

Monthly Model Watch #2

October 2, 2026 · MarketMania Research · Monthly series · window Sep 1-30, 2026 UTC

RESEARCH SNAPSHOT			
20,682 directional calls scored	7 models (stable lineup)	of 39,806 mature forecasts	window Sep 1-30, 2026 UTC
hit = direction rule: tp1/tp2->hit, sl->miss, expiry->sign of gross pnl	FH gate 1h / 4h / 1d	base field hit 45.5%	cutoff Thu Oct 1, 2026, 16:00 UTC
market: BTC net +6.42% (prior +24.95%)	ann vol 41.2% (was 41.5%)	TOP5 volume \$80.01B, +21% m/m	pairwise corr 0.78 (was 0.79)

KEY FINDING · OBSERVATION (one monthly window)

Field accuracy went 46.3% -> 45.5% (-0.8pp) from August to September with 1 of 7 models improving, a single model line led the four weeks, and no monthly title was awarded.

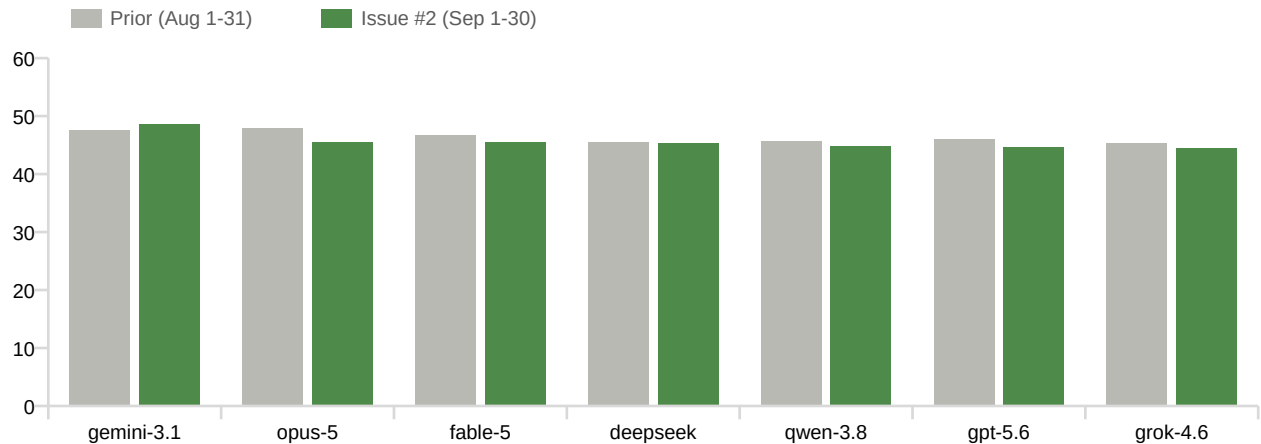
WEEKLY EVIDENCE -> MONTHLY VERDICT (rule below)

Metric	W1 Aug 31-Sep 6	W2 Sep 7-13	W3 Sep 14-20	W4 Sep 21-27	Month	Verdict
Leader (hit-rate)	gemini-3.1	gemini-3.1	gemini-3.1	gemini-3.1	gemini-3.1	Confirmed (4/4)
Most readable coin	BNB	BNB	ETH	XRP	XRP	Rejected (1/4)
Field hit below 50%	42.8%	39.6%	50.7%	46.8%	45.5%	Confirmed (3/4)

Verdict rule: each row states one reading of the month; a week agrees when it shows the same reading (same sign, or the same name). Confirmed = 3 or 4 of the 4 weeks agree with the month; Inconclusive = 2; Rejected = 0 or 1. Weekly cells come from the frozen weekly packs (7-line roster); the month is the pooled Sep 1-30 pull. W1 straddles the month edge (Aug 31).

