Monthly Model Watch #3

November 5, 2026 · MarketMania Research · Monthly series · window Oct 1-31, 2026 UTC

RESEARCH SNAPSHOT			
22,579 directional calls scored	**7** models (lineage spliced: opus, gpt, grok)	of **40,889** mature forecasts	window **Oct 1-31, 2026 UTC**
hit = direction rule: tp1/tp2->hit, sl->miss, expiry->sign of gross pnl	FH gate **1h / 4h / 1d**	base field hit **48.6%**	cutoff **Sun Nov 1, 2026, 16:00 UTC**
market: BTC net **-7.22%** (prior +6.42%)	ann vol **28.3%** (was 41.2%)	TOP5 volume **$68.94B**, -14% m/m	pairwise corr **0.87** (was 0.78)

KEY FINDING · OBSERVATION (one monthly window)

Field accuracy went 45.5% -> 48.6% (+3.1pp) from September to October with 6 of 7 models improving, 4 distinct model lines led the four weeks, and no monthly title was awarded.

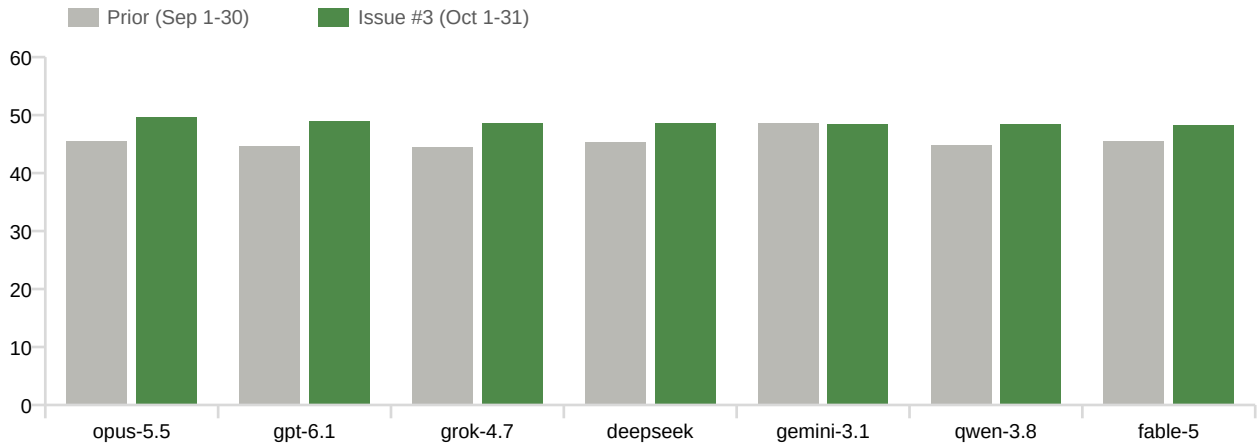
WEEKLY EVIDENCE -> MONTHLY VERDICT (rule below)

Metric	W1 Oct 5-11	W2 Oct 12-18	W3 Oct 19-25	W4 Oct 26-Nov 1	Month	Verdict
Leader (hit-rate)	opus-5.5	deepseek	gemini-3.1	grok-4.7	opus-5.5	Rejected (1/4)
Most readable coin	BTC	XRP	XRP	SOL	XRP	Inconclusive (2/4)
Field hit below 50%	43.8%	46.6%	57.0%	41.6%	48.6%	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 Oct 1-31 pull. W4 straddles the month edge (Nov 1); Oct 1-4 belong to the month only.
The model cutover of Oct 17, 2026 09:01 UTC (gpt-5.6-sol -> gpt-6.1-sol, claude-opus-5 -> claude-opus-5.5, grok-4.6 -> grok-4.7) falls inside W2; W1 runs on the earlier builds, W3 and W4 on the new ones, the month column joins both -- all per-line rows are model lines, not single builds.

