.8%
+0.010..+0.015	40	97.5%	97.5%	2.5%
>+0.015	25	96.0%	96.0%	4.0%

5.3 — 3 minute(s) left, holdout

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate
<=-0.015	26	0.0%	100.0%	0.0%
-0.015..-0.010	52	0.0%	100.0%	0.0%
-0.010..-0.0075	56	1.8%	98.2%	1.8%
-0.0075..-0.005	70	5.7%	94.3%	5.7%
-0.005..-0.002	187	13.9%	86.1%	13.9%
-0.002..0	136	33.1%	66.9%	33.1%
0..+0.002	205	56.6%	56.6%	43.4%
+0.002..+0.005	160	83.1%	83.1%	16.9%
+0.005..+0.0075	74	93.2%	93.2%	6.8%
+0.0075..+0.010	59	96.6%	96.6%	3.4%
+0.010..+0.015	43	97.7%	97.7%	2.3%
>+0.015	29	100.0%	100.0%	0.0%

5.4 — 2 minute(s) left, holdout

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate
<=-0.015	28	0.0%	100.0%	0.0%
-0.015..-0.010	55	0.0%	100.0%	0.0%
-0.010..-0.0075	64	1.6%	98.4%	1.6%
-0.0075..-0.005	68	7.4%	92.6%	7.4%
-0.005..-0.002	178	9.6%	90.4%	9.6%
-0.002..0	126	33.3%	66.7%	33.3%
0..+0.002	213	54.5%	54.5%	45.5%
+0.002..+0.005	150	87.3%	87.3%	12.7%
+0.005..+0.0075	72	95.8%	95.8%	4.2%
+0.0075..+0.010	68	98.5%	98.5%	1.5%
+0.010..+0.015	40	97.5%	97.5%	2.5%
>+0.015	35	100.0%	100.0%	0.0%

5.5 — 1 minute(s) left, holdout

Move bucket	n	Finished UP	Move-sign accuracy	Flip rate
<=-0.015	30	0.0%	100.0%	0.0%
-0.015..-0.010	59	1.7%	98.3%	1.7%
-0.010..-0.0075	61	1.6%	98.4%	1.6%
-0.0075..-0.005	70	2.9%	97.1%	2.9%
-0.005..-0.002	176	4.5%	95.5%	4.5%
-0.002..0	137	32.1%	67.9%	32.1%
0..+0.002	200	61.0%	61.0%	39.0%
+0.002..+0.005	147	89.1%	89.1%	10.9%
+0.005..+0.0075	70	98.6%	98.6%	1.4%
+0.0075..+0.010	65	96.9%	96.9%	3.1%
+0.010..+0.015	43	97.7%	97.7%	2.3%
>+0.015	39	100.0%	100.0%	0.0%

6. Baseline B — Kalshi % alone as a probability: calibration (holdout)

Kalshi % holdout scores: 5 min: acc 79.9%, Brier 0.1302 · 4 min: acc 83.5%, Brier 0.1192 · 3 min: acc 85.2%, Brier 0.1053 · 2 min: acc 88.5%, Brier 0.0858 · 1 min: acc 92.0%, Brier 0.0584. Calibration bins are on P(up) = the Kalshi Yes price itself (not flipped). A perfectly calibrated market has Actual UP ≈ Mean predicted in every bin.

6.1 — 5 minute(s) left, holdout calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.1	203	4.3%	4.4%	+0.1 pts
0.1–0.2	102	14.8%	13.7%	-1.1 pts
0.2–0.3	82	24.8%	17.1%	-7.8 pts
0.3–0.4	82	34.5%	24.4%	-10.1 pts
0.4–0.5	74	44.1%	50.0%	+5.9 pts
0.5–0.6	89	54.2%	43.8%	-10.4 pts
0.6–0.7	86	64.4%	59.3%	-5.1 pts
0.7–0.8	82	74.7%	75.6%	+0.9 pts
0.8–0.9	99	85.1%	86.9%	+1.8 pts
0.9–1.0	198	95.4%	96.0%	+0.5 pts

6.2 — 4 minute(s) left, holdout calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.1	231	3.2%	3.5%	+0.2 pts
0.1–0.2	104	14.4%	13.5%	-0.9 pts
0.2–0.3	72	24.2%	15.3%	-8.9 pts
0.3–0.4	63	33.8%	27.0%	-6.8 pts
0.4–0.5	76	44.6%	34.2%	-10.4 pts
0.5–0.6	73	54.1%	49.3%	-4.8 pts
0.6–0.7	72	64.3%	69.4%	+5.1 pts
0.7–0.8	72	74.4%	66.7%	-7.8 pts
0.8–0.9	83	84.8%	86.7%	+1.9 pts
0.9–1.0	251	96.1%	95.6%	-0.4 pts

6.3 — 3 minute(s) left, holdout calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.1	292	3.2%	3.4%	+0.2 pts
0.1–0.2	82	14.4%	13.4%	-1.0 pts
0.2–0.3	58	24.4%	15.5%	-8.8 pts
0.3–0.4	56	33.6%	32.1%	-1.4 pts
0.4–0.5	71	44.3%	35.2%	-9.1 pts
0.5–0.6	50	54.6%	50.0%	-4.6 pts
0.6–0.7	57	65.1%	63.2%	-1.9 pts
0.7–0.8	69	74.9%	73.9%	-0.9 pts
0.8–0.9	75	84.7%	80.0%	-4.7 pts
0.9–1.0	287	96.6%	96.5%	-0.1 pts

6.4 — 2 minute(s) left, holdout calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.1	345	2.7%	2.6%	-0.1 pts
0.1–0.2	77	13.9%	11.7%	-2.2 pts
0.2–0.3	44	23.5%	18.2%	-5.3 pts
0.3–0.4	42	35.0%	23.8%	-11.2 pts
0.4–0.5	35	44.1%	31.4%	-12.7 pts
0.5–0.6	48	54.7%	47.9%	-6.8 pts
0.6–0.7	43	65.1%	65.1%	-0.0 pts
0.7–0.8	54	75.4%	74.1%	-1.3 pts

