

Kalshi BTC 15-Minute — The Last 5 Minutes

Volume 2 — Local Mathematics: every model tested on all 6,482 windows since launch

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

Dataset: Aug 13, 2026 (series launch) → Oct 20, 2026, 11:30 PM PT · Up 3,247 / Down 3,235

Sources: Kalshi markets + 1-minute candlesticks, Coinbase BTC-USD 1-minute candles (local data only)

Kalshi BTC 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)	6,482
Up / Down	3,247 / 3,235
First window	2026-08-13T23:45:00Z UTC (Aug 13, 4:45 PM PT) — series launch
Last window	2026-10-21T06:30:00Z UTC (Oct 20, 11:30 PM PT)
Train (first 70%)	4,537 windows, 2026-08-13T23:45:00Z → 2026-09-30T18:15:00Z UTC
Holdout (last 30%)	1,945 windows, 2026-09-30T18:30:00Z → 2026-10-21T06:30:00Z UTC
Sources	data_btcfull.json (6,427) + data_btc_ext_2026-10-20.json (23) + 32 rows reconstructed from Kalshi_BTC_15m_EveryMinute.xlsx (beat/result/final \$) = 6,482. Ticker generator validated 6,450/6,450 against the JSON tickers; Kalshi candles found for 6,482/6,482 tickers.
Spot	candles_btc_full.json — Coinbase BTC-USD 1-minute opens, 98,455 minutes. Exactly 1 missing minute in the whole span (Oct 2, 16:47 UTC); it was forward-filled only for path/volatility features (it is not itself a 5–1-minute checkpoint).
Kalshi %	kalshi_candles_btc_full.json — 1-min candle Yes open at the checkpoint minute (fallbacks: that candle's close, previous candle's close), flipped to the lead side where a "lead-side %" is shown. As a probability, $P(\text{up})$ = the raw Yes price.

Definitions: at minutes-left m , offset = $(15-m) \times 60$ s after window open; **lead \$** = Coinbase open at open+offset – Kalshi strike (floor_strike); result = Kalshi official result (yes→up).

2. Method

- **Accuracy** = fraction of windows where the predicted side (lead 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**: $P(\text{up}) = \Phi(\text{lead} / (\sigma \cdot \sqrt{m}))$, m = minutes left. σ = per-minute \$ volatility, estimated three ways: (i) **global** — standard deviation of all 1-minute Coinbase \$ changes inside the train span = **\$42.08/min**; (ii) **rolling** — realized vol of the 60 minutes *before* the window opened (per window, no lookahead); (iii) **in-window** — realized vol of the 1-minute changes from window open up to the checkpoint.
- **Settlement-averaging variant**: Kalshi settles on an average over 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.
- **Logistic regression**: sklearn is not installed in this environment, so logistic regression was implemented directly: 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 (e.g. pre-window vol for the first windows) 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 σ = **\$42.08 per minute** ($\sqrt{t}$ scaling: $\sigma\sqrt{5}$ = \$94 at 5 min left, $\sigma\sqrt{1}$ = \$42 at 1 min left). Per-window sigma distributions:

Sigma estimate	Sample	Median \$/min	25th pct	75th pct
Rolling pre-window 60m	all	30.30	20.18	41.14
Rolling pre-window 60m	train	30.93	20.63	42.37
Rolling pre-window 60m	holdout	28.88	19.29	38.55
In-window realized vol (to 5 min left)	all	26.49	17.13	39.63
In-window realized vol (to 4 min left)	all	26.76	17.26	39.51
In-window realized vol (to 3 min left)	all	26.89	17.43	39.61
In-window realized vol (to 2 min left)	all	27.04	17.48	39.60
In-window realized vol (to 1 min left)	all	27.22	17.68	39.58

Reading: per-window sigma varies a lot around the global \$42 (rolling pre-60m median $\approx$ \$30, interquartile roughly \$20–\$41), which is why the rolling probit beats the global probit below — a single fixed σ is mis-scaled in both quiet and wild windows.

4. Baseline A — lead sign by signed lead bucket (FULL sample, all 6482 windows)

Lead-sign accuracy overall by minute (full sample): 5 min: 81.3% · 4 min: 83.6% · 3 min: 85.6% · 2 min: 88.5% · 1 min: 94.0%.

4.1 — 5 minute(s) left, full sample

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
≤ -200	288	2.8%	97.2%	2.8%
-200..-150	244	4.5%	95.5%	4.5%
-150..-100	426	8.2%	91.8%	8.2%
-100..-75	353	8.8%	91.2%	8.8%
-75..-50	463	15.6%	84.4%	15.6%
-50..-25	700	24.4%	75.6%	24.4%
-25..-10	504	31.7%	68.3%	31.7%
-10..0	366	47.5%	52.5%	47.5%
0..+10	300	58.3%	58.3%	41.7%
+10..+25	515	69.9%	69.9%	30.1%
+25..+50	637	75.4%	75.4%	24.6%
+50..+75	439	85.6%	85.6%	14.4%
+75..+100	342	93.3%	93.3%	6.7%
+100..+150	387	95.1%	95.1%	4.9%
+150..+200	215	95.3%	95.3%	4.7%
$> +200$	303	99.7%	99.7%	0.3%

4.2 — 4 minute(s) left, full sample

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
≤ -200	316	0.9%	99.1%	0.9%
-200..-150	233	1.3%	98.7%	1.3%
-150..-100	444	5.0%	95.0%	5.0%
-100..-75	379	7.7%	92.3%	7.7%

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
-75..-50	455	14.1%	85.9%	14.1%
-50..-25	674	22.1%	77.9%	22.1%
-25..-10	461	31.5%	68.5%	31.5%
-10..0	348	44.3%	55.7%	44.3%
0..+10	336	58.0%	58.0%	42.0%
+10..+25	484	73.1%	73.1%	26.9%
+25..+50	604	78.0%	78.0%	22.0%
+50..+75	430	89.5%	89.5%	10.5%
+75..+100	331	92.4%	92.4%	7.6%
+100..+150	439	96.6%	96.6%	3.4%
+150..+200	235	97.9%	97.9%	2.1%
>+200	313	100.0%	100.0%	0.0%

4.3 — 3 minute(s) left, full sample

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
<=-200	344	0.9%	99.1%	0.9%
-200..-150	222	0.9%	99.1%	0.9%
-150..-100	462	3.2%	96.8%	3.2%
-100..-75	343	3.8%	96.2%	3.8%
-75..-50	480	8.8%	91.2%	8.8%
-50..-25	671	18.3%	81.7%	18.3%
-25..-10	451	35.3%	64.7%	35.3%
-10..0	359	44.0%	56.0%	44.0%
0..+10	304	59.5%	59.5%	40.5%
+10..+25	449	71.5%	71.5%	28.5%
+25..+50	588	83.5%	83.5%	16.5%
+50..+75	437	89.9%	89.9%	10.1%
+75..+100	363	95.6%	95.6%	4.4%
+100..+150	427	98.1%	98.1%	1.9%
+150..+200	240	99.2%	99.2%	0.8%
>+200	342	100.0%	100.0%	0.0%