TL;DR

- **OBSERVATION -- No monthly title.** claude-opus-5.5 (spliced) leads the month descriptively at 49.6% [47.7%, 51.5%]; the gap to #2 (gpt-6.1-sol (spliced), 48.9%) is 0.7pp. 6 of 7 models improved against September, and 4 distinct model lines held the weekly top spot across the four weeks.
- **No ticker cleared 51%.** XRP is the month's most readable coin (50.2%, n=4,431) and BNB the hardest at 47.5%; 0 of the 5 assets cleared 51% over the month. Pair of the month: claude-opus-5.5 (spliced) x XRP 52.5% (n=499); worst: claude-fable-5 x BNB 46.0% (n=592).
- **Mean weekly rank ran 2.75 to 5.25 across the four weeks.** deepseek-v4-pro swung across 6 rank positions and gpt-6.1-sol (spliced) across 1. Consecutive-week Spearman came in at W1-W2 -0.68, W2-W3 -0.36, W3-W4 -0.11.
- **The field read volatile weeks better.** Split by the pack's own volatility rule the field hit 44.3% (n=9,617) in the two calm weeks against 51.2% (n=10,457) in the two volatile weeks, a difference of +6.9pp. Cross-TF/FH lifts ran -1.3pp to +2.8pp. On trend days, 4h TF-agreement hit 53.3% (n=808) against 42.8% (n=1,862) on flat days.



Directional hit-rate (%) by model, the prior window against the whole month, ranked by this month's rate. Short names: opus-5.5 = claude-opus-5.5, fable-5 = claude-fable-5, gemini-3.1 = gemini-3.1-pro, gpt-6.1 = gpt-6.1-sol, qwen-3.8 = qwen-3.8-max, deepseek = deepseek-v4-pro. Field hit 45.5% -> 48.6%. The prior window is the full September: window A of the same pull, slots 2026-09-01 to 2026-10-01 (first slot actually seen 2026-09-01 00:01 UTC), matured at the prior cutoff Thu Oct 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as October -- 20,682 directional calls against 22,579 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 (Sun Nov 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. 3 model lines are lineage splices this month (claude-opus-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17)).

Leaderboard of the month

Model	N	Cov.	Hit rate	95% CI	M/m pp
claude-opus-5.5 (spliced)	2,612	44.7%	49.6%	47.7%-51.5%	+4.2
gpt-6.1-sol (spliced)	3,627	62.0%	48.9%	47.3%-50.5%	+4.4
grok-4.7 (spliced)	3,069	53.5%	48.6%	46.9%-50.4%	+4.2
deepseek-v4-pro	3,485	59.5%	48.5%	46.9%-50.2%	+3.3
gemini-3.1-pro	3,439	58.7%	48.4%	46.7%-50.1%	-0.1
qwen-3.8-max	3,344	56.9%	48.4%	46.7%-50.0%	+3.6
claude-fable-5	3,003	51.2%	48.2%	46.4%-50.0%	+2.8

No monthly title is awarded this issue. Title rule: **>=5pp gap AND non-overlapping 95% CI, N>=10** per side. claude-opus-5.5 (spliced)'s lead over gpt-6.1-sol (spliced) is 0.7pp 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 Sep 1-30, n=20,682 directional calls there against 22,579 here, so it is a step between windows of different length. claude-opus-5.5 (spliced) = claude-opus-5.5 (incl. opus-5-era, Oct 1-17). gpt-6.1-sol (spliced) = gpt-6.1-sol (incl. 5.6-era, Oct 1-17). grok-4.7 (spliced) = grok-4.7 (incl. 4.6-era, Oct 1-17). In the prior window the lineage splice folds 6,300 claude-opus-5 rows into claude-opus-5.5, 6,300 gpt-5.6-sol rows into gpt-6.1-sol and 6,300 grok-4.6 rows into

grok-4.7 (18,900 rows in all).