TL;DR

- **OBSERVATION -- No monthly title.** gemini-3.1-pro leads the month descriptively at 48.5% [46.8%, 50.2%]; the gap to #2 (claude-opus-5, 45.4%) is 3.1pp. 1 of 7 models improved against August, and a single model line held the weekly top spot across the four weeks.
- **No ticker cleared 51%.** XRP is the month's most readable coin (47.0%, n=4,308) and SOL the hardest at 44.0%; 0 of the 5 assets cleared 51% over the month. Pair of the month: gemini-3.1-pro x SOL 49.4% (n=702); worst: grok-4.6 x SOL 40.9% (n=503).
- **Mean weekly rank ran 1.00 to 5.25 across the four weeks.** claude-opus-5 swung across 5 rank positions and gemini-3.1-pro across 0. Consecutive-week Spearman came in at W1-W2 0.54, W2-W3 -0.11, W3-W4 0.57.
- **The field read volatile weeks better.** Split by the pack's own volatility rule the field hit 43.4% (n=9,426) in the two calm weeks against 48.2% (n=9,430) in the two volatile weeks, a difference of +4.8pp. Cross-TF/FH lifts ran -1.5pp to +0.2pp. On trend days, 4h TF-agreement hit 55.2% (n=1,026) against 32.5% (n=1,279) on flat days.



Directional hit-rate (%) by model, the prior window against the whole month, ranked by this month's rate. Short names: opus-5 = claude-opus-5, fable-5 = claude-fable-5, gemini-3.1 = gemini-3.1-pro, gpt-5.6 = gpt-5.6-sol, qwen-3.8 = qwen-3.8-max, deepseek = deepseek-v4-pro. Field hit 46.3% -> 45.5%. The prior window is the full August: window A of the same pull, slots 2026-08-01 to 2026-09-01 (first slot actually seen 2026-08-01 00:01 UTC), matured at the prior cutoff Tue Sep 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as September -- 19,383 directional calls against 20,682 in the month. The leaderboard's M/m column is computed on unrounded rates and can differ by 0.1pp from the difference of the rounded columns.

Why it matters

MarketMania scores 7 model lines against the same market, hour after hour. Monthly Model Watch asks the weekly questions at monthly n -- who led, which tickers were readable, who saw the turn first, does self-agreement help -- and then asks the three questions a single week cannot answer: which coin the field reads best over a month, how far a weekly rank moves inside one month, and how the field's accuracy differs between the calm and volatile weeks the pack labels. Each block gets its own table, its own N and its own caveats.

How to read this

Every block below is scored on the same pool: matured directional calls inside the 1h/4h/1d FH gate, hit by the methodology v1.1 direction rule, over the whole month at one cutoff (Thu Oct 1, 2026, 16:00 UTC). The monthly title is awarded only on a >=5pp gap AND non-overlapping 95% Wilson CIs with N>=10 per side -- a descriptive lead is not a title. Reversals use the 4h grid with a >=2.0% counter-move sustained 8h, BTC/ETH only in v1. Cross-TF/FH blocks ask whether a model agreeing with itself does any better, and are split trend vs flat. The by-week blocks come from the four frozen weekly packs at their own published cutoffs. No model line is lineage-spliced this month (the prior window carries the grok and qwen splices).

Leaderboard of the month

Model	N	Cov.	Hit rate	95% CI	M/m pp
gemini-3.1-pro	3,284	57.6%	48.5%	46.8%-50.2%	+0.9
claude-opus-5	2,532	44.5%	45.4%	43.5%-47.3%	-2.4
claude-fable-5	2,945	51.7%	45.4%	43.6%-47.2%	-1.4
deepseek-v4-pro	3,352	59.1%	45.3%	43.6%-46.9%	-0.3
qwen-3.8-max	3,165	56.0%	44.7%	43.0%-46.5%	-0.9
gpt-5.6-sol	3,133	55.0%	44.5%	42.7%-46.2%	-1.4
grok-4.6	2,271	40.0%	44.4%	42.4%-46.4%	-0.9

No monthly title is awarded this issue. Title rule: >=5pp gap AND non-overlapping 95% CI, N>=10 per side. gemini-3.1-pro's lead over claude-opus-5 is 3.1pp and their CIs overlap, so neither condition is met. M/m pp = this month's hit rate minus the same line's hit rate in the pack's prior window Aug 1-31, n=19,383 directional calls there against 20,682 here, so it is a step between windows of different length. In the prior window the lineage splice folds 4,520 grok-4.5 rows into grok-4.6 and 870 qwen-3.7-max rows into qwen-3.8-max (5,390 rows in all).

Ticker of the month

Symbol	N	Hit rate	95% CI	Prior
XRP	4,308	47.0%	45.6%-48.5%	46.7%
BTC	3,987	46.0%	44.4%-47.5%	46.5%
ETH	3,973	45.4%	43.8%-46.9%	42.9%
BNB	3,994	45.1%	43.6%-46.7%	48.1%
SOL	4,420	44.0%	42.6%-45.5%	47.3%

XRP was the field's most readable ticker (47.0%) and SOL the hardest (44.0%); 0 of 5 assets cleared 51% this month. Prior = the same per-symbol hit rate over Aug 1-31, on n=3,695 to 4,068 per coin. This issue's ordering is XRP > BTC > ETH > BNB > SOL.