4.4 — 2 minute(s) left, full sample

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
<=-200	359	0.6%	99.4%	0.6%
-200..-150	239	0.0%	100.0%	0.0%
-150..-100	433	1.6%	98.4%	1.6%
-100..-75	379	4.0%	96.0%	4.0%
-75..-50	473	6.1%	93.9%	6.1%
-50..-25	657	13.4%	86.6%	13.4%
-25..-10	463	25.5%	74.5%	25.5%
-10..0	323	49.5%	50.5%	49.5%
0..+10	333	63.7%	63.7%	36.3%
+10..+25	403	76.9%	76.9%	23.1%
+25..+50	605	87.4%	87.4%	12.6%
+50..+75	427	94.6%	94.6%	5.4%
+75..+100	334	97.0%	97.0%	3.0%
+100..+150	431	98.8%	98.8%	1.2%
+150..+200	241	100.0%	100.0%	0.0%

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
>+200	382	100.0%	100.0%	0.0%

4.5 — 1 minute(s) left, full sample

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate (result opposite lead)
<=-200	377	0.0%	100.0%	0.0%
-200..-150	241	0.0%	100.0%	0.0%
-150..-100	419	0.2%	99.8%	0.2%
-100..-75	352	0.3%	99.7%	0.3%
-75..-50	461	0.4%	99.6%	0.4%
-50..-25	669	4.6%	95.4%	4.6%
-25..-10	473	16.7%	83.3%	16.7%
-10..0	349	37.8%	62.2%	37.8%
0..+10	344	73.5%	73.5%	26.5%
+10..+25	425	90.6%	90.6%	9.4%
+25..+50	531	98.5%	98.5%	1.5%
+50..+75	448	99.8%	99.8%	0.2%
+75..+100	311	100.0%	100.0%	0.0%
+100..+150	422	100.0%	100.0%	0.0%
+150..+200	265	100.0%	100.0%	0.0%
>+200	395	100.0%	100.0%	0.0%

5. Baseline A (holdout) — lead buckets on the 1,945 holdout windows only

5.1 — 5 minute(s) left, holdout

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
<=-200	68	0.0%	100.0%	0.0%
-200..-150	71	2.8%	97.2%	2.8%
-150..-100	133	7.5%	92.5%	7.5%
-100..-75	95	8.4%	91.6%	8.4%
-75..-50	136	12.5%	87.5%	12.5%
-50..-25	215	22.3%	77.7%	22.3%
-25..-10	149	30.9%	69.1%	30.9%
-10..0	101	47.5%	52.5%	47.5%
0..+10	103	56.3%	56.3%	43.7%
+10..+25	177	66.1%	66.1%	33.9%
+25..+50	201	80.1%	80.1%	19.9%
+50..+75	156	82.1%	82.1%	17.9%
+75..+100	99	94.9%	94.9%	5.1%
+100..+150	114	93.0%	93.0%	7.0%
+150..+200	49	91.8%	91.8%	8.2%
>+200	78	100.0%	100.0%	0.0%

5.2 — 4 minute(s) left, holdout

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
<=-200	83	0.0%	100.0%	0.0%
-200..-150	67	1.5%	98.5%	1.5%
-150..-100	134	4.5%	95.5%	4.5%
-100..-75	101	3.0%	97.0%	3.0%

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
-75..-50	135	11.1%	88.9%	11.1%
-50..-25	225	20.4%	79.6%	20.4%
-25..-10	124	34.7%	65.3%	34.7%
-10..0	99	46.5%	53.5%	46.5%
0..+10	99	57.6%	57.6%	42.4%
+10..+25	169	70.4%	70.4%	29.6%
+25..+50	194	74.7%	74.7%	25.3%
+50..+75	152	86.8%	86.8%	13.2%
+75..+100	104	96.2%	96.2%	3.8%
+100..+150	117	95.7%	95.7%	4.3%
+150..+200	58	98.3%	98.3%	1.7%
>+200	84	100.0%	100.0%	0.0%

5.3 — 3 minute(s) left, holdout

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
<=-200	97	1.0%	99.0%	1.0%
-200..-150	56	1.8%	98.2%	1.8%
-150..-100	133	0.8%	99.2%	0.8%
-100..-75	104	2.9%	97.1%	2.9%
-75..-50	147	9.5%	90.5%	9.5%
-50..-25	212	16.0%	84.0%	16.0%
-25..-10	136	31.6%	68.4%	31.6%
-10..0	107	40.2%	59.8%	40.2%
0..+10	97	66.0%	66.0%	34.0%
+10..+25	147	66.7%	66.7%	33.3%
+25..+50	197	87.8%	87.8%	12.2%
+50..+75	131	89.3%	89.3%	10.7%
+75..+100	118	96.6%	96.6%	3.4%
+100..+150	123	97.6%	97.6%	2.4%
+150..+200	48	100.0%	100.0%	0.0%
>+200	92	100.0%	100.0%	0.0%

5.4 — 2 minute(s) left, holdout

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
<=-200	103	1.0%	99.0%	1.0%
-200..-150	60	0.0%	100.0%	0.0%
-150..-100	122	0.8%	99.2%	0.8%
-100..-75	115	2.6%	97.4%	2.6%
-75..-50	148	5.4%	94.6%	5.4%
-50..-25	195	10.3%	89.7%	10.3%
-25..-10	138	23.9%	76.1%	23.9%
-10..0	101	45.5%	54.5%	45.5%
0..+10	109	59.6%	59.6%	40.4%
+10..+25	139	71.2%	71.2%	28.8%
+25..+50	193	92.2%	92.2%	7.8%
+50..+75	135	95.6%	95.6%	4.4%
+75..+100	115	97.4%	97.4%	2.6%
+100..+150	123	99.2%	99.2%	0.8%
+150..+200	54	100.0%	100.0%	0.0%

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
>+200	95	100.0%	100.0%	0.0%