Ticker of the month

Symbol	N	Hit rate	95% CI	Prior
XRP	4,431	50.2%	48.8%-51.7%	47.0%
BTC	4,703	49.6%	48.2%-51.0%	46.0%
ETH	4,475	48.0%	46.5%-49.5%	45.4%
SOL	4,590	47.8%	46.4%-49.2%	44.0%
BNB	4,380	47.5%	46.0%-48.9%	45.1%

XRP was the field's most readable ticker (50.2%) and BNB the hardest (47.5%); 0 of 5 assets cleared 51% this month. Prior = the same per-symbol hit rate over Sep 1-30, on n=3,973 to 4,420 per coin. This issue's ordering is XRP > BTC > ETH > SOL > BNB.

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
claude-opus-5.5 (spliced)	XRP	499	52.5%
claude-fable-5	XRP	550	50.9%
deepseek-v4-pro	BTC	716	50.8%
grok-4.7 (spliced)	XRP	612	50.5%
claude-opus-5.5 (spliced)	SOL	526	50.2%
gpt-6.1-sol (spliced)	BTC	756	50.1%

Model (5 worst)	Symbol	N	Hit rate
claude-fable-5	SOL	635	46.9%
deepseek-v4-pro	SOL	682	46.8%
gemini-3.1-pro	SOL	701	46.7%
qwen-3.8-max	BNB	661	46.3%
claude-fable-5	BNB	592	46.0%

claude-opus-5.5 (spliced) x XRP (52.5%, n=499) is the month's best pair and claude-fable-5 x BNB (46.0%, n=592) the weakest. Pair history accumulates across issues -- read this as the third 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	Fri Oct 2, 12:00	Short	-2.28%	0	NOBODY
BTC	Tue Oct 13, 00:00	Long	+2.40%	0	NOBODY
BTC	Thu Oct 22, 00:00	Short	-2.32%	1	gpt-6.1-sol (spliced)
BTC	Wed Oct 28, 20:00	Short	-2.36%	0	NOBODY
BTC	Thu Oct 29, 08:00	Long	+2.27%	0	NOBODY
BTC	Thu Oct 29, 16:00	Short	-4.55%	14	gpt-6.1-sol (spliced)
ETH	Fri Oct 2, 12:00	Short	-2.75%	0	NOBODY
ETH	Mon Oct 12, 16:00	Short	-3.33%	7	deepseek-v4-pro
ETH	Tue Oct 13, 00:00	Long	+3.24%	0	NOBODY
ETH	Thu Oct 15, 08:00	Long	+2.60%	0	NOBODY
ETH	Fri Oct 16, 12:00	Short	-2.56%	13	gpt-6.1-sol (spliced)
ETH	Sun Oct 18, 20:00	Long	+3.52%	0	NOBODY
ETH	Tue Oct 20, 08:00	Long	+2.34%	8	gemini-3.1-pro
ETH	Sat Oct 24, 16:00	Long	+2.21%	1	gemini-3.1-pro
ETH	Wed Oct 28, 20:00	Short	-3.46%	2	gpt-6.1-sol (spliced)
ETH	Thu Oct 29, 08:00	Long	+2.36%	0	NOBODY
ETH	Thu Oct 29, 16:00	Short	-4.84%	14	gemini-3.1-pro