Pair of the month

Top 6 of the 12 model x ticker pairs tracked this month (best hit-rate), then the 5 worst in their own table -- the pair list is split rather than paged so neither table breaks across a page. Pairs shown require N of 8 or more calls.

Model (top 6)	Symbol	N	Hit rate
gemini-3.1-pro	SOL	702	49.4%
gemini-3.1-pro	BTC	640	49.1%
gemini-3.1-pro	XRP	673	48.9%
grok-4.6	XRP	482	48.5%
gemini-3.1-pro	ETH	622	47.6%
gemini-3.1-pro	BNB	647	47.4%

Model (5 worst)	Symbol	N	Hit rate
deepseek-v4-pro	SOL	691	43.3%
grok-4.6	BTC	407	43.2%
gpt-5.6-sol	SOL	657	42.5%
qwen-3.8-max	BNB	594	42.1%
grok-4.6	SOL	503	40.9%

gemini-3.1-pro x SOL (49.4%, n=702) is the month's best pair and grok-4.6 x SOL (40.9%, n=503) the weakest. Pair history accumulates across issues -- read this as the second monthly data point, not a ranking.

Who saw the reversal first

Reversal definition: 4h grid; trend = sign of prior 24h; counter-move of 2.0% or more (BTC/ETH) sustained 8h. A model 'calls' the reversal if it has a matured directional hit call in the new direction, FH 4h or 1d, in the slot window [T-12h, T+FH].

Symbol	Time (UTC)	New side	Move (8h)	Callers	First caller
BTC	Wed Sep 23, 08:00	Short	-2.90%	1	gemini-3.1-pro
ETH	Thu Sep 3, 08:00	Long	+4.61%	0	NOBODY
ETH	Fri Sep 4, 08:00	Short	-2.13%	0	NOBODY
ETH	Fri Sep 11, 08:00	Long	+4.19%	11	gemini-3.1-pro
ETH	Fri Sep 11, 16:00	Short	-2.11%	0	NOBODY
ETH	Sun Sep 20, 12:00	Long	+2.22%	17	gemini-3.1-pro
ETH	Wed Sep 23, 08:00	Short	-3.73%	6	gpt-5.6-sol

The BTC short turn drew 1 caller, first gemini-3.1-pro on a 1d/4h call at Sep 24, 00:01 with confidence 65; the ETH long turn drew NOBODY; the ETH short turn drew NOBODY; the ETH long turn drew 11 callers, first gemini-3.1-pro on a 1d/1d call at Sep 12, 00:01 with confidence 70; the ETH short turn drew NOBODY; the ETH long turn drew 17 callers, first gemini-3.1-pro on a 1d/1d call at Sep 20, 00:01 with confidence 75; the ETH short turn drew 6 callers, first gpt-5.6-sol on a 4h/1h call at Sep 23, 08:01 with confidence 67. Scope in v1 remains BTC/ETH only; 7 events in this window, against 2 / 2 / 1 / 2 in the four published weekly issues (W1 Aug 31-Sep 6, W2 Sep 7-13, W3 Sep 14-20, W4 Sep 21-27), and 4 of the 7 found at least one caller. The window reaches past

the turn, so a counted call is not necessarily ahead of it: the first counted call came before the turn in 1 case (ETH Sep 20 by 11h 59min) and after it in 3 cases (BTC Sep 23 by 16h 01min; ETH Sep 11 by 16h 01min; ETH Sep 23 by 1 min) -- slot times, not the answers' ready times; calls in the old direction are not counted.

Cross-TF and cross-FH confirmation: does agreeing with yourself help?

Block	FH	vs.	Agree n	Agree hit	Dis n	Dis hit	Base hit	Lift
cross-TF	4h	TF 4h vs 1h	2,305	42.6%	68	57.4%	43.9% (3,139)	-1.4
cross-TF	1d	TF 1d vs 4h	423	34.8%	21	57.1%	36.2% (665)	-1.5
cross-FH	4h	by FH 1h	2,059	44.1%	92	51.1%	43.9% (3,139)	+0.2
cross-FH	1d	by FH 4h	341	36.4%	28	42.9%	36.2% (665)	+0.1

* = N<10 (insufficient); insufficient cells are never used to rank models. Mixed pairs (the same model's partner call at that slot is sideways, not directional): 761 / 221 / 982 / 296; primaries with no partner call stored at that slot, in no bucket: 5 / 0 / 6 / 0 -- so agree + disagree + mixed + unpaired = the base n of each family. Pooled across all four cell families the lifts run -1.5pp to +0.2pp this month.