5.5 — 1 minute(s) left, holdout

Lead bucket \$	n	Finished UP	Lead-sign accuracy	Flip rate
<=-200	106	0.0%	100.0%	0.0%
-200..-150	64	0.0%	100.0%	0.0%
-150..-100	114	0.0%	100.0%	0.0%
-100..-75	109	0.0%	100.0%	0.0%
-75..-50	128	0.0%	100.0%	0.0%
-50..-25	216	3.2%	96.8%	3.2%
-25..-10	150	12.0%	88.0%	12.0%
-10..0	99	24.2%	75.8%	24.2%
0..+10	111	74.8%	74.8%	25.2%
+10..+25	134	91.8%	91.8%	8.2%
+25..+50	171	98.2%	98.2%	1.8%
+50..+75	153	100.0%	100.0%	0.0%
+75..+100	101	100.0%	100.0%	0.0%
+100..+150	128	100.0%	100.0%	0.0%
+150..+200	66	100.0%	100.0%	0.0%
>+200	95	100.0%	100.0%	0.0%

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

Kalshi % holdout scores: 5 min: acc 81.2%, Brier 0.1303 · 4 min: acc 83.3%, Brier 0.1146 · 3 min: acc 86.2%, Brier 0.0940 · 2 min: acc 89.3%, Brier 0.0739 · 1 min: acc 95.5%, Brier 0.0343. 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	397	4.4%	4.0%	-0.4 pts
0.1–0.2	182	14.8%	15.9%	+1.1 pts
0.2–0.3	138	24.5%	26.8%	+2.4 pts
0.3–0.4	112	34.2%	30.4%	-3.9 pts
0.4–0.5	113	44.7%	42.5%	-2.2 pts
0.5–0.6	111	54.5%	52.3%	-2.3 pts
0.6–0.7	127	64.7%	59.8%	-4.9 pts
0.7–0.8	152	74.5%	73.7%	-0.8 pts
0.8–0.9	194	84.6%	82.0%	-2.7 pts
0.9–1.0	419	95.5%	94.7%	-0.7 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	452	3.5%	2.9%	-0.7 pts
0.1–0.2	169	14.5%	11.2%	-3.2 pts
0.2–0.3	134	24.7%	28.4%	+3.6 pts
0.3–0.4	111	34.4%	30.6%	-3.8 pts
0.4–0.5	88	44.6%	52.3%	+7.7 pts
0.5–0.6	75	54.5%	53.3%	-1.1 pts
0.6–0.7	116	64.6%	60.3%	-4.3 pts

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap (actual – predicted)
0.7–0.8	133	74.5%	69.2%	-5.4 pts
0.8–0.9	171	84.6%	77.2%	-7.4 pts
0.9–1.0	496	96.3%	97.2%	+0.9 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	549	2.9%	2.2%	-0.7 pts
0.1–0.2	145	14.2%	13.1%	-1.1 pts
0.2–0.3	99	24.4%	26.3%	+1.9 pts
0.3–0.4	95	34.2%	36.8%	+2.7 pts
0.4–0.5	74	44.1%	45.9%	+1.9 pts
0.5–0.6	72	54.5%	50.0%	-4.5 pts
0.6–0.7	82	64.1%	59.8%	-4.3 pts
0.7–0.8	106	74.6%	64.2%	-10.4 pts
0.8–0.9	135	85.0%	83.7%	-1.3 pts
0.9–1.0	588	97.1%	97.6%	+0.5 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	666	2.1%	2.4%	+0.3 pts
0.1–0.2	102	14.2%	14.7%	+0.5 pts
0.2–0.3	71	24.2%	21.1%	-3.1 pts
0.3–0.4	57	33.5%	33.3%	-0.2 pts
0.4–0.5	67	44.6%	46.3%	+1.7 pts
0.5–0.6	65	54.3%	49.2%	-5.1 pts
0.6–0.7	56	64.6%	64.3%	-0.3 pts
0.7–0.8	67	74.6%	64.2%	-10.4 pts
0.8–0.9	87	85.3%	79.3%	-6.0 pts
0.9–1.0	707	97.8%	97.6%	-0.2 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	803	1.0%	1.4%	+0.4 pts
0.1–0.2	71	13.9%	11.3%	-2.7 pts
0.2–0.3	35	24.6%	8.6%	-16.1 pts
0.3–0.4	41	35.2%	22.0%	-13.2 pts
0.4–0.5	24	44.0%	41.7%	-2.3 pts
0.5–0.6	26	54.3%	61.5%	+7.2 pts
0.6–0.7	30	64.4%	46.7%	-17.7 pts
0.7–0.8	29	74.8%	72.4%	-2.4 pts
0.8–0.9	50	84.7%	90.0%	+5.3 pts
0.9–1.0	836	99.0%	99.2%	+0.1 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	1221	4.7%	4.3%	-0.3 pts
0.1–0.2	677	14.5%	14.3%	-0.2 pts
0.2–0.3	484	24.6%	24.8%	+0.2 pts
0.3–0.4	426	34.5%	35.7%	+1.2 pts
0.4–0.5	371	44.5%	41.5%	-3.0 pts
0.5–0.6	375	54.6%	51.5%	-3.1 pts
0.6–0.7	415	65.0%	60.2%	-4.8 pts
0.7–0.8	496	74.5%	73.8%	-0.7 pts
0.8–0.9	631	84.6%	85.4%	+0.8 pts
0.9–1.0	1386	95.3%	95.5%	+0.2 pts

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

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	1497	4.0%	3.2%	-0.8 pts
0.1–0.2	581	14.4%	13.8%	-0.6 pts
0.2–0.3	423	24.2%	27.0%	+2.7 pts
0.3–0.4	367	34.3%	33.2%	-1.1 pts
0.4–0.5	298	44.5%	42.3%	-2.2 pts
0.5–0.6	316	54.3%	53.8%	-0.5 pts
0.6–0.7	364	64.6%	61.0%	-3.6 pts
0.7–0.8	415	74.7%	71.8%	-2.9 pts
0.8–0.9	532	84.8%	81.2%	-3.6 pts
0.9–1.0	1689	95.9%	96.8%	+0.9 pts

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

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	1828	3.1%	2.7%	-0.5 pts
0.1–0.2	472	14.3%	13.6%	-0.8 pts
0.2–0.3	336	24.4%	28.6%	+4.2 pts
0.3–0.4	291	34.3%	40.9%	+6.5 pts
0.4–0.5	253	44.4%	40.7%	-3.7 pts
0.5–0.6	266	54.6%	51.1%	-3.5 pts
0.6–0.7	258	64.7%	58.1%	-6.6 pts
0.7–0.8	321	74.5%	68.5%	-5.9 pts
0.8–0.9	434	85.2%	80.9%	-4.4 pts
0.9–1.0	2023	96.9%	96.8%	-0.0 pts