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.9	77	84.8%	83.1%	-1.7 pts
0.9–1.0	332	97.1%	96.4%	-0.8 pts

6.5 — 1 minute(s) left, holdout calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.0–0.1	418	2.2%	1.2%	-1.0 pts
0.1–0.2	54	14.0%	7.4%	-6.6 pts
0.2–0.3	23	23.5%	26.1%	+2.6 pts
0.3–0.4	33	34.3%	30.3%	-4.0 pts
0.4–0.5	21	44.8%	28.6%	-16.2 pts
0.5–0.6	30	55.4%	53.3%	-2.0 pts
0.6–0.7	34	64.8%	50.0%	-14.8 pts
0.7–0.8	33	74.4%	75.8%	+1.3 pts
0.8–0.9	53	84.4%	84.9%	+0.5 pts
0.9–1.0	398	97.8%	97.5%	-0.4 pts

7. Baseline B — Kalshi % calibration (FULL sample)

7.1 — 5 minute(s) left, full sample calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	653	4.5%	3.1%	-1.4 pts
0.1–0.2	376	14.4%	13.3%	-1.1 pts
0.2–0.3	273	24.5%	20.9%	-3.6 pts
0.3–0.4	251	34.3%	32.7%	-1.6 pts
0.4–0.5	247	44.2%	45.7%	+1.6 pts
0.5–0.6	260	54.2%	52.3%	-1.9 pts
0.6–0.7	230	64.4%	64.8%	+0.4 pts
0.7–0.8	301	74.4%	76.4%	+2.0 pts
0.8–0.9	387	84.7%	83.7%	-1.0 pts
0.9–1.0	678	95.2%	96.0%	+0.9 pts

7.2 — 4 minute(s) left, full sample calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	782	3.8%	3.3%	-0.5 pts
0.1–0.2	365	14.2%	13.2%	-1.1 pts
0.2–0.3	239	24.2%	21.8%	-2.4 pts
0.3–0.4	225	34.4%	32.0%	-2.4 pts
0.4–0.5	223	44.3%	39.0%	-5.3 pts
0.5–0.6	194	54.6%	58.8%	+4.1 pts
0.6–0.7	224	64.4%	68.3%	+3.9 pts
0.7–0.8	244	74.7%	74.6%	-0.1 pts
0.8–0.9	327	85.1%	84.1%	-1.0 pts
0.9–1.0	833	95.8%	96.4%	+0.6 pts

7.3 — 3 minute(s) left, full sample calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	955	3.4%	3.2%	-0.1 pts

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.1–0.2	295	14.1%	13.6%	-0.5 pts
0.2–0.3	222	24.6%	18.9%	-5.6 pts
0.3–0.4	175	33.6%	34.3%	+0.7 pts
0.4–0.5	186	44.2%	38.2%	-6.0 pts
0.5–0.6	165	54.5%	56.4%	+1.8 pts
0.6–0.7	176	64.6%	65.9%	+1.3 pts
0.7–0.8	205	74.7%	79.0%	+4.4 pts
0.8–0.9	305	85.0%	83.3%	-1.7 pts
0.9–1.0	972	96.4%	97.0%	+0.6 pts

7.4 — 2 minute(s) left, full sample calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	1163	2.9%	3.1%	+0.2 pts
0.1–0.2	242	13.7%	12.0%	-1.7 pts
0.2–0.3	151	23.9%	25.8%	+1.9 pts
0.3–0.4	143	34.6%	33.6%	-1.0 pts
0.4–0.5	119	44.5%	44.5%	+0.1 pts
0.5–0.6	137	54.6%	48.9%	-5.7 pts
0.6–0.7	134	64.9%	66.4%	+1.5 pts
0.7–0.8	168	75.3%	75.6%	+0.3 pts
0.8–0.9	242	84.9%	82.6%	-2.2 pts
0.9–1.0	1157	97.0%	97.1%	+0.2 pts

7.5 — 1 minute(s) left, full sample calibration

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	1396	2.3%	1.9%	-0.4 pts
0.1–0.2	161	13.9%	11.8%	-2.1 pts
0.2–0.3	91	23.8%	27.5%	+3.6 pts
0.3–0.4	92	34.4%	37.0%	+2.6 pts
0.4–0.5	82	44.5%	48.8%	+4.3 pts
0.5–0.6	78	55.1%	55.1%	+0.1 pts
0.6–0.7	93	64.6%	62.4%	-2.2 pts
0.7–0.8	128	74.6%	70.3%	-4.3 pts
0.8–0.9	161	84.5%	79.5%	-5.0 pts
0.9–1.0	1374	97.7%	98.1%	+0.4 pts

8. Complete model comparison — every model, every minute (train and holdout)

Models: **Move sign (hard)** = 0/1 forecast from move sign (Brier = 1 – accuracy). **Move-bucket empirical** = train up-rate inside the window's signed move bucket (Laplace-smoothed), applied to holdout. **Probit** variants per §2. **L0–L6** = cumulative logistic sets (§10). **Pooled** = one logistic fitted on all five minutes stacked, features move + minutes-left + move×minutes.

8.1 — 5 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Move sign (hard)	80.7%	0.1934	78.6%	0.2142	+0.0840
Kalshi % alone	81.7%	0.1282	79.9%	0.1302	+0.0000

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Move-bucket empirical (train-fit)	80.7%	0.1388	78.6%	0.1488	+0.0186
Probit global sigma	80.7%	0.1377	78.6%	0.1435	+0.0132
Probit rolling pre-60m sigma	80.7%	0.1370	78.6%	0.1431	+0.0129
Probit in-window sigma	80.7%	0.1383	78.6%	0.1459	+0.0156
Probit global sigma + settlement-avg	80.7%	0.1372	78.6%	0.1431	+0.0128
Probit rolling sigma + settlement-avg	80.7%	0.1369	78.6%	0.1432	+0.0130
Probit in-window sigma + settlement-avg	80.7%	0.1387	78.6%	0.1465	+0.0163
L0 move only	80.7%	0.1369	78.6%	0.1432	+0.0130
L1 + momentum/path changes (chg 1m, chg 3m)	80.7%	0.1368	79.6%	0.1434	+0.0131
L2 + distance travelled / in-window vol (path, RV, range)	80.9%	0.1364	79.0%	0.1435	+0.0132
L3 + Kalshi % (logit)	82.2%	0.1265	81.3%	0.1284	-0.0019
L4 + Kalshi % change (1m, 3m)	82.2%	0.1262	81.4%	0.1284	-0.0019
L5 + Kalshi volume (log vol, vol change)	82.2%	0.1262	81.4%	0.1284	-0.0019
L6 + pre-window regime (60m return, 60m vol)	82.0%	0.1261	81.0%	0.1288	-0.0014
Pooled logistic move + minutes-left (+move x min)	80.7%	0.1370	78.6%	0.1434	+0.0132