The BTC short turn drew NOBODY; the BTC long turn drew NOBODY; the BTC short turn drew 1 caller, first gpt-6.1-sol (spliced) on a 4h/1h call at Oct 22, 00:01 with confidence 59; the BTC short turn drew NOBODY; the BTC long turn drew NOBODY; the BTC short turn drew 14 callers, first gpt-6.1-sol (spliced) on a 1d/4h call at Oct 30, 00:01 with confidence 70; the ETH short turn drew NOBODY; the ETH short turn drew 7 callers, first deepseek-v4-pro on a 4h/1h call at Oct 12, 12:01 with confidence 58; the ETH long turn drew NOBODY; the ETH long turn drew NOBODY; the ETH short turn drew 13 callers, first gpt-6.1-sol (spliced) on a 1d/1d call at Oct 16, 00:01 with confidence 76; the ETH long turn drew NOBODY; the ETH long turn drew 8 callers, first gemini-3.1-pro on a 1d/4h call at Oct 21, 00:01 with confidence 65; the ETH long turn drew 1 caller, first gemini-3.1-pro on a 4h/1h call at Oct 24, 16:01 with confidence 60; the ETH short turn drew 2 callers, first gpt-6.1-sol (spliced) on a 1d/1d call at Oct 29, 00:01 with confidence 61; the ETH long turn drew NOBODY; the ETH short turn drew 14 callers, first gemini-3.1-pro on a 1d/4h call at Oct 30, 00:01 with confidence 75. Scope in v1 remains BTC/ETH only; 17 events in this window, against 0 / 6 / 3 / 6 in the four published weekly issues (W1 Oct 5-11, W2 Oct 12-18, W3 Oct 19-25, W4 Oct 26-Nov 1), and 8 of the 17 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 2 cases (ETH Oct 12 by 3h 59min; ETH Oct 16 by 11h 59min) and after it in 6 cases (BTC Oct 22 by 1 min; BTC Oct 29 by 8h 01min; ETH Oct 20 by 16h 01min; ETH Oct 24 by 1 min; ETH Oct 28 by 4h 01min; ETH Oct 29 by 8h 01min) -- 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	2,670	46.0%	140	57.9%	47.3% (3,712)	-1.3
cross-TF	1d	TF 1d vs 4h	460	49.4%	13	0.0%	46.9% (621)	+2.5
cross-FH	4h	by FH 1h	2,452	46.7%	153	51.6%	47.3% (3,712)	-0.6
cross-FH	1d	by FH 4h	397	49.6%	14	7.1%	46.9% (621)	+2.8

* = 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): 897 / 148 / 1101 / 210; 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 -1.3pp to +2.8pp this month.

Regime control. Methodology calls for a trend-vs-flat split on this table. This window has 9 trend days of 31 (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| 1.115%). On trend days, 4h TF-agreement hit 53.3% (n=808) against 42.8% (n=1,862) on flat days.

Side-mix and the month's failure

Model	N	Long	Short	Sideways	L:S
claude-fable-5	5,860	29.4%	21.8%	48.8%	1.35
claude-opus-5.5 (spliced)	5,840	23.3%	21.5%	55.3%	1.08
deepseek-v4-pro	5,860	25.0%	34.4%	40.5%	0.73
gemini-3.1-pro	5,855	28.1%	30.6%	41.3%	0.92
gpt-6.1-sol (spliced)	5,855	22.1%	39.8%	38.0%	0.56
grok-4.7 (spliced)	5,737	20.4%	33.1%	46.5%	0.62
qwen-3.8-max	5,882	28.2%	28.7%	43.1%	0.98

L:S = each model's own long-share divided by its short-share; the field average of the per-model ratios is 0.89 this month. 'Sideways' was 44.8% of mature forecasts (18,310 of 40,889). Consensus skew: 2,663 symbol/FH/TF cells had 5 or more models on the same side and hit 48.8% (n=16,912) against the 48.6% field base.

Model	Symbol	Side	Conf	FH / TF	Slot (UTC)	Exit	Net PnL
gemini-3.1-pro	BNB	Short	90	1h / 1h	Oct 30, 23:01	EXPIRY	-$0.65

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

claude-opus-5.5 (spliced) is both this month's hit-rate leader (49.6%) and the lowest-Brier line (Brier 0.2600, gap +9.0pp). Per-model ok-rate this window ran 97.31% to 100.00%. The full confidence-bucket analysis and the week-by-week gap drift live in Monthly Calibration #3.

Asset league: which coin the field reads best

A published week gives each coin 877 to 1,215 calls; the month gives 4,380 to 4,703 per coin. Across the month the field's hit rate ran from 50.2% on XRP down to 47.5% on BNB, a spread of +2.8pp against a 48.6% 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	4,431	50.2%	48.8%-51.7%	1
BTC	4,703	49.6%	48.2%-51.0%	2
ETH	4,475	48.0%	46.5%-49.5%	3
SOL	4,590	47.8%	46.4%-49.2%	4
BNB	4,380	47.5%	46.0%-48.9%	5