Regime control. Methodology calls for a trend-vs-flat split on this table. This window has 10 trend days of 30 (a day is a trend day when its absolute BTC 1d move is 1.5% or more, else a flat day), and the four weeks label as W1 volatile, W2 calm, W3 volatile, W4 calm under the pack's rule (a week is volatile when its mean absolute BTC 1d move is at or above the median of the four weeks, else calm; median mean |BTC 1d move| 1.482%). On trend days, 4h TF-agreement hit 55.2% (n=1,026) against 32.5% (n=1,279) on flat days.

Side-mix and the month's failure

Model	N	Long	Short	Sideways	L:S
claude-fable-5	5,700	34.0%	17.7%	48.3%	1.92
claude-opus-5	5,696	29.9%	14.5%	55.5%	2.05
deepseek-v4-pro	5,675	33.0%	26.1%	40.9%	1.26
gemini-3.1-pro	5,699	32.4%	25.2%	42.4%	1.28
gpt-5.6-sol	5,697	27.8%	27.2%	45.0%	1.02
grok-4.6	5,684	19.3%	20.6%	60.1%	0.93
qwen-3.8-max	5,655	33.3%	22.7%	44.0%	1.46

L:S = each model's own long-share divided by its short-share; the field average of the per-model ratios is 1.42 this month.

'Sideways' was 48.0% of mature forecasts (19,124 of 39,806). Consensus skew: 2,473 symbol/FH/TF cells had 5 or more models on the same side and hit 45.0% (n=15,689) against the 45.5% field base.

Model	Symbol	Side	Conf	FH / TF	Slot (UTC)	Exit	Net PnL
gpt-5.6-sol	BNB	Short	84	1h / 1h	Sep 16, 09:01	EXPIRY	-\$0.67

Fail of the month. The month's single highest-confidence individual miss, exit reason expiry. Listed as a single card, never as a model ranking.

Calibration bridge

The hit-rate leader this month is gemini-3.1-pro (48.5%); the lowest-Brier line is qwen-3.8-max (Brier 0.2701, gap +14.7pp). Per-model ok-rate this window ran 99.06% to 99.98%. The full confidence-bucket analysis and the week-by-week gap drift live in Monthly Calibration #2.

Asset league: which coin the field reads best

A published week gives each coin 780 to 1,190 calls; the month gives 3,973 to 4,420 per coin. Across the month the field's hit rate ran from 47.0% on XRP down to 44.0% on SOL, a spread of +3.0pp against a 45.5% field base. The ordering below is the month pull; the third table shows how often that ordering held from week to week.

Symbol	N	Hit rate	95% CI	Rank
XRP	4,308	47.0%	45.6%-48.5%	1
BTC	3,987	46.0%	44.4%-47.5%	2
ETH	3,973	45.4%	43.8%-46.9%	3
BNB	3,994	45.1%	43.6%-46.7%	4
SOL	4,420	44.0%	42.6%-45.5%	5

Ranked by hit rate, ties broken by n descending; Wilson 95% CIs are descriptive and 9 of the 10 pairwise CI comparisons in this table overlap. Source: the month pull split by slot week, at the monthly cutoff.

Symbol	1h	4h	1d	Best FH
XRP	47.6% (2,726)	46.8% (1,343)	42.3% (239)	1h
BTC	48.3% (2,538)	44.0% (1,243)	29.1% (206)	1h
ETH	46.9% (2,525)	44.0% (1,223)	36.0% (225)	1h
BNB	48.7% (2,499)	39.5% (1,249)	37.4% (246)	1h
SOL	45.4% (2,720)	43.5% (1,403)	33.3% (297)	1h

Cells are hit rate with n in parentheses; * = N<10 (insufficient) and such cells never win the Best FH column, which is left n/a when no horizon clears the floor. Same pool and cutoff as the table above.