8.2 — 4 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Move sign (hard)	81.8%	0.1817	79.8%	0.2024	+0.0832
Kalshi % alone	84.4%	0.1147	83.5%	0.1192	+0.0000
Move-bucket empirical (train-fit)	81.8%	0.1302	79.8%	0.1390	+0.0198
Probit global sigma	81.8%	0.1278	79.8%	0.1343	+0.0152
Probit rolling pre-60m sigma	81.8%	0.1270	79.8%	0.1339	+0.0148
Probit in-window sigma	81.8%	0.1279	79.8%	0.1363	+0.0171
Probit global sigma + settlement-avg	81.8%	0.1273	79.8%	0.1340	+0.0149
Probit rolling sigma + settlement-avg	81.8%	0.1271	79.8%	0.1343	+0.0151
Probit in-window sigma + settlement-avg	81.8%	0.1287	79.8%	0.1372	+0.0181
L0 move only	81.8%	0.1274	79.8%	0.1346	+0.0154
L1 + momentum/path changes (chg 1m, chg 3m)	81.6%	0.1271	80.6%	0.1349	+0.0157
L2 + distance travelled / in-window vol (path, RV, range)	81.8%	0.1270	80.4%	0.1348	+0.0157
L3 + Kalshi % (logit)	84.4%	0.1138	83.0%	0.1182	-0.0010
L4 + Kalshi % change (1m, 3m)	84.4%	0.1137	83.3%	0.1183	-0.0008
L5 + Kalshi volume (log vol, vol change)	84.4%	0.1137	83.3%	0.1183	-0.0008
L6 + pre-window regime (60m return, 60m vol)	84.4%	0.1136	83.4%	0.1185	-0.0007

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Pooled logistic move + minutes-left (+move x min)	81.8%	0.1270	79.8%	0.1344	+0.0152

8.3 — 3 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Move sign (hard)	83.4%	0.1661	81.8%	0.1823	+0.0770
Kalshi % alone	86.8%	0.0981	85.2%	0.1053	+0.0000
Move-bucket empirical (train-fit)	83.4%	0.1182	81.8%	0.1263	+0.0210
Probit global sigma	83.4%	0.1146	81.8%	0.1236	+0.0183
Probit rolling pre-60m sigma	83.4%	0.1132	81.8%	0.1238	+0.0185
Probit in-window sigma	83.4%	0.1150	81.8%	0.1253	+0.0200
Probit global sigma + settlement-avg	83.4%	0.1143	81.8%	0.1239	+0.0186
Probit rolling sigma + settlement-avg	83.4%	0.1136	81.8%	0.1250	+0.0196
Probit in-window sigma + settlement-avg	83.4%	0.1163	81.8%	0.1274	+0.0220
L0 move only	83.4%	0.1144	81.8%	0.1239	+0.0186
L1 + momentum/path changes (chg 1m, chg 3m)	83.5%	0.1141	82.5%	0.1241	+0.0187
L2 + distance travelled / in-window vol (path, RV, range)	83.7%	0.1140	82.0%	0.1244	+0.0190
L3 + Kalshi % (logit)	86.6%	0.0973	85.9%	0.1038	-0.0016
L4 + Kalshi % change (1m, 3m)	86.5%	0.0973	85.9%	0.1037	-0.0016
L5 + Kalshi volume (log vol, vol change)	86.5%	0.0973	85.9%	0.1037	-0.0016
L6 + pre-window regime (60m return, 60m vol)	86.4%	0.0971	85.8%	0.1039	-0.0014
Pooled logistic move + minutes-left (+move x min)	83.4%	0.1142	81.8%	0.1240	+0.0186

8.4 — 2 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Move sign (hard)	85.5%	0.1454	83.0%	0.1696	+0.0837
Kalshi % alone	87.9%	0.0849	88.5%	0.0858	+0.0000
Move-bucket empirical (train-fit)	85.5%	0.1047	83.0%	0.1147	+0.0289
Probit global sigma	85.5%	0.1020	83.0%	0.1094	+0.0236
Probit rolling pre-60m sigma	85.5%	0.1031	83.0%	0.1111	+0.0253
Probit in-window sigma	85.5%	0.1046	83.0%	0.1139	+0.0281
Probit global sigma + settlement-avg	85.5%	0.1024	83.0%	0.1105	+0.0246
Probit rolling sigma + settlement-avg	85.5%	0.1047	83.0%	0.1134	+0.0276
Probit in-window sigma + settlement-avg	85.5%	0.1072	83.0%	0.1173	+0.0315
L0 move only	85.5%	0.1021	83.0%	0.1100	+0.0242
L1 + momentum/path changes (chg 1m, chg 3m)	85.4%	0.1020	83.0%	0.1101	+0.0243

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
L2 + distance travelled / in-window vol (path, RV, range)	85.6%	0.1019	83.2%	0.1101	+0.0243
L3 + Kalshi % (logit)	87.8%	0.0842	88.9%	0.0842	-0.0017
L4 + Kalshi % change (1m, 3m)	88.4%	0.0840	88.8%	0.0841	-0.0018
L5 + Kalshi volume (log vol, vol change)	88.4%	0.0840	88.8%	0.0841	-0.0018
L6 + pre-window regime (60m return, 60m vol)	88.0%	0.0838	89.1%	0.0840	-0.0018
Pooled logistic move + minutes-left (+move x min)	85.5%	0.1019	83.0%	0.1099	+0.0241