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

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	2222	2.3%	2.8%	+0.5 pts
0.1–0.2	342	14.2%	18.1%	+3.9 pts
0.2–0.3	246	24.7%	21.1%	-3.5 pts
0.3–0.4	196	34.0%	37.8%	+3.8 pts
0.4–0.5	193	44.4%	44.6%	+0.1 pts
0.5–0.6	191	54.1%	55.0%	+0.8 pts
0.6–0.7	207	64.4%	62.8%	-1.6 pts
0.7–0.8	214	74.6%	67.8%	-6.8 pts
0.8–0.9	316	84.9%	76.9%	-8.0 pts

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.9–1.0	2355	97.9%	97.2%	-0.7 pts

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

Kalshi P(up) bin	n	Mean predicted	Actual UP	Gap
0.0–0.1	2669	1.1%	1.3%	+0.2 pts
0.1–0.2	198	14.1%	11.6%	-2.5 pts
0.2–0.3	146	24.2%	21.9%	-2.2 pts
0.3–0.4	120	34.7%	30.8%	-3.8 pts
0.4–0.5	95	44.3%	46.3%	+2.0 pts
0.5–0.6	118	54.6%	50.8%	-3.8 pts
0.6–0.7	98	64.5%	57.1%	-7.4 pts
0.7–0.8	131	74.9%	72.5%	-2.4 pts
0.8–0.9	172	84.9%	91.3%	+6.4 pts
0.9–1.0	2735	99.0%	99.0%	-0.0 pts

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

Models: **Lead sign (hard)** = 0/1 forecast from lead sign (Brier = 1 – accuracy). **Lead-bucket empirical** = train up-rate inside the window's signed lead 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 lead + minutes-left + lead×minutes.

8.1 — 5 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Lead sign (hard)	81.4%	0.1865	81.0%	0.1897	+0.0594
Kalshi % alone	81.4%	0.1283	81.2%	0.1303	+0.0000
Lead-bucket empirical (train-fit)	81.4%	0.1338	81.0%	0.1371	+0.0067
Probit global sigma	81.4%	0.1402	81.0%	0.1441	+0.0138
Probit rolling pre-60m sigma	81.4%	0.1322	81.0%	0.1354	+0.0051
Probit in-window sigma	81.4%	0.1327	81.0%	0.1372	+0.0069
Probit global sigma + settlement-avg	81.4%	0.1388	81.0%	0.1425	+0.0122
Probit rolling sigma + settlement-avg	81.4%	0.1318	81.0%	0.1350	+0.0047
Probit in-window sigma + settlement-avg	81.4%	0.1329	81.0%	0.1374	+0.0071
L0 lead only	81.4%	0.1353	81.0%	0.1383	+0.0080
L1 + momentum/path changes (chg 1m, chg 3m)	81.4%	0.1352	81.2%	0.1383	+0.0080
L2 + distance travelled / in-window vol (path, RV, range)	81.0%	0.1352	80.9%	0.1384	+0.0081
L3 + Kalshi % (logit)	81.7%	0.1280	81.4%	0.1305	+0.0002
L4 + Kalshi % change (1m, 3m)	81.6%	0.1278	81.1%	0.1311	+0.0007
L5 + Kalshi volume (log vol, vol change)	81.5%	0.1278	81.2%	0.1312	+0.0009
L6 + pre-window regime (60m return, 60m vol)	81.5%	0.1278	81.2%	0.1312	+0.0009
Pooled logistic lead + minutes-left (+lead x min)	81.4%	0.1357	81.0%	0.1389	+0.0086

8.2 — 4 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
Lead sign (hard)	83.9%	0.1613	83.0%	0.1702	+0.0556
Kalshi % alone	84.0%	0.1132	83.3%	0.1146	+0.0000
Lead-bucket empirical (train-fit)	83.9%	0.1182	83.0%	0.1222	+0.0077
Probit global sigma	83.9%	0.1255	83.0%	0.1285	+0.0139
Probit rolling pre-60m sigma	83.9%	0.1177	83.0%	0.1199	+0.0053
Probit in-window sigma	83.9%	0.1176	83.0%	0.1195	+0.0050
Probit global sigma + settlement-avg	83.9%	0.1237	83.0%	0.1264	+0.0118
Probit rolling sigma + settlement-avg	83.9%	0.1172	83.0%	0.1196	+0.0050
Probit in-window sigma + settlement-avg	83.9%	0.1178	83.0%	0.1198	+0.0052
L0 lead only	83.8%	0.1196	82.9%	0.1221	+0.0076
L1 + momentum/path changes (chg 1m, chg 3m)	84.0%	0.1195	82.8%	0.1220	+0.0075
L2 + distance travelled / in-window vol (path, RV, range)	84.1%	0.1193	82.8%	0.1223	+0.0077
L3 + Kalshi % (logit)	84.2%	0.1128	83.1%	0.1147	+0.0001
L4 + Kalshi % change (1m, 3m)	84.0%	0.1127	83.0%	0.1147	+0.0002
L5 + Kalshi volume (log vol, vol change)	83.9%	0.1127	83.1%	0.1149	+0.0003
L6 + pre-window regime (60m return, 60m vol)	83.9%	0.1127	83.1%	0.1149	+0.0003
Pooled logistic lead + minutes-left (+lead x min)	83.8%	0.1191	82.9%	0.1216	+0.0071

8.3 — 3 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
Lead sign (hard)	85.3%	0.1468	86.3%	0.1373	+0.0433
Kalshi % alone	85.7%	0.0985	86.2%	0.0940	+0.0000
Lead-bucket empirical (train-fit)	85.3%	0.1043	86.3%	0.1032	+0.0092
Probit global sigma	85.3%	0.1099	86.3%	0.1111	+0.0171
Probit rolling pre-60m sigma	85.3%	0.1041	86.3%	0.1011	+0.0071
Probit in-window sigma	85.3%	0.1040	86.3%	0.1000	+0.0060
Probit global sigma + settlement-avg	85.3%	0.1075	86.3%	0.1079	+0.0139
Probit rolling sigma + settlement-avg	85.3%	0.1037	86.3%	0.1002	+0.0062
Probit in-window sigma + settlement-avg	85.3%	0.1044	86.3%	0.0999	+0.0059
L0 lead only	85.3%	0.1043	86.4%	0.1033	+0.0093
L1 + momentum/path changes (chg 1m, chg 3m)	85.4%	0.1042	86.3%	0.1035	+0.0095
L2 + distance travelled / in-window vol (path, RV, range)	85.3%	0.1041	86.0%	0.1035	+0.0095
L3 + Kalshi % (logit)	85.9%	0.0979	86.1%	0.0942	+0.0002

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
L4 + Kalshi % change (1m, 3m)	85.7%	0.0976	86.3%	0.0949	+0.0009
L5 + Kalshi volume (log vol, vol change)	85.8%	0.0975	86.1%	0.0952	+0.0012
L6 + pre-window regime (60m return, 60m vol)	85.8%	0.0975	86.1%	0.0952	+0.0012
Pooled logistic lead + minutes-left (+lead x min)	85.3%	0.1039	86.2%	0.1023	+0.0083