Ranked by hit rate, ties broken by n descending; Wilson 95% CIs are descriptive and 10 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	50.4% (2,762)	50.2% (1,435)	47.9% (234)	1h
BTC	51.5% (2,876)	47.1% (1,556)	44.3% (271)	1h
ETH	48.8% (2,748)	45.6% (1,447)	52.5% (280)	1d
SOL	48.4% (2,849)	45.0% (1,483)	56.6% (258)	1d
BNB	49.2% (2,707)	44.1% (1,446)	48.0% (227)	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 Oct 5-11	W2 Oct 12-18	W3 Oct 19-25	W4 Oct 26-Nov 1	Month
XRP	42.1% (5)	49.0% (1)	61.2% (1)	42.7% (2)	50.2% (1)
BTC	46.1% (1)	44.0% (5)	57.4% (3)	42.3% (3)	49.6% (2)
ETH	43.9% (3)	47.4% (3)	54.4% (4)	41.8% (4)	48.0% (3)
SOL	44.2% (2)	47.4% (2)	53.3% (5)	44.0% (1)	47.8% (4)
BNB	42.4% (4)	44.8% (4)	58.6% (2)	36.8% (5)	47.5% (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 1 of 4 weeks and 4 for the month; best horizon 1d (56.6% on n=258).

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 2.75 to 5.25, the widest span is 6 rank positions and the narrowest 1, and 4 distinct model lines held the weekly top spot across the four weeks.

Model	W1	W2	W3	W4	Month	Mean rank	Range
claude-opus-5.5 (spliced)	1 (47.0%)	5 (46.3%)	2 (58.0%)	6 (40.9%)	1 (49.6%)	3.50	5
gpt-6.1-sol (spliced)	3 (43.5%)	3 (46.5%)	3 (57.6%)	2 (42.5%)	2 (48.9%)	2.75	1
grok-4.7 (spliced)	5 (43.3%)	2 (47.0%)	7 (55.9%)	1 (42.5%)	3 (48.6%)	3.75	6
deepseek-v4-pro	7 (42.0%)	1 (48.2%)	4 (56.5%)	4 (41.5%)	4 (48.5%)	4.00	6
gemini-3.1-pro	4 (43.4%)	6 (45.8%)	1 (58.3%)	3 (41.7%)	5 (48.4%)	3.50	5
qwen-3.8-max	6 (43.1%)	4 (46.3%)	6 (56.3%)	5 (41.3%)	6 (48.4%)	5.25	2
claude-fable-5	2 (45.0%)	7 (45.6%)	5 (56.4%)	7 (40.4%)	7 (48.2%)	5.25	5

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 Oct 5-11, W2 Oct 12-18, W3 Oct 19-25, W4 Oct 26-Nov 1); 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 -0.68, W2-W3 -0.36, W3-W4 -0.11 (ties get average ranks). 4 distinct model lines 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 1.115% 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 Oct 5-11	1.130%	2	5	volatile
W2 Oct 12-18	1.113%	3	4	calm
W3 Oct 19-25	1.117%	1	6	volatile
W4 Oct 26-31	1.107%	1	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 (1.115%). BTC only: this is not the market check's "Mean |1d move| (5 assets)" row, which averages five coins over the whole window. W4 holds Oct 26-31 here: the month pull ends on Oct 31.

Model	Calm n	Calm hit	Volatile n	Volatile hit	Diff pp
claude-opus-5.5 (spliced)	1,090	43.7%	1,217	53.3%	+9.7
claude-fable-5	1,259	43.0%	1,381	51.3%	+8.4
gemini-3.1-pro	1,440	43.8%	1,592	51.5%	+7.8
gpt-6.1-sol (spliced)	1,563	44.7%	1,696	51.5%	+6.8
grok-4.7 (spliced)	1,439	44.8%	1,383	50.9%	+6.2
qwen-3.8-max	1,383	44.4%	1,536	50.3%	+5.9
deepseek-v4-pro	1,443	45.3%	1,652	49.8%	+4.5
Field (all models)	**9,617**	**44.3%**	**10,457**	**51.2%**	**+6.9**

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 Oct 5 to Nov 1 (2,505 calls on Oct 1-4) are not in either regime cell, so the two columns do not sum to the month.