8.5 — 1 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
Move sign (hard)	88.2%	0.1180	86.0%	0.1404	+0.0820
Kalshi % alone	91.2%	0.0612	92.0%	0.0584	+0.0000
Move-bucket empirical (train-fit)	88.2%	0.0863	86.0%	0.0989	+0.0406
Probit global sigma	88.2%	0.0829	86.0%	0.0949	+0.0366
Probit rolling pre-60m sigma	88.2%	0.0839	86.0%	0.0975	+0.0392
Probit in-window sigma	88.2%	0.0856	86.0%	0.1003	+0.0419
Probit global sigma + settlement-avg	88.2%	0.0876	86.0%	0.1026	+0.0442
Probit rolling sigma + settlement-avg	88.2%	0.0895	86.0%	0.1061	+0.0477
Probit in-window sigma + settlement-avg	88.2%	0.0921	86.0%	0.1094	+0.0510
L0 move only	88.2%	0.0828	86.0%	0.0947	+0.0363
L1 + momentum/path changes (chg 1m, chg 3m)	88.4%	0.0820	86.1%	0.0954	+0.0370
L2 + distance travelled / in-window vol (path, RV, range)	88.4%	0.0820	86.0%	0.0954	+0.0370
L3 + Kalshi % (logit)	91.6%	0.0584	92.2%	0.0570	-0.0014
L4 + Kalshi % change (1m, 3m)	91.9%	0.0580	92.1%	0.0582	-0.0001
L5 + Kalshi volume (log vol, vol change)	91.8%	0.0578	92.2%	0.0587	+0.0003
L6 + pre-window regime (60m return, 60m vol)	92.0%	0.0572	91.7%	0.0590	+0.0006
Pooled logistic move + minutes-left (+move x min)	88.2%	0.0833	86.0%	0.0949	+0.0365

9. Probit deep dive — what the Brownian math does and does not do (holdout)

Minutes left	Global σ Brier	Rolling pre-60 σ Brier	In-window σ Brier	Global + settle-avg	Rolling + settle-avg	In-window + settle-avg	Kalshi % Brier (reference)
5	0.1435	0.1431	0.1459	0.1431	0.1432	0.1465	0.1302
4	0.1343	0.1339	0.1363	0.1340	0.1343	0.1372	0.1192
3	0.1236	0.1238	0.1253	0.1239	0.1250	0.1274	0.1053
2	0.1094	0.1111	0.1139	0.1105	0.1134	0.1173	0.0858
1	0.0949	0.0975	0.1003	0.1026	0.1061	0.1094	0.0584

Verdict on the probit family (mean holdout Brier across 5–1 min): global σ 0.1212 $\rightarrow$ rolling σ 0.1219 $\rightarrow$ rolling + settlement-averaging 0.1244, vs Kalshi % 0.0998. For copper the per-window σ swap is worth almost nothing (rolling – global by minute: 5 min -0.0004, 4 min -0.0004, 3 min +0.0002, 2 min +0.0017, 1 min +0.0026; in-window σ is worse than global at most minutes) — copper's per-window vol estimates are noisier relative to signal than BTC's, and the move itself is measured on a proxy. Settlement-averaging is worse still: at 1 minute left it **hurts** every σ choice (global 0.0949 $\rightarrow$ 0.1026; rolling 0.0975 $\rightarrow$ 0.1061), the opposite of BTC. Sharpening the probability around a proxy move is overconfidence when the settlement index (Pyth) can sit a few ticks away from the proxy — the averaging term shrinks the variance the formula *thinks* remains, but the proxy-vs-Pyth basis noise does not shrink with it. And every probit variant is far behind the Kalshi price at every minute — a much bigger gap than BTC's, because the Kalshi market prices the actual settlement source while the probit only sees the proxy move.

Accuracy note: every probit variant makes the **same sign calls as the move** ($P \geq 0.5$ ■ $\text{move} \geq 0$), so probit accuracy equals move-sign accuracy at every minute by construction: 5 min 78.6% · 4 min 79.8% · 3 min 81.8% · 2 min 83.0% · 1 min 86.0% (holdout). Sigma choice changes only the probabilities, never the direction.

10. Logistic regression — cumulative feature additions (holdout)

Each row adds one feature block to the row above. Δ columns are vs the previous set (previous row) — that is the marginal value of that block.

10.1 — 5 minute(s) left

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 move-only	Δ Brier vs Kalshi %
L0 move only	78.6%	0.1432	—	—	+0.0000	+0.0130
L1 + momentum/path changes (chg 1m, chg 3m)	79.6%	0.1434	+1.0 pts	+0.0001	+0.0001	+0.0131
L2 + distance travelled / in-window vol (path, RV, range)	79.0%	0.1435	-0.5 pts	+0.0001	+0.0002	+0.0132
L3 + Kalshi % (logit)	81.3%	0.1284	+2.3 pts	-0.0151	-0.0149	-0.0019
L4 + Kalshi % change (1m, 3m)	81.4%	0.1284	+0.1 pts	+0.0000	-0.0149	-0.0019
L5 + Kalshi volume (log vol, vol change)	81.4%	0.1284	+0.0 pts	+0.0000	-0.0149	-0.0019
L6 + pre-window regime (60m return, 60m vol)	81.0%	0.1288	-0.4 pts	+0.0004	-0.0144	-0.0014

10.2 — 4 minute(s) left

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 move-only	Δ Brier vs Kalshi %
L0 move only	79.8%	0.1346	—	—	+0.0000	+0.0154
L1 + momentum/path changes (chg 1m, chg 3m)	80.6%	0.1349	+0.8 pts	+0.0003	+0.0003	+0.0157
L2 + distance travelled / in-window vol (path, RV, range)	80.4%	0.1348	-0.2 pts	-0.0001	+0.0002	+0.0157
L3 + Kalshi % (logit)	83.0%	0.1182	+2.6 pts	-0.0166	-0.0164	-0.0010
L4 + Kalshi % change (1m, 3m)	83.3%	0.1183	+0.4 pts	+0.0001	-0.0163	-0.0008
L5 + Kalshi volume (log vol, vol change)	83.3%	0.1183	+0.0 pts	+0.0000	-0.0163	-0.0008
L6 + pre-window regime (60m return, 60m vol)	83.4%	0.1185	+0.1 pts	+0.0001	-0.0161	-0.0007