8.4 — 2 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
Lead sign (hard)	88.4%	0.1159	88.6%	0.1136	+0.0397
Kalshi % alone	89.0%	0.0802	89.3%	0.0739	+0.0000
Lead-bucket empirical (train-fit)	88.5%	0.0850	88.2%	0.0834	+0.0095
Probit global sigma	88.4%	0.0926	88.6%	0.0917	+0.0178
Probit rolling pre-60m sigma	88.4%	0.0861	88.6%	0.0828	+0.0089
Probit in-window sigma	88.4%	0.0864	88.6%	0.0824	+0.0085
Probit global sigma + settlement-avg	88.4%	0.0890	88.6%	0.0868	+0.0129
Probit rolling sigma + settlement-avg	88.4%	0.0856	88.6%	0.0816	+0.0077
Probit in-window sigma + settlement-avg	88.4%	0.0871	88.6%	0.0823	+0.0084
L0 lead only	88.6%	0.0863	88.9%	0.0835	+0.0096
L1 + momentum/path changes (chg 1m, chg 3m)	88.5%	0.0863	88.7%	0.0832	+0.0093
L2 + distance travelled / in-window vol (path, RV, range)	88.4%	0.0862	88.7%	0.0831	+0.0091
L3 + Kalshi % (logit)	89.0%	0.0797	89.3%	0.0736	-0.0004
L4 + Kalshi % change (1m, 3m)	89.0%	0.0795	89.1%	0.0742	+0.0003
L5 + Kalshi volume (log vol, vol change)	89.0%	0.0792	89.2%	0.0746	+0.0007
L6 + pre-window regime (60m return, 60m vol)	89.0%	0.0792	89.2%	0.0746	+0.0007
Pooled logistic lead + minutes-left (+lead x min)	88.5%	0.0857	88.9%	0.0823	+0.0084

8.5 — 1 minute(s) left

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	Δ Brier vs Kalshi (holdout)
Lead sign (hard)	93.5%	0.0650	95.3%	0.0468	+0.0125
Kalshi % alone	94.2%	0.0405	95.5%	0.0343	+0.0000
Lead-bucket empirical (train-fit)	93.5%	0.0475	95.3%	0.0408	+0.0065
Probit global sigma	93.5%	0.0593	95.3%	0.0576	+0.0233
Probit rolling pre-60m sigma	93.5%	0.0515	95.3%	0.0455	+0.0113
Probit in-window sigma	93.5%	0.0502	95.3%	0.0434	+0.0092
Probit global sigma + settlement-avg	93.5%	0.0492	95.3%	0.0446	+0.0104

Model	Train acc	Train Brier	Holdout acc	Holdout Brier	ΔBrier vs Kalshi (holdout)
Probit rolling sigma + settlement-avg	93.5%	0.0485	95.3%	0.0398	+0.0056
Probit in-window sigma + settlement-avg	93.5%	0.0490	95.3%	0.0396	+0.0053
L0 lead only	93.8%	0.0473	94.9%	0.0439	+0.0096
L1 + momentum/path changes (chg 1m, chg 3m)	93.7%	0.0473	94.8%	0.0441	+0.0098
L2 + distance travelled / in-window vol (path, RV, range)	93.7%	0.0471	94.8%	0.0439	+0.0097
L3 + Kalshi % (logit)	94.2%	0.0405	95.5%	0.0343	+0.0000
L4 + Kalshi % change (1m, 3m)	94.4%	0.0402	95.6%	0.0339	-0.0003
L5 + Kalshi volume (log vol, vol change)	94.4%	0.0402	95.7%	0.0343	+0.0000
L6 + pre-window regime (60m return, 60m vol)	94.4%	0.0402	95.7%	0.0343	+0.0000
Pooled logistic lead + minutes-left (+lead x min)	93.7%	0.0510	95.0%	0.0486	+0.0143

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	0.1441	0.1354	0.1372	0.1425	0.1350	0.1374	0.1303
4	0.1285	0.1199	0.1195	0.1264	0.1196	0.1198	0.1146
3	0.1111	0.1011	0.1000	0.1079	0.1002	0.0999	0.0940
2	0.0917	0.0828	0.0824	0.0868	0.0816	0.0823	0.0739
1	0.0576	0.0455	0.0434	0.0446	0.0398	0.0396	0.0343

Verdict on the probit family (mean holdout Brier across 5–1 min): global σ 0.1066 → rolling σ 0.0969 → rolling + settlement-averaging 0.0953, vs Kalshi % 0.0894. Per-window sigma (rolling or in-window) is a REAL improvement over one global σ at every minute. Settlement-averaging HELPS the probit too — small at 5–2 min, largest at 1 min where the effective variance drops to a third (rolling: 0.0455 → 0.0398) — exactly the direction the averaging math predicts. But even the best probit is WORSE than simply using the Kalshi price at every minute: the market price already contains the Brownian arithmetic plus order-flow/settlement information the formula cannot see. The probit is a good sanity model and a fallback when no market price exists — it is not an upgrade on the market.

Accuracy note: every probit variant makes the **same sign calls as the lead** ($P \geq 0.5 \blacksquare \text{lead} \geq 0$), so probit accuracy equals lead-sign accuracy at every minute by construction: 5 min 81.0% · 4 min 83.0% · 3 min 86.3% · 2 min 88.6% · 1 min 95.3% (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	Δacc vs previous	ΔBrier vs previous	ΔBrier vs L0 lead-only	ΔBrier vs Kalshi %
Lead sign (hard)	81.0%	0.1897	—	—	+0.0514	+0.0594
Lead-bucket empirical (train-fit)	81.0%	0.1371	+0.0 pts	-0.0527	-0.0012	+0.0067
L0 lead only	81.0%	0.1383	-0.1 pts	+0.0012	+0.0000	+0.0080

Feature set (cumulative)	Holdout acc	Holdout Brier	Δ acc vs previous	Δ Brier vs previous	Δ Brier vs L0 lead-only	Δ Brier vs Kalshi %
L1 + momentum/path changes (chg 1m, chg 3m)	81.2%	0.1383	+0.2 pts	+0.0000	+0.0000	+0.0080
L2 + distance travelled / in-window vol (path, RV, range)	80.9%	0.1384	-0.3 pts	+0.0001	+0.0001	+0.0081
L3 + Kalshi % (logit)	81.4%	0.1305	+0.5 pts	-0.0079	-0.0078	+0.0002
L4 + Kalshi % change (1m, 3m)	81.1%	0.1311	-0.3 pts	+0.0006	-0.0072	+0.0007
L5 + Kalshi volume (log vol, vol change)	81.2%	0.1312	+0.2 pts	+0.0002	-0.0070	+0.0009
L6 + pre-window regime (60m return, 60m vol)	81.2%	0.1312	+0.0 pts	+0.0000	-0.0070	+0.0009