Market check: October ran calmer than September

OBSERVATION — mean |1d move| 1.91% -> 1.57% (-17% rel), BTC realized vol 34.9% -> 32.4% (ann., hourly); field directional accuracy 45.5% -> 48.6% (+3.1pp), 6 of 7 models improved, sim win-rate up for 6 of 7, field sim PnL -$2554.15 -> -$1648.92 (Sep 1-30 vs Oct 1-31; descriptive, one pair of windows, not a claim).
Robustness: raw price-sign accuracy 48.4% -> 48.1% — the opposite direction to the trade-based hit rule.

Measure	Sep 1-30	Oct 1-31	Change
Mean |1d move| (5 assets)	1.91%	1.57%	-17% rel
BTC realized vol (ann., hourly)	34.9%	32.4%	-2.5pp
Field directional accuracy	45.5%	48.6%	+3.1pp
Models improving hit-rate	--	6 of 7	--
Raw price-sign accuracy	48.4%	48.1%	-0.3pp
Field sim win-rate	38.4%	39.6%	+1.2pp
Field sim net PnL	-$2554.15	-$1648.92	--

Window A is the pack's prior window -- the full September -- so this slice and every month-over-month column in this report compare October against the full September. 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: claude-opus-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17). 10,395 rows were remapped inside the month against 18,900 in the prior window.

Practical implications

- On this month's leaderboard no title was awarded, the top two CIs overlap, and 4 distinct model lines led the four Mon-Sun weeks from Oct 5 to Nov 1.
- XRP ranked #1 for the month on n=4,431; the week-by-week league shows it ranked #1 in 2 of the four published weeks.
- The field differs by +6.9pp between the two regimes on n=9,617 calm vs n=10,457 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 September: every "prior" column, the gray series of the page-1 chart and the market-check A column describe Sep 1-30. It is window A of the same pull, slots 2026-09-01 to 2026-10-01 (first slot actually seen 2026-09-01 00:01 UTC), matured at the prior cutoff Thu Oct 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as October -- 20,682 directional calls against 22,579 in the month. Month-over-month deltas are a step between two windows of different length (30 and 31 days).
- Two cutoffs live in this issue. The month is frozen at Sun Nov 1, 2026, 16:00 UTC; the four weekly blocks are frozen at their own published cutoffs (W1 Oct 5-11, W2 Oct 12-18, W3 Oct 19-25, W4 Oct 26-Nov 1). 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 (22,579 vs 20,635 directional, plus 2,505 on the 4 remainder days); W4 also runs to Nov 1, a day the month pull does not hold, so its pack counts 5,094 directional calls against 4,533 for Oct 26-31 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; 17 events in this window still cannot characterise detector performance. Cross-TF and cross-FH disagreement cells are small (n=140, 13, 153, 14); cells under N=10 are insufficient and never used to rank models.
- Model lines in this window: claude-fable-5, claude-opus-5.5 (spliced), deepseek-v4-pro, gemini-3.1-pro, gpt-6.1-sol (spliced), grok-4.7 (spliced), qwen-3.8-max. Series density and lineage notes (monthly #3): (1) rolling-1w forecasts emit daily since Aug 22, 2026 and rolling-1M daily since Aug 23, 2026, so across this 31-day window daily coverage is FULL as measured -- the 1w series has slots on 31 of 31 days and the 1M series on 31 of 31 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 2,160, 1M 2,135). (2) Model lineage is a splice inside this window. Three lines flipped at Oct 17, 2026 09:01 UTC, day 17 of 31 of the month: "gpt-6.1-sol (incl. 5.6-era, Oct 1-17)" (3,465 rows from the gpt-5.6-sol era, 3,045 from gpt-6.1-sol); "claude-opus-5.5 (incl. opus-5-era, Oct 1-17)" (3,465 rows from the claude-opus-5 era, 3,045 from claude-opus-5.5); "grok-4.7 (incl. 4.6-era, Oct 1-17)" (3,465 rows from the grok-4.6 era, 3,045 from grok-4.7). Every per-model aggregate sees the merged head id (the raw id is kept in the pack); 10,395 rows were remapped inside the month, and in the prior window 6,300 rows of each of the three lines (18,900 in all) are earlier-build rows, so each month-over-month row for those lines joins a spliced month to a single-build prior month -- a remapped comparison is not the same model build. (3) The two hourly slots before the flip (Oct 17, 07:01 and 08:01 UTC) hold 60 invalid rows (claude-opus-5 20, gpt-5.6-sol 20, grok-4.6 20), counted in the invalid counter as they are; in those two hours the other 4 lines (claude-fable-5, deepseek-v4-pro, gemini-3.1-pro, qwen-3.8-max) carry ok rows. (4) Past lineage flips: the grok 4.5 -> 4.6 flip (Aug 24, 2026 09:22 UTC) and the qwen-3.7-max -> qwen-3.8-max succession (Aug 5, 2026) both sit before 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 September: it covers 30 days (Sep 1-30), and the TOP5 volume figure compares a 31-day sum with a 30-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 (Sun Nov 1, 2026, 16:00 UTC) -- the D2 definition pinned in the weekly series; this issue prints 82,334. 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 (Oct 26-Nov 1) reproduces the published weekly #13 on 7 of 7 checks, and the prior window of this pack (Sep 1-30, prior cutoff Thu Oct 1, 2026, 16:00 UTC) reproduces the published Monthly #2 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	45,540	Mature (scored pool)	40,889
OK in gate	40,889	-- of them directional	22,579
Out of gate (1w / 1M)	2,160 / 2,135	-- of them sideways	18,310
Invalid	356	Pending (next issue)	0
Late closes	0	Uptime, all grids	1,178 / 1,178 slots
Source file	monthly_metrics_2026-10.json	Report cutoff	Sun Nov 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-11-03 12:39 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 22,579 scored observations. MARKETMANIA RESEARCH TO DATE (as of cutoff Nov 1, 16:00 UTC): 82,334 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 69 published reports.