10.3 — 3 minute(s) left

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 move-only	Δ Brier vs Kalshi %
L0 move only	81.8%	0.1239	—	—	+0.0000	+0.0186
L1 + momentum/path changes (chg 1m, chg 3m)	82.5%	0.1241	+0.7 pts	+0.0001	+0.0001	+0.0187
L2 + distance travelled / in-window vol (path, RV, range)	82.0%	0.1244	-0.5 pts	+0.0003	+0.0005	+0.0190
L3 + Kalshi % (logit)	85.9%	0.1038	+3.9 pts	-0.0206	-0.0201	-0.0016
L4 + Kalshi % change (1m, 3m)	85.9%	0.1037	+0.0 pts	-0.0000	-0.0202	-0.0016
L5 + Kalshi volume (log vol, vol change)	85.9%	0.1037	+0.0 pts	+0.0000	-0.0202	-0.0016
L6 + pre-window regime (60m return, 60m vol)	85.8%	0.1039	-0.1 pts	+0.0002	-0.0200	-0.0014

10.4 — 2 minute(s) left

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 move-only	Δ Brier vs Kalshi %
L0 move only	83.0%	0.1100	—	—	+0.0000	+0.0242
L1 + momentum/path changes (chg 1m, chg 3m)	83.0%	0.1101	+0.0 pts	+0.0000	+0.0000	+0.0243
L2 + distance travelled / in-window vol (path, RV, range)	83.2%	0.1101	+0.2 pts	+0.0000	+0.0001	+0.0243
L3 + Kalshi % (logit)	88.9%	0.0842	+5.7 pts	-0.0260	-0.0259	-0.0017
L4 + Kalshi % change (1m, 3m)	88.8%	0.0841	-0.1 pts	-0.0001	-0.0260	-0.0018
L5 + Kalshi volume (log vol, vol change)	88.8%	0.0841	+0.0 pts	+0.0000	-0.0260	-0.0018
L6 + pre-window regime (60m return, 60m vol)	89.1%	0.0840	+0.3 pts	-0.0000	-0.0260	-0.0018

10.5 — 1 minute(s) left

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 move-only	Δ Brier vs Kalshi %
L0 move only	86.0%	0.0947	—	—	+0.0000	+0.0363
L1 + momentum/path changes (chg 1m, chg 3m)	86.1%	0.0954	+0.2 pts	+0.0007	+0.0007	+0.0370
L2 + distance travelled / in-window vol (path, RV, range)	86.0%	0.0954	-0.2 pts	+0.0000	+0.0008	+0.0370
L3 + Kalshi % (logit)	92.2%	0.0570	+6.2 pts	-0.0384	-0.0376	-0.0014
L4 + Kalshi % change (1m, 3m)	92.1%	0.0582	-0.1 pts	+0.0012	-0.0364	-0.0001
L5 + Kalshi volume (log vol, vol change)	92.2%	0.0587	+0.1 pts	+0.0005	-0.0359	+0.0003
L6 + pre-window regime (60m return, 60m vol)	91.7%	0.0590	-0.5 pts	+0.0003	-0.0357	+0.0006

10.6 — Pooled model (move + minutes-left, one logistic for all five minutes)

Minutes left	Pooled holdout acc	Pooled holdout Brier	Per-minute L0 acc	Per-minute L0 Brier
5	78.6%	0.1434	78.6%	0.1432
4	79.8%	0.1344	79.8%	0.1346
3	81.8%	0.1240	81.8%	0.1239
2	83.0%	0.1099	83.0%	0.1100
1	86.0%	0.0949	86.0%	0.0947

Verdict on pooling: NULL. The pooled move+minutes model tracks per-minute L0 almost exactly and never beats it by the ≥ 0.002 standard at any minute.

11. High-confidence subsets — best model at each minute (holdout)

"Best model" = lowest holdout Brier at that minute among all 17 models in §8. (Selecting "best" on the holdout itself is mildly optimistic — treat these accuracies as upper bounds, and compare with the Kalshi-% table in §12, which involves no selection.) Confidence = $\max(P, 1-P)$.

Minutes left	Best model	Holdout Brier	Threshold	n	Coverage (% of holdout)	Accuracy on subset
5	L3 + Kalshi % (logit)	0.1284	$\geq 85\%$	526	47.9%	94.5%
5	L3 + Kalshi % (logit)	0.1284	$\geq 90\%$	424	38.7%	95.8%
5	L3 + Kalshi % (logit)	0.1284	$\geq 95\%$	274	25.0%	98.2%
4	L3 + Kalshi % (logit)	0.1182	$\geq 85\%$	595	54.2%	93.9%
4	L3 + Kalshi % (logit)	0.1182	$\geq 90\%$	500	45.6%	96.2%
4	L3 + Kalshi % (logit)	0.1182	$\geq 95\%$	365	33.3%	97.5%
3	L4 + Kalshi % change (1m, 3m)	0.1037	$\geq 85\%$	673	61.3%	95.2%
3	L4 + Kalshi % change (1m, 3m)	0.1037	$\geq 90\%$	597	54.4%	96.5%
3	L4 + Kalshi % change (1m, 3m)	0.1037	$\geq 95\%$	463	42.2%	98.1%
2	L6 + pre-window regime (60m return, 60m vol)	0.0840	$\geq 85\%$	739	67.4%	96.2%
2	L6 + pre-window regime (60m return, 60m vol)	0.0840	$\geq 90\%$	653	59.5%	97.4%
2	L6 + pre-window regime (60m return, 60m vol)	0.0840	$\geq 95\%$	532	48.5%	98.1%
1	L3 + Kalshi % (logit)	0.0570	$\geq 85\%$	865	78.9%	98.2%
1	L3 + Kalshi % (logit)	0.0570	$\geq 90\%$	814	74.2%	98.3%
1	L3 + Kalshi % (logit)	0.0570	$\geq 95\%$	726	66.2%	99.2%