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 lead-only	Δ Brier vs Kalshi %
Lead sign (hard)	83.0%	0.1702	—	—	+0.0480	+0.0556
Lead-bucket empirical (train-fit)	83.0%	0.1222	+0.0 pts	-0.0479	+0.0001	+0.0077
L0 lead only	82.9%	0.1221	-0.1 pts	-0.0001	+0.0000	+0.0076
L1 + momentum/path changes (chg 1m, chg 3m)	82.8%	0.1220	-0.1 pts	-0.0001	-0.0001	+0.0075
L2 + distance travelled / in-window vol (path, RV, range)	82.8%	0.1223	+0.0 pts	+0.0003	+0.0002	+0.0077
L3 + Kalshi % (logit)	83.1%	0.1147	+0.3 pts	-0.0076	-0.0074	+0.0001
L4 + Kalshi % change (1m, 3m)	83.0%	0.1147	-0.1 pts	+0.0000	-0.0074	+0.0002
L5 + Kalshi volume (log vol, vol change)	83.1%	0.1149	+0.1 pts	+0.0002	-0.0072	+0.0003
L6 + pre-window regime (60m return, 60m vol)	83.1%	0.1149	+0.0 pts	+0.0000	-0.0072	+0.0003

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 lead-only	Δ Brier vs Kalshi %
Lead sign (hard)	86.3%	0.1373	—	—	+0.0340	+0.0433
Lead-bucket empirical (train-fit)	86.3%	0.1032	+0.0 pts	-0.0341	-0.0001	+0.0092
L0 lead only	86.4%	0.1033	+0.2 pts	+0.0001	+0.0000	+0.0093
L1 + momentum/path changes (chg 1m, chg 3m)	86.3%	0.1035	-0.2 pts	+0.0002	+0.0002	+0.0095
L2 + distance travelled / in-window vol (path, RV, range)	86.0%	0.1035	-0.3 pts	-0.0000	+0.0002	+0.0095
L3 + Kalshi % (logit)	86.1%	0.0942	+0.1 pts	-0.0093	-0.0091	+0.0002
L4 + Kalshi % change (1m, 3m)	86.3%	0.0949	+0.2 pts	+0.0007	-0.0084	+0.0009
L5 + Kalshi volume (log vol, vol change)	86.1%	0.0952	-0.2 pts	+0.0003	-0.0081	+0.0012
L6 + pre-window regime (60m return, 60m vol)	86.1%	0.0952	+0.0 pts	+0.0000	-0.0081	+0.0012

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 lead-only	Δ Brier vs Kalshi %
Lead sign (hard)	88.6%	0.1136	—	—	+0.0301	+0.0397
Lead-bucket empirical (train-fit)	88.2%	0.0834	-0.5 pts	-0.0302	-0.0001	+0.0095
L0 lead only	88.9%	0.0835	+0.7 pts	+0.0001	+0.0000	+0.0096
L1 + momentum/path changes (chg 1m, chg 3m)	88.7%	0.0832	-0.2 pts	-0.0003	-0.0003	+0.0093
L2 + distance travelled / in-window vol (path, RV, range)	88.7%	0.0831	+0.0 pts	-0.0001	-0.0004	+0.0091
L3 + Kalshi % (logit)	89.3%	0.0736	+0.6 pts	-0.0095	-0.0099	-0.0004
L4 + Kalshi % change (1m, 3m)	89.1%	0.0742	-0.2 pts	+0.0006	-0.0093	+0.0003
L5 + Kalshi volume (log vol, vol change)	89.2%	0.0746	+0.1 pts	+0.0004	-0.0089	+0.0007
L6 + pre-window regime (60m return, 60m vol)	89.2%	0.0746	+0.0 pts	+0.0000	-0.0089	+0.0007

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 lead-only	Δ Brier vs Kalshi %
Lead sign (hard)	95.3%	0.0468	—	—	+0.0029	+0.0125
Lead-bucket empirical (train-fit)	95.3%	0.0408	+0.0 pts	-0.0060	-0.0031	+0.0065
L0 lead only	94.9%	0.0439	-0.5 pts	+0.0031	+0.0000	+0.0096
L1 + momentum/path changes (chg 1m, chg 3m)	94.8%	0.0441	-0.1 pts	+0.0002	+0.0002	+0.0098
L2 + distance travelled / in-window vol (path, RV, range)	94.8%	0.0439	+0.0 pts	-0.0001	+0.0000	+0.0097
L3 + Kalshi % (logit)	95.5%	0.0343	+0.8 pts	-0.0097	-0.0096	+0.0000
L4 + Kalshi % change (1m, 3m)	95.6%	0.0339	+0.1 pts	-0.0003	-0.0100	-0.0003
L5 + Kalshi volume (log vol, vol change)	95.7%	0.0343	+0.1 pts	+0.0004	-0.0096	+0.0000
L6 + pre-window regime (60m return, 60m vol)	95.7%	0.0343	+0.0 pts	+0.0000	-0.0096	+0.0000

10.6 — Pooled model (lead + 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	81.0%	0.1389	81.0%	0.1383
4	82.9%	0.1216	82.9%	0.1221
3	86.2%	0.1023	86.4%	0.1033
2	88.9%	0.0823	88.9%	0.0835
1	95.0%	0.0486	94.9%	0.0439

Verdict on pooling: NULL-to-negative. The pooled lead+minutes model never beats the per-minute lead-only logistic on Brier at any minute — minutes-left is already fully handled by fitting each minute separately, and pooling just blurs the per-minute lead scaling.