Symbol	W1 Aug 31-Sep 6	W2 Sep 7-13	W3 Sep 14-20	W4 Sep 21-27	Month
XRP	44.2% (2)	42.2% (2)	50.4% (2)	50.0% (1)	47.0% (1)
BTC	40.6% (4)	41.4% (3)	50.3% (3)	48.9% (2)	46.0% (2)
ETH	43.1% (3)	29.5% (5)	57.2% (1)	45.8% (4)	45.4% (3)
BNB	47.3% (1)	46.9% (1)	46.3% (5)	38.2% (5)	45.1% (4)
SOL	38.7% (5)	36.0% (4)	49.5% (4)	48.7% (3)	44.0% (5)

Cells are hit rate with the week's rank in parentheses. The four weekly columns are the PUBLISHED weekly numbers -- each week at its own frozen cutoff; the Month column is the single month pull. SOL ranked first in 0 of 4 weeks and 5 for the month; best horizon 1h (45.4% on n=2,720).

Rank stability across the four weeks

Every model's rank in each of the four published weeks, its rank for the pooled month, its mean rank and the span between its best and worst week, with the hit rate next to every rank. Mean rank runs 1.00 to 5.25, the widest span is 5 rank positions and the narrowest 0, and a single model line held the weekly top spot across the four weeks.

Model	W1	W2	W3	W4	Month	Mean rank	Range
gemini-3.1-pro	1 (46.5%)	1 (41.8%)	1 (52.9%)	1 (51.4%)	1 (48.5%)	1.00	0
claude-opus-5	5 (42.4%)	7 (36.5%)	3 (51.6%)	2 (47.6%)	2 (45.4%)	4.25	5
claude-fable-5	3 (42.8%)	6 (36.5%)	2 (52.3%)	5 (46.7%)	3 (45.4%)	4.00	4
deepseek-v4-pro	2 (44.2%)	2 (41.4%)	7 (48.4%)	7 (43.9%)	4 (45.3%)	4.50	5
qwen-3.8-max	4 (42.5%)	4 (40.0%)	4 (50.8%)	6 (44.5%)	5 (44.7%)	4.50	2
gpt-5.6-sol	6 (41.6%)	3 (40.1%)	5 (49.7%)	4 (46.8%)	6 (44.5%)	4.50	3
grok-4.6	7 (37.6%)	5 (39.6%)	6 (49.6%)	3 (46.9%)	7 (44.4%)	5.25	4

Rank = position in that window's leaderboard, which the pack sorts by hit rate descending; hit rate in parentheses. Weeks are the published issues at their own frozen cutoffs (W1 Aug 31-Sep 6, W2 Sep 7-13, W3 Sep 14-20, W4 Sep 21-27); the Month column is the single month pull, so the month rank is not the mean of the four weekly ranks. Consecutive-week Spearman rho: W1-W2 0.54, W2-W3 -0.11, W3-W4 0.57 (ties get average ranks). A single model line held the weekly top spot. Sorted by month rank.

Calm vs volatile weeks

The pack labels each of the four weeks from the BTC daily series: a week is volatile when its mean absolute BTC 1d move is at or above the median of the four weeks, else calm, against a median mean |BTC 1d move| of 1.482% across the four weeks; the trend/flat day counts come from the same BTC series under the daily rule (a day is a trend day when its absolute BTC 1d move is 1.5% or more, else a flat day). The rule labels 2 weeks calm and 2 weeks volatile, which splits the four weeks into two regime cells.

Week	Mean BTC 1d move	Trend days	Flat days	Label
W1 Sep 1-6	1.555%	3	3	volatile
W2 Sep 7-13	0.864%	2	5	calm
W3 Sep 14-20	1.803%	3	4	volatile
W4 Sep 21-27	1.410%	2	5	calm

Mean |BTC 1d move| = the mean absolute daily BTC move of that week, from the entry-price series the pack reconstructs; the label is the pack's rule against the four-week median (1.482%). BTC only: this is not the market check's "Mean |1d move| (5 assets)" row, which averages five coins over the whole window. W1 holds Sep 1-6 here: the month pull starts on Sep 1.