12. High-confidence subsets — comparators (holdout): Kalshi % alone and L3 (move+path+Kalshi %)

Kalshi % alone

Minutes left	Threshold	n	Coverage	Accuracy on subset
5	$\geq 85\%$	516	47.0%	94.8%
5	$\geq 90\%$	408	37.2%	95.3%
5	$\geq 95\%$	269	24.5%	97.8%
4	$\geq 85\%$	593	54.1%	94.4%
4	$\geq 90\%$	492	44.8%	95.7%
4	$\geq 95\%$	368	33.5%	97.3%
3	$\geq 85\%$	668	60.9%	95.4%

Minutes left	Threshold	n	Coverage	Accuracy on subset
3	≥90%	591	53.9%	96.6%
3	≥95%	452	41.2%	98.2%
2	≥85%	774	70.6%	95.7%
2	≥90%	688	62.7%	96.7%
2	≥95%	568	51.8%	98.2%
1	≥85%	882	80.4%	97.5%
1	≥90%	822	74.9%	98.2%
1	≥95%	745	67.9%	98.7%

L3 + Kalshi % (logit)

Minutes left	Threshold	n	Coverage	Accuracy on subset
5	≥85%	526	47.9%	94.5%
5	≥90%	424	38.7%	95.8%
5	≥95%	274	25.0%	98.2%
4	≥85%	595	54.2%	93.9%
4	≥90%	500	45.6%	96.2%
4	≥95%	365	33.3%	97.5%
3	≥85%	670	61.1%	95.4%
3	≥90%	596	54.3%	96.5%
3	≥95%	466	42.5%	98.3%
2	≥85%	748	68.2%	96.0%
2	≥90%	653	59.5%	97.5%
2	≥95%	528	48.1%	98.5%
1	≥85%	865	78.9%	98.2%
1	≥90%	814	74.2%	98.3%
1	≥95%	726	66.2%	99.2%

13. Flip risk — P(result opposite to the move sign)

Overall flip rate = 1 – move-sign accuracy. Full sample: 5 min: 20.0% · 4 min: 18.8% · 3 min: 17.1% · 2 min: 15.3% · 1 min: 12.5%. Holdout: 5 min: 21.4% · 4 min: 20.2% · 3 min: 18.2% · 2 min: 17.0% · 1 min: 14.0%.

13.1 Flip rate by signed move bucket × minutes left (FULL sample, %)

Move bucket	5 min	4 min	3 min	2 min	1 min	n at 5 min
<=-0.015	1.3	1.2	1.8	0.0	0.0	76
-0.015..-0.010	2.2	1.9	0.0	1.1	0.6	136
-0.010..-0.0075	3.2	5.0	2.3	2.9	1.9	190
-0.0075..-0.005	7.8	6.1	5.4	3.6	1.3	230
-0.005..-0.002	18.4	15.1	14.8	12.5	6.6	576
-0.002..0	32.5	35.1	33.0	30.5	27.9	496
0..+0.002	38.8	38.9	39.7	39.6	35.7	766
+0.002..+0.005	19.6	16.8	14.0	9.6	8.1	570
+0.005..+0.0075	7.0	5.5	4.9	3.6	0.8	228
+0.0075..+0.010	2.9	3.0	2.2	1.5	0.9	174
+0.010..+0.015	3.6	4.1	2.0	0.6	0.6	140
>+0.015	0.0	1.1	0.0	0.0	0.0	74

13.2 Flip rate by signed move bucket × minutes left (HOLDOUT, %)

Move bucket	5 min	4 min	3 min	2 min	1 min	n at 5 min (holdout)
<=-0.015	0.0	0.0	0.0	0.0	0.0	16
-0.015..-0.010	2.4	0.0	0.0	0.0	1.7	41
-0.010..-0.0075	5.2	3.4	1.8	1.6	1.6	58
-0.0075..-0.005	9.5	9.8	5.7	7.4	2.9	63
-0.005..-0.002	17.7	14.6	13.9	9.6	4.5	192
-0.002..0	31.2	34.9	33.1	33.3	32.1	154
0..+0.002	43.8	43.5	43.4	45.5	39.0	226
+0.002..+0.005	20.8	19.6	16.9	12.7	10.9	173
+0.005..+0.0075	9.5	6.0	6.8	4.2	1.4	63
+0.0075..+0.010	1.8	3.8	3.4	1.5	3.1	57
+0.010..+0.015	2.9	2.5	2.3	2.5	2.3	34
>+0.015	0.0	4.0	0.0	0.0	0.0	20

13.3 Flip rate by |move| tier (FULL sample, %)

	move	tier	5 min	4 min	3 min	2 min	1 min
0-0.002	36.3 (n=1262)	37.4 (n=1219)	37.1 (n=1154)	36.1 (n=1141)	32.6 (n=1122)		
0.002-0.005	19.0 (n=1146)	15.9 (n=1119)	14.4 (n=1096)	11.1 (n=1044)	7.3 (n=1052)		
0.005-0.0075	7.4 (n=458)	5.8 (n=466)	5.2 (n=504)	3.6 (n=500)	1.1 (n=464)		
0.0075-0.010	3.0 (n=364)	4.1 (n=368)	2.2 (n=362)	2.2 (n=406)	1.4 (n=427)		
0.010-0.015	2.9 (n=276)	3.0 (n=305)	0.9 (n=322)	0.9 (n=343)	0.6 (n=338)		
0.015+	0.7 (n=150)	1.1 (n=179)	0.9 (n=218)	0.0 (n=222)	0.0 (n=253)		

Reading the flip tables: as with BTC, flip risk is a function of move SIZE first and the clock second — but copper's floor is higher. A sub-0.002 move flips 36.3% of the time at 5 minutes left and still 32.6% at 1 minute left; even a 0.002–0.005 move flips 7.3% at 1 minute. By |move| ≥0.005 the last minute is close to settled (≤1.4% flip), and at ≥0.010 it is effectively settled (0.6% flip at 1 min, 0.0% for 0.015+). There is **no BTC-style 2→1 minute flip collapse**: BTC's \$25–50 tier fell 13.0% → 3.3% and its \$50–75 tier 5.8% → 0.3% between 2 and 1 minutes left, while copper's mid tiers fall much less (0.005–0.0075: 3.6% → 1.1%; 0.0075–0.010: 2.2% → 1.4%). That missing collapse is consistent with §9: the late-window safety BTC gets from settlement averaging is eaten for copper by proxy-vs-Pyth basis noise, which does not average away.