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	Kalshi % alone	0.1303	≥85%	1020	52.4%	93.4%
5	Kalshi % alone	0.1303	≥90%	830	42.7%	95.3%
5	Kalshi % alone	0.1303	≥95%	491	25.2%	97.8%
4	Kalshi % alone	0.1146	≥85%	1137	58.5%	95.4%
4	Kalshi % alone	0.1146	≥90%	960	49.4%	97.1%
4	Kalshi % alone	0.1146	≥95%	669	34.4%	98.5%
3	Kalshi % alone	0.0940	≥85%	1311	67.4%	96.6%
3	Kalshi % alone	0.0940	≥90%	1148	59.0%	97.5%
3	Kalshi % alone	0.0940	≥95%	890	45.8%	98.2%
2	L3 + Kalshi % (logit)	0.0736	≥85%	1402	72.1%	97.4%
2	L3 + Kalshi % (logit)	0.0736	≥90%	1263	64.9%	98.4%
2	L3 + Kalshi % (logit)	0.0736	≥95%	1035	53.2%	98.9%
1	L4 + Kalshi % change (1m, 3m)	0.0339	≥85%	1718	88.3%	98.5%
1	L4 + Kalshi % change (1m, 3m)	0.0339	≥90%	1632	83.9%	98.9%
1	L4 + Kalshi % change (1m, 3m)	0.0339	≥95%	1523	78.3%	99.3%

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

Kalshi % alone

Minutes left	Threshold	n	Coverage	Accuracy on subset
5	≥85%	1020	52.4%	93.4%
5	≥90%	830	42.7%	95.3%
5	≥95%	491	25.2%	97.8%
4	≥85%	1137	58.5%	95.4%
4	≥90%	960	49.4%	97.1%
4	≥95%	669	34.4%	98.5%
3	≥85%	1311	67.4%	96.6%
3	≥90%	1148	59.0%	97.5%
3	≥95%	890	45.8%	98.2%
2	≥85%	1498	77.0%	96.4%
2	≥90%	1379	70.9%	97.6%
2	≥95%	1162	59.7%	98.7%
1	≥85%	1713	88.1%	98.5%
1	≥90%	1644	84.5%	98.9%
1	≥95%	1541	79.2%	99.3%

L3 + Kalshi % (logit)

Minutes left	Threshold	n	Coverage	Accuracy on subset
5	≥85%	1005	51.7%	93.3%
5	≥90%	809	41.6%	95.2%
5	≥95%	514	26.4%	97.3%

Minutes left	Threshold	n	Coverage	Accuracy on subset
4	≥85%	1118	57.5%	95.8%
4	≥90%	944	48.5%	97.1%
4	≥95%	666	34.2%	98.6%
3	≥85%	1246	64.1%	97.0%
3	≥90%	1073	55.2%	97.8%
3	≥95%	834	42.9%	98.4%
2	≥85%	1402	72.1%	97.4%
2	≥90%	1263	64.9%	98.4%
2	≥95%	1035	53.2%	98.9%
1	≥85%	1689	86.8%	98.6%
1	≥90%	1610	82.8%	98.9%
1	≥95%	1510	77.6%	99.4%

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

Overall flip rate = 1 – lead-sign accuracy. Full sample: 5 min: 18.7% · 4 min: 16.4% · 3 min: 14.4% · 2 min: 11.5% · 1 min: 6.0%. Holdout: 5 min: 19.0% · 4 min: 17.0% · 3 min: 13.7% · 2 min: 11.4% · 1 min: 4.7%.

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

Lead bucket \$	5 min	4 min	3 min	2 min	1 min	n at 5 min
<=-200	2.8	0.9	0.9	0.6	0.0	288
-200..-150	4.5	1.3	0.9	0.0	0.0	244
-150..-100	8.2	5.0	3.2	1.6	0.2	426
-100..-75	8.8	7.7	3.8	4.0	0.3	353
-75..-50	15.6	14.1	8.8	6.1	0.4	463
-50..-25	24.4	22.1	18.3	13.4	4.6	700
-25..-10	31.7	31.5	35.3	25.5	16.7	504
-10..0	47.5	44.3	44.0	49.5	37.8	366
0..+10	41.7	42.0	40.5	36.3	26.5	300
+10..+25	30.1	26.9	28.5	23.1	9.4	515
+25..+50	24.6	22.0	16.5	12.6	1.5	637
+50..+75	14.4	10.5	10.1	5.4	0.2	439
+75..+100	6.7	7.6	4.4	3.0	0.0	342
+100..+150	4.9	3.4	1.9	1.2	0.0	387
+150..+200	4.7	2.1	0.8	0.0	0.0	215
>+200	0.3	0.0	0.0	0.0	0.0	303

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

Lead bucket \$	5 min	4 min	3 min	2 min	1 min	n at 5 min (holdout)
<=-200	0.0	0.0	1.0	1.0	0.0	68
-200..-150	2.8	1.5	1.8	0.0	0.0	71
-150..-100	7.5	4.5	0.8	0.8	0.0	133
-100..-75	8.4	3.0	2.9	2.6	0.0	95
-75..-50	12.5	11.1	9.5	5.4	0.0	136
-50..-25	22.3	20.4	16.0	10.3	3.2	215
-25..-10	30.9	34.7	31.6	23.9	12.0	149
-10..0	47.5	46.5	40.2	45.5	24.2	101
0..+10	43.7	42.4	34.0	40.4	25.2	103
+10..+25	33.9	29.6	33.3	28.8	8.2	177

Lead bucket \$	5 min	4 min	3 min	2 min	1 min	n at 5 min (holdout)
+25..+50	19.9	25.3	12.2	7.8	1.8	201
+50..+75	17.9	13.2	10.7	4.4	0.0	156
+75..+100	5.1	3.8	3.4	2.6	0.0	99
+100..+150	7.0	4.3	2.4	0.8	0.0	114
+150..+200	8.2	1.7	0.0	0.0	0.0	49
>+200	0.0	0.0	0.0	0.0	0.0	78

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

	lead	tier	5 min	4 min	3 min	2 min	1 min
\$0–10	44.9 (n=666)	43.1 (n=684)	42.4 (n=663)	42.8 (n=656)	32.2 (n=692)		
\$10–25	30.9 (n=1019)	29.1 (n=945)	31.9 (n=900)	24.4 (n=866)	13.1 (n=898)		
\$25–50	24.5 (n=1337)	22.1 (n=1278)	17.5 (n=1257)	13.0 (n=1262)	3.3 (n=1201)		
\$50–75	15.0 (n=902)	12.3 (n=885)	9.4 (n=919)	5.8 (n=900)	0.3 (n=909)		
\$75–100	7.8 (n=695)	7.6 (n=710)	4.1 (n=706)	3.5 (n=713)	0.2 (n=663)		
\$100–150	6.6 (n=813)	4.2 (n=882)	2.6 (n=889)	1.4 (n=864)	0.1 (n=841)		
\$150–200	4.6 (n=459)	1.7 (n=469)	0.9 (n=462)	0.0 (n=480)	0.0 (n=506)		
\$200+	1.5 (n=591)	0.5 (n=629)	0.4 (n=686)	0.3 (n=741)	0.0 (n=772)		

Reading the flip tables: flip risk is a function of lead SIZE, not of the clock alone. A sub-\$10 lead flips roughly 4 times in 10 even at 5 minutes left and still ~1 time in 3 at 1 minute left (32.2%); a \$75+ lead at 5 minutes left flips under 1 time in 10, and a \$150+ lead at 2–1 minutes left is nearly settled (flip ■2%). The 5→1 minute improvement is real but second-order next to lead size: inside the small-lead tiers the flip rate falls only modestly (a sub-\$10 lead still flips 32.2% of the time at 1 minute left), while for leads of \$25+ the final minute is dramatically safer than the minute before it (e.g. \$25–50: 13.0% flip at 2 min → 3.3% at 1 min; \$50–75: 5.8% → 0.3%) — the signature of settlement averaging over the final minute, which a lead at the start of that minute only has to survive on average, not tick-by-tick.