Model	Calm n	Calm hit	Volatile n	Volatile hit	Diff pp
claude-fable-5	1,360	42.1%	1,325	49.1%	+7.0
claude-opus-5	1,130	42.6%	1,173	48.3%	+5.8
qwen-3.8-max	1,457	42.4%	1,422	47.9%	+5.5
deepseek-v4-pro	1,504	42.7%	1,551	47.7%	+5.0
gemini-3.1-pro	1,467	47.0%	1,530	50.8%	+3.9
gpt-5.6-sol	1,453	43.5%	1,397	47.2%	+3.7
grok-4.6	1,055	43.3%	1,032	45.4%	+2.0
Field (all models)	9,426	43.4%	9,430	48.2%	+4.8

Definition: diff_pp = volatile hit_rate - calm hit_rate, in percentage points. Sorted by Diff pp descending, Field row last. * = N<10 (insufficient). Rows outside the four Mon-Sun weeks from Aug 31 to Sep 27 (1,826 calls on Sep 28-30) are not in either regime cell, so the two columns do not sum to the month.

Market check: September ran calmer on vol, wider on price

OBSERVATION — mean |1d move| 1.98% -> 1.91% (-4% rel), BTC realized vol 36.9% -> 34.9% (ann., hourly); field directional accuracy 46.3% -> 45.5% (-0.8pp), 1 of 7 models improved, sim win-rate up for 6 of 7, field sim PnL -\$1054.79 -> -\$2554.15 (Aug 1-31 vs Sep 1-30; descriptive, one pair of windows, not a claim).

Robustness: raw price-sign accuracy 46.0% -> 48.4% — the opposite direction to the trade-based hit rule.

Measure	Aug 1-31	Sep 1-30	Change
Mean 1d move (5 assets)	1.98%	1.91%	-4% rel
BTC realized vol (ann., hourly)	36.9%	34.9%	-2.0pp
Field directional accuracy	46.3%	45.5%	-0.8pp
Models improving hit-rate	--	1 of 7	--
Raw price-sign accuracy	46.0%	48.4%	+2.4pp
Field sim win-rate	35.8%	38.4%	+2.6pp
Field sim net PnL	-\$1054.79	-\$2554.15	--

Window A is the pack's prior window -- the full August -- so this slice and every month-over-month column in this report compare September against the full August. Descriptive, one pair of windows; the hit rule is methodology v1.1 (trade-based) and the raw price-sign row is the robustness check (hourly closes at :00 against slots at :01). BTC realized vol here is the hourly-return series from the alive slice; the snapshot row's figure is the audited daily-candle estimate -- different estimators, both reported as measured. Lineage splices in this window: none. 0 rows were remapped inside the month against 5,390 in the prior window.

Practical implications

- On this month's leaderboard no title was awarded, the top two CIs overlap, and a single model line led the four Mon-Sun weeks from Aug 31 to Sep 27.
- XRP ranked #1 for the month on $n=4,308$; the week-by-week league shows it ranked #1 in 1 of the four published weeks.
- The field differs by +4.8pp between the two regimes on $n=9,426$ calm vs $n=9,430$ volatile; the rule that draws the line is the pack's own four-week median, recomputed every month.