14. Plain verdicts — what helps, what is NULL

Feature / model idea	Verdict	Evidence (holdout)
Move sign (direction)	Real, but the weakest leg for copper	Accuracy 78.6% at 5 min → 86.0% at 1 min — and unlike BTC, the Kalshi price's own sign beats it at every minute, by 6.0 pts at 1 min. The proxy move alone leaves real accuracy on the table late.
Kalshi % as probability	Best single probability source at every minute	Brier 5 min 0.1302, 4 min 0.1192, 3 min 0.1053, 2 min 0.0858, 1 min 0.0584. Calibration (\$6/\$7) is good in the big bins. The only models that edge it are the L3-family logistics that contain it, and by less than the 0.002 'helps' bar at each minute (5 min -0.0019, 4 min -0.0010, 3 min -0.0016, 2 min -0.0017, 1 min -0.0014) — a consistent but sub-threshold margin, paired with real accuracy gains at 5–3 min.
Kalshi % added to logistic (L3 vs L2)	HELPS the logistic — by importing the market	Brier deltas L3–L2: 5 min -0.0151, 4 min -0.0166, 3 min -0.0206, 2 min -0.0260, 1 min -0.0384. By far the largest single-block gain anywhere in this report. After L3, the logistic sits essentially on top of the market price with the move as a small correction.
Rolling / in-window σ in probit (vs global σ)	NULL for copper (helped BTC)	Rolling–global Brier: 5 min -0.0004, 4 min -0.0004, 3 min +0.0002, 2 min +0.0017, 1 min +0.0026; in-window–global: 5 min +0.0024, 4 min +0.0019, 3 min +0.0017, 2 min +0.0045, 1 min +0.0054. All inside noise, mixed signs. Per-window vol does not rescale the copper probit usefully.

Feature / model idea	Verdict	Evidence (holdout)
Settlement-averaging (probit variant)	HURTS for copper, most at 1 min (helped BTC)	Rolling+avg – rolling Brier: 5 min +0.0001, 4 min +0.0004, 3 min +0.0012, 2 min +0.0023, 1 min +0.0086. The $\sigma^2(m-2/3)$ sharpening over-trusts a proxy move whose basis vs Pyth does not shrink in the final minute. Do not carry the BTC settlement-averaging adjustment over to copper.
Momentum / path (move change 1m & 3m, L1)	NULL	L1–L0 Brier: 5 min +0.0001, 4 min +0.0003, 3 min +0.0001, 2 min +0.0000, 1 min +0.0007. Accuracy ticks up at 5 min but Brier does not move — the probabilities are not better.
Distance travelled / in-window RV / range (L2)	NULL	L2–L1 Brier: 5 min +0.0001, 4 min -0.0001, 3 min +0.0003, 2 min +0.0000, 1 min +0.0000.
Kalshi % change, 1m & 3m (L4)	NULL	L4–L3 Brier: 5 min +0.0000, 4 min +0.0001, 3 min -0.0000, 2 min -0.0001, 1 min +0.0012.
Kalshi candle volume / volume change (L5)	NULL	L5–L4 Brier: 5 min +0.0000, 4 min +0.0000, 3 min +0.0000, 2 min +0.0000, 1 min +0.0005 — zero to four ten-thousandths.
Pre-window 60m return + vol regime (L6)	NULL	L6–L5 Brier: 5 min +0.0004, 4 min +0.0001, 3 min +0.0002, 2 min -0.0000, 1 min +0.0003.
Pooled move + minutes-left logistic	NULL	Never beats per-minute L0 by the ≥ 0.002 standard (\$10.6).
Move-bucket empirical probabilities	Weakest probability fallback	Brier 5 min 0.1488, 4 min 0.1390, 3 min 0.1263, 2 min 0.1147, 1 min 0.0989 — worse than even the global- σ probit at most minutes, because the move buckets inherit the full proxy noise with no market information.

Bottom line for the last 5 minutes (copper)

- For copper, the market is the model.** Kalshi % is the best single probability at every minute (0.1302 → 0.0584 Brier, 5→1 min), and its sign beats the proxy move's sign at every minute — by a wide late margin (88.5% vs 83.0% at 2 min; 92.0% vs 86.0% at 1 min). Kalshi prices Pyth, the settlement source; our move prices a Binance/Yahoo proxy.
- The one combination worth keeping is move + Kalshi % (the L3 logistic).** It edges Kalshi alone at all five minutes on Brier (by 0.0010–0.0018 — under the 0.002 bar, so call it marginal, not proven) and beats it more clearly on accuracy at 5–3 min (81.3% vs 79.9% at 5 min). Everything added after Kalshi % — its changes, volume, pre-window regime — is null.
- Do not import BTC's probit upgrades.** Per-window σ is null for copper and settlement-averaging actively hurts at 1 minute. The plain global- σ probit ($\sigma = 0.0020$) is the best of a weak family here, and the whole family trails the market by 0.013–0.037 Brier.
- Confidence lives in |move| size + market agreement.** At $\geq 90\%$ model confidence the holdout accuracy in \$11–12 is materially higher, but copper needs a genuinely big move to earn it: a sub-0.002 move is still a ~33% flip at 1 minute left, against $\sigma\sqrt{5} \approx 0.0045$ of remaining 5-minute variance. Small copper moves late are NOT the near-locks that small BTC leads become.
- Flip risk falls with the clock inside a bucket, but bucket dominates and the floor is high** (\$13). Size any last-5-minutes copper rule by |move| bucket first, minute second — and defer to the Kalshi price when the two disagree (\$15).

15. Copper vs BTC, and the basis-noise test

BTC reference numbers are the holdout results from the BTC last-5-minutes report on the same methodology (6,482 windows, 70/30 chronological holdout): lead-sign accuracy 81.0/83.0/86.3/88.6/95.3% and Kalshi % Brier 0.1303/0.1146/0.0940/0.0739/0.0343 at 5/4/3/2/1 minutes left.