Issue #3. 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 third issue reads October against the full September: claude-opus-5.5 (spliced) at the top (49.6%; 45.4% in the prior window), 8 of 17 qualifying turns found a caller, 0 of 5 tickers above 51%, 4 distinct model lines leading the four weeks and a calm-vs-volatile field difference of +6.9pp. 3 model lines are lineage splices this month (claude-opus-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17)). Engine 1.1 has powered the sandbox since Aug 18, i.e. before this window and before the prior window; no cross-engine PnL comparisons are claimed.

What we're testing next

- Monthly #4 (November): the same three series over Nov 1-30, cut on Dec 1 plus the maturity lag, so its month-over-month comparison runs against this October -- the first month whose prior window carries the cutover splice.
- Monthly Config Watch #3: the same October window through the config search engine, published as the fourth monthly report of this issue.
- Quarterly Stability Index: the next repeat-prompt stability run is the next cross-quarter point on model stability; its date is announced with the run.

Related research

Report	Direct PDF link
Monthly Consensus Watch #3	https://marketmania.ai/research/reports/consensus-watch-monthly-2026-10.pdf
Monthly Calibration #3	https://marketmania.ai/research/reports/calibration-monthly-2026-10.pdf
Weekly Model Watch #13	https://marketmania.ai/research/reports/model-watch-2026-10-26.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_2026m10,
  title  = {Monthly Model Watch #3: monthly leaderboard, asset league, rank stability and regime split,
Oct 1-31 2026},
  author = {{MarketMania Research}},
  year   = {2026}, month = {November}, day = {5},
  url    = {https://marketmania.ai/research/reports/model-watch-monthly-2026-10.pdf},
  note   = {Methodology v1.1 (2026-08-10), hash e66c7e8c864a2233; window Oct 1-31, 2026 UTC; source
monthly_metrics_2026-10.json}
}
```