Limitations

- The prior window is the full August: every "prior" column, the gray series of the page-1 chart and the market-check A column describe Aug 1-31. It is window A of the same pull, slots 2026-08-01 to 2026-09-01 (first slot actually seen 2026-08-01 00:01 UTC), matured at the prior cutoff Tue Sep 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as September -- 19,383 directional calls against 20,682 in the month. Month-over-month deltas are a step between two windows of different length (31 and 30 days).
- Two cutoffs live in this issue. The month is frozen at Thu Oct 1, 2026, 16:00 UTC; the four weekly blocks are frozen at their own published cutoffs (W1 Aug 31-Sep 6, W2 Sep 7-13, W3 Sep 14-20, W4 Sep 21-27). A forecast still pending at its week's cutoff is immature in the weekly pack and mature in the month pull, so the month does not equal the sum of the weeks (20,682 vs 19,544 directional, plus 1,826 on the 3 remainder days); W1 also opens on Aug 31, a day the month pull does not hold, so its pack counts 4,520 directional calls against 3,832 for Sep 1-6 inside the month pull -- that gap is arithmetic, not a data problem.
- Observations inside one window are not independent, and the 95% Wilson CIs shown throughout are descriptive, not inferential. Ranks are ordinal: a 0.3pp gap and a 6pp gap both read as one rank position, so the rank-stability table prints the hit rate next to every rank.
- The reversal detector remains v1 and BTC/ETH-only, on price series reconstructed from trade entry prices; 7 events in this window still cannot characterise detector performance. Cross-TF and cross-FH disagreement cells are small ($n=68, 21, 92, 28$); cells under $N=10$ are insufficient and never used to rank models.
- Model lines in this window: claude-fable-5, claude-opus-5, deepseek-v4-pro, gemini-3.1-pro, gpt-5.6-sol, grok-4.6, qwen-3.8-max. Series density and lineage notes (monthly #2): (1) rolling-1w forecasts emit daily since Aug 22, 2026 and rolling-1M daily since Aug 23, 2026, so across this 30-day window daily coverage is FULL as measured -- the 1w series has slots on 30 of 30 days and the 1M series on 30 of 30 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 2,096, 1M 2,089). (2) No model line is lineage-spliced inside this window. (3) Every per-model aggregate in this issue sees the merged head id; the raw id is kept in the pack, and 0 rows were remapped inside the month against 5,390 in the prior window.
- Market-state row is computed on a single-exchange (binance) daily candle series, as in the weekly series after the audit that found the raw table mixes two exchanges. The prior-month block is the full August: it covers 31 days (Aug 1-31), and the TOP5 volume figure compares a 30-day sum with a 31-day sum -- a level difference, not a like-for-like change.
- Research-to-date counter: directional forecasts resolved inside the FH gate (1h/4h/1d) since Jul 11, counted at the issue cutoff (Thu Oct 1, 2026, 16:00 UTC) -- the D2 definition pinned in the weekly series; this issue prints 59,755. The pack carries no prior-cutoff pin for a monthly window (the prior-cutoff pin is null by construction), so the control this issue is the weekly reproduction: the W4 block of this pack (Sep 21-27) reproduces the published weekly #8 on 7 of 7 checks, and the prior window of this pack (Aug 1-31, prior cutoff Tue Sep 1, 2026, 16:00 UTC) reproduces the published Monthly #1 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	44,100	Mature (scored pool)	39,806
OK in gate	39,806	-- of them directional	20,682
Out of gate (1w / 1M)	2,096 / 2,089	-- of them sideways	19,124
Invalid	109	Pending (next issue)	0
Late closes	0	Uptime, all grids	1,140 / 1,140 slots
Source file	monthly_metrics_2026-09.json	Report cutoff	Thu Oct 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-10-02 08:05 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 20,682 scored observations. MARKETMANIA RESEARCH TO DATE (as of cutoff Oct 1, 16:00 UTC): 59,755 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 41 published reports.

Issue #2. Monthly Model Watch is a living monthly comparison; each issue appends one more calendar month of leaderboard, asset-league, rank-stability and regime data. This second issue reads September against the full August: gemini-3.1-pro at the top (48.5%; 47.6% in the prior window), 4 of 7 qualifying turns found a caller, 0 of 5 tickers above 51%, a single model line leading the four weeks and a calm-vs-volatile field difference of +4.8pp. No model line is lineage-spliced this month (the prior window carries the grok and qwen splices). Engine 1.1 has powered the sandbox since Aug 18, i.e. before this window and inside the prior window; no cross-engine PnL comparisons are claimed.

What we're testing next

- Monthly #3 (October): the same three series over Oct 1-31, cut on Nov 1 plus the maturity lag, so its month-over-month comparison runs against this September.
- Monthly Config Watch #2: the same September window through the config search engine (w298), published as the fourth monthly report of this issue.
- Quarterly Stability Index: Run 3, the next repeat-prompt stability run, is planned for the first half of October and is the first cross-quarter point on model stability.

Related research

Report	Direct PDF link
Monthly Consensus Watch #2	https://marketmania.ai/research/reports/consensus-watch-monthly-2026-09.pdf
Monthly Calibration #2	https://marketmania.ai/research/reports/calibration-monthly-2026-09.pdf
Weekly Model Watch #8	https://marketmania.ai/research/reports/model-watch-2026-09-21.pdf

The three monthly reports publish together as one issue each month; each links straight to the others' PDF and to the last weekly issue of the same series. Direct links are the posting rule from wave 2 on.

CITE THIS REPORT

```
@misc{mm_modelwatch_monthly_2026m09,
  title = {Monthly Model Watch #2: monthly leaderboard, asset league, rank stability and regime split, Sep 1-30 2026},
  author = {{MarketMania Research}},
  year = {2026}, month = {October}, day = {2},
  url = {https://marketmania.ai/research/reports/model-watch-monthly-2026-09.pdf},
  note = {Methodology v1.1 (2026-08-10), hash e66c7e8c864a2233; window Sep 1-30, 2026 UTC; source monthly_metrics_2026-09.json}
}
```