Minutes left	Copper move-sign acc (holdout)	BTC lead-sign acc (holdout)	Gap	Copper Kalshi % acc	BTC Kalshi % acc	Copper Kalshi Brier	BTC Kalshi Brier
5	78.6%	81.0%	-2.4 pts	79.9%	81.2%	0.1302	0.1303
4	79.8%	83.0%	-3.2 pts	83.5%	83.3%	0.1192	0.1146
3	81.8%	86.3%	-4.5 pts	85.2%	86.2%	0.1053	0.0940
2	83.0%	88.6%	-5.6 pts	88.5%	89.3%	0.0858	0.0739

Minutes left	Copper move-sign acc (holdout)	BTC lead-sign acc (holdout)	Gap	Copper Kalshi % acc	BTC Kalshi % acc	Copper Kalshi Brier	BTC Kalshi Brier
1	86.0%	95.3%	-9.3 pts	92.0%	95.5%	0.0584	0.0343

On the FULL sample, copper move-sign accuracy at the sheet checkpoints is 81.2% at 4 min and 84.7% at 2 min, vs BTC's 83.6% / 88.5% — the same softness seen in the earlier sheet agreement figures (copper 81.2% / 84.7% on the sheet's own basis). The gap is small at 5 minutes and **widens sharply into the close** on the move sign (−2.4 pts at 5 min → −9.3 pts at 1 min, holdout), while the Kalshi-price gap stays narrow (copper Kalshi Brier is essentially BTC's at 5–3 min and only clearly worse at 2–1 min). In other words: copper spot really does move almost like BTC into the close — what degrades is our *proxy measurement* of it against the Pyth settlement, and the market price, which tracks Pyth, does not suffer from it.

15.1 Accuracy by |move| size — the basis-noise split (holdout)

Minutes left		move	<0.002: n	acc	0.002–0.005: n	acc		move	≥0.005: n	acc
5	380	61.3%	365	80.8%	352	94.9%				
4	372	59.9%	335	83.3%	390	95.6%				
3	341	60.7%	347	84.7%	409	96.8%				
2	339	59.0%	328	89.0%	430	97.4%				
1	337	63.8%	323	92.6%	437	98.2%				

Same split, FULL sample:

Minutes left		move	<0.002: n	acc	0.002–0.005: n	acc		move	≥0.005: n	acc
5	1262	63.7%	1146	81.0%	1248	95.7%				
4	1219	62.6%	1119	84.1%	1318	96.0%				
3	1154	62.9%	1096	85.6%	1406	97.2%				
2	1141	63.9%	1044	88.9%	1471	98.0%				
1	1122	67.4%	1052	92.7%	1482	99.1%				

15.2 When the Kalshi price and the proxy move disagree, who is right? (holdout)

Disagreement = sign of (Kalshi Yes − 0.5) ≠ sign of the self-consistent move. The Kalshi market prices Pyth (the settlement source); the move is measured on the Binance/Yahoo proxy.

Minutes left	Disagreements (holdout)	Kalshi side right	Move side right	Disagreements (full)	Kalshi right (full)
5	145	55.2%	44.8%	396	55.3%
4	129	65.9%	34.1%	381	64.0%
3	126	65.1%	34.9%	376	66.8%
2	134	72.4%	27.6%	356	67.1%
1	132	75.0%	25.0%	380	68.9%

Reading §15: the basis-noise explanation holds up. (i) Copper's deficit vs BTC is concentrated exactly where it should be if the proxy is the problem — in the move sign, late, and in small moves: sub-0.002 moves are barely better than a coin flip even at 1 minute, while ≥0.005 moves score in the 90s. (ii) When market and move disagree, the market side wins clearly at every minute — the opposite pattern from a market that is merely echoing spot. (iii) The Kalshi-price quality gap vs BTC is small, so the underlying copper market is not dramatically less predictable — the *measurement* is. Practical consequence: for copper in the last 5 minutes, weight the Kalshi % over the move sign whenever they conflict, and treat any sub-0.002 "lead" as no lead at all.

16. Limitations & honesty notes

- Spot proxy: the self-consistent move removes the proxy's *level* basis vs Pyth, but not its *change* noise over 1–14 minutes (Binance perp microstructure, Yahoo HG=F staleness in the tail, forward-filled minutes). This is the dominant error source in every move-based number above and is why the market price outperforms here more than for BTC

(Coinbase vs CF Benchmarks is a much tighter basis).

- Yahoo tail: windows after Binance Vision coverage ends (Oct 19) use HG=F futures minutes with forward-fills up to 60 min; their in-window realized-vol and path features are coarser. They sit mostly in the holdout, which if anything makes holdout move-based scores conservative.
- Kalshi % is the 1-minute candle **open** (per the established sheet definition), not a trade-by-trade mid; intra-minute information is not in these files. Early-series candles are sometimes sparse (some candles missing entirely); the fallback chain recovered a price for every checkpoint at 5–1 min (0 missing).
- Settlement-averaging assumes a 60-second averaging window and Brownian increments; it is presented as a model. For copper the data reject it at 1 minute — possibly because Pyth's own update/aggregation differs, possibly because basis noise dominates; either way the empirical answer is not to use it here.
- Copper trades weekdays with daily breaks: the window *sequence* is chronological but wall-clock gaps vary, so "pre-window 60m" features span different real time after breaks. The 70/30 split is by window order, as in the BTC report.
- Holdout is a single chronological split (last 30%, Oct 4 → Oct 21 PT). Market conditions in the holdout are one regime; Brier differences under ~ 0.002 should be read as noise — several copper margins (L3 vs Kalshi, rolling vs global σ) sit exactly in that band and are labelled accordingly.
- "Best model" in §11 was selected on the same holdout it is scored on; §12 exists precisely to give the no-selection comparison.

Generated locally from the files listed in §1. Scripts: `analyze_last5_copper.py` (+ `make_last5_copper_report.py`). No trades placed; no live jobs or crons touched.