14. Plain verdicts — what helps, what is NULL

Feature / model idea	Verdict	Evidence (holdout)
Lead sign (direction)	Real, and it is the whole directional signal	Accuracy 81.0% at 5 min → 95.3% at 1 min. Nothing below changes direction decisions for the better.
Kalshi % as probability	Best single probability source at (nearly) every minute	Brier 5 min 0.1303, 4 min 0.1146, 3 min 0.0940, 2 min 0.0739, 1 min 0.0343. Well calibrated in \$6/\$7 (most bins within a few points). Only beaten trivially twice: L3 at 2 min by 0.0003 and L4 at 1 min by 0.0004 — noise, not an edge.
Kalshi % added to logistic (L3 vs L2)	HELPS the logistic — by importing the market	Brier deltas L3–L2: 5 min -0.0079, 4 min -0.0076, 3 min -0.0093, 2 min -0.0095, 1 min -0.0097. After L3 the logistic = Kalshi alone (±0.0005): the model adds nothing on top of the market price, it just learns to copy it.
Rolling / in-window σ in probit (vs global σ)	HELPS (within the probit family)	Rolling–global Brier: 5 min -0.0087, 4 min -0.0086, 3 min -0.0100, 2 min -0.0089, 1 min -0.0120. In-window σ is about as good. Still loses to Kalshi % at every minute.
Settlement-averaging (probit variant)	HELPS the probit, most at 1 min	Rolling+avg – rolling Brier: 5 min -0.0004, 4 min -0.0003, 3 min -0.0009, 2 min -0.0012, 1 min -0.0057. Direction of effect matches the $\sigma^2(m-2/3)$ math. Does not close the gap to the market price.
Momentum / path (lead change 1m & 3m, L1)	NULL	L1–L0 Brier: 5 min +0.0000, 4 min -0.0001, 3 min +0.0002, 2 min -0.0003, 1 min +0.0002; accuracy changes ≤0.3 pts, mixed signs. Confirms the earlier momentum finding inside the last 5 minutes too: the path here does not matter, only the current lead.
Distance travelled / in-window RV / range (L2)	NULL	L2–L1 Brier: 5 min +0.0001, 4 min +0.0003, 3 min -0.0000, 2 min -0.0001, 1 min -0.0001. (In-window vol helps inside the <i>probit formula</i> as a σ estimate — as a logistic feature it adds nothing, because the lead already prices it.)

Feature / model idea	Verdict	Evidence (holdout)
Kalshi % change, 1m & 3m (L4)	NULL (one trivial exception)	L4–L3 Brier: 5 min +0.0006, 4 min +0.0000, 3 min +0.0007, 2 min +0.0006, 1 min -0.0003. The 1-min -0.0003 is the only negative delta and is noise-sized; accuracy at 5–3 min actually falls.
Kalshi candle volume / volume change (L5)	NULL	L5–L4 Brier: 5 min +0.0002, 4 min +0.0002, 3 min +0.0003, 2 min +0.0004, 1 min +0.0004. Volume is a confidence sorter in some earlier tests; as a probability feature in the last 5 minutes it earns nothing.
Pre-window 60m return + vol regime (L6)	NULL	L6–L5 Brier: 5 min +0.0000, 4 min +0.0000, 3 min +0.0000, 2 min +0.0000, 1 min +0.0000 — exactly 0.0000 at most minutes (coefficients shrink to ~nothing). The hour before the window tells you nothing about its last 5 minutes beyond what σ already carries.
Pooled lead + minutes-left logistic	NULL / worse	Never beats per-minute L0 on Brier (§10.6).
Lead-bucket empirical probabilities	Decent fallback, worse than market & probit-rolling	Brier 5 min 0.1371, 4 min 0.1222, 3 min 0.1032, 2 min 0.0834, 1 min 0.0408 vs Kalshi 0.1303, 0.1146, 0.0940, 0.0739, 0.0343.

Bottom line for the last 5 minutes

- Direction is the lead; probability is the market.** Lead-sign holdout accuracy runs 81.0% (5 min) → 95.3% (1 min), and no feature tested — momentum, path, distance, volume, Kalshi-% changes, pre-window regime — improves on it honestly.
- For a probability, read the Kalshi price.** It is the best-calibrated, lowest-Brier source at 5, 4 and 3 minutes, and statistically tied with the best logistic at 2 and 1 minutes. Models that include it converge to it; models that exclude it lose to it.
- The honest mathematical model is the rolling- σ probit with settlement averaging** (mean holdout Brier 0.0953 vs Kalshi 0.0894): use it as a cross-check/sanity band around the market price and as the fallback if a market price is ever missing — not as a replacement.
- Confidence lives in lead size + market agreement, not in extra features.** The high-confidence tables (§11–12) show the usable rule: when the best model/market is $\geq 90\%$ confident, holdout accuracy is materially higher, at the coverage cost printed in the tables; when the lead is under $\sim \$25$, no model tested manufactures 90% reliability at 5 minutes left — those windows are genuinely close to coin flips relative to the remaining variance ($\sigma\sqrt{5} \approx \$94$).
- Flip risk halves from 5→1 minute inside a bucket, but bucket (lead size) dominates the clock (§13).** Any last-5-minutes rule should be sized by lead bucket first, minute second.

15. Limitations & honesty notes

- Spot proxy: Coinbase 1-minute opens vs Kalshi's CF Benchmarks settlement basis — tiny-move windows can differ by the basis; this affects all models equally and is the same basis as the existing sheets.
- 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.
- Settlement-averaging assumes a 60-second averaging window and Brownian increments; the exact CF Benchmarks/Kalshi methodology may average differently — the variant is presented as a model, and its holdout gain, while consistent, is small.
- Holdout is a single chronological split (last 30%, Sep 30 → Oct 20). Market conditions in the holdout are one regime; Brier differences under ~ 0.002 should be read as noise.
- "Best model" in §11 was selected on the same holdout it is scored on; §12 exists precisely to give the no-selection comparison.
- Probabilities from logistic models are fitted values on standardized features with L2 shrinkage; results were checked to be insensitive to the penalty choice at the precision reported.

Generated locally from the files listed in §1. Script: `analyze_last5.py` (+ `make_last5_report.py`). No trades placed; no live jobs or crons touched.