

Monthly Consensus Watch #2

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

RESEARCH SNAPSHOT			
17,625 LOO-scoreable calls	7 models (stable lineup)	3,057 thin/tied excluded	window Sep 1-30, 2026 UTC
method = leave-one-out majority, 3+ directional peers	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)

Herdling ran at 99.2% of leave-one-out scoreable calls across the month and the pooled lift came in at -9.7pp: agree-hit 45.1% against 54.7% on disagreement, on a disagree bucket of n=137.

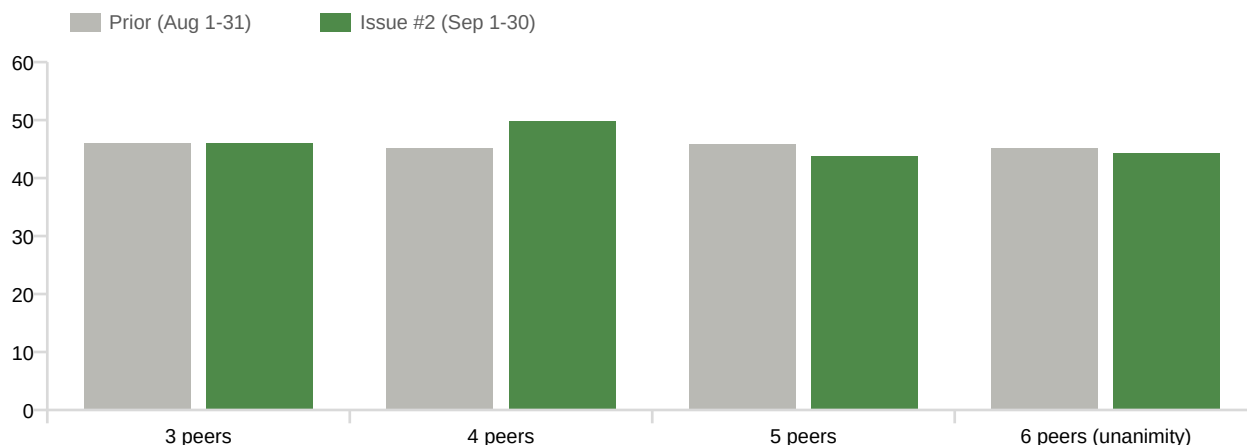
WEEKLY EVIDENCE -> MONTHLY VERDICT (rule below)

Metric	W1 Aug 31-Sep 6	W2 Sep 7-13	W3 Sep 14-20	W4 Sep 21-27	Month	Verdict
LOO lift, pp	-27.6	-30.4	+12.1	-14.0	-9.7	Confirmed (3/4)
Herd share >= 95%	99.3%	99.3%	98.9%	99.3%	99.2%	Confirmed (4/4)
Unanimity hit vs field hit	-7.1	-5.3	+0.8	+1.0	-1.2	Inconclusive (2/4)

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 -- 99.2% of the field's leave-one-out calls ran with the herd over the month.** Of 17,625 LOO-scoreable directional calls, 17,488 (99.2%) sided with peers' majority; agree-hit was 45.1% [44.3%, 45.8%] vs. 54.7% [46.4%, 62.8%] for the 137 disagreeers, a pooled lift of -9.7pp. Descriptive, one month, not a claim of a durable edge.
- **The margin curve read 45.9% / 49.7% / 43.8% / 44.3% from 3 peers to unanimity.** 2 of the 4 tiers sit above the 45.5% field base. The same four tiers read 45.9% / 45.1% / 45.8% / 45.1% in the prior window Aug 1-31, on n=1,964 / 2,630 / 4,482 / 6,664 there.
- **Week by week the lift changed sign 2 times.** The four published weeks ran Aug 31-Sep 6 -27.6pp, Sep 7-13 -30.4pp, Sep 14-20 +12.1pp, Sep 21-27 -14.0pp; the month, which also counts the 3 days outside those weeks, pools at -9.7pp and the prior window read +1.9pp. The month is one pooled pull, not the mean of the weeks: the weekly range is -30.4pp to +12.1pp.
- **1,965 cells had 6+ directional models on one side across the month.** The 5 worst sat at 69.6-70.4 mean confidence and resolved 0.0%-14.3%; the prior window Aug 1-31 had 1,672 such cells, hitting 45.4% on n=10,984.



Hit rate (%) by number of agreeing LOO peers, the prior window against the whole month. Monthly n: 3 peers 1,712 / 4 peers 2,400 / 5 peers 3,972 / 6 peers 9,317. Prior n: 3 peers 1,964 / 4 peers 2,630 / 5 peers 4,482 / 6 peers 6,664. Field base this month 45.5%, prior window 46.3%. 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 prior 6-peer tier is the published Monthly #1's six-peer bucket (n=6,664), the same strict leave-one-out definition as this month's; the pack's prior block carries the wider unanimity rule instead (every directional call of a cell with 6 or more models on one side, n=10,984, 45.4%), which is the Unanimity column of the week-by-week table, not a margin tier.

Why it matters

MarketMania publishes forecasts from 7 frontier model lines into the same slots every hour; a natural question for anyone reading the feed is whether a model's call is worth more when it agrees with what everyone else is saying, or whether that agreement is just everyone making the same mistake together. The four weekly issues set against this month answered it with lifts of Aug 31-Sep 6 -27.6pp, Sep 7-13 -30.4pp, Sep 14-20 +12.1pp, Sep 21-27 -14.0pp, and the pooled month reads -9.7pp. This monthly issue asks the same question at monthly n, across four weeks the pack labels as 2 calm and 2 volatile, using the same leave-one-out design built to avoid comparing a model to an index that includes its own vote.

How to read this

For every model M and every forecast it made, we rebuild that slot's consensus using only the OTHER active models in the same symbol / forecast-horizon (FH) / timeframe (TF) cell -- M's own call never counts toward its own consensus. M is scored agree if its side (long or short) matches the strict majority of those peers, and disagree if it sits alone against that majority; a cell needs at least 3 directional peers to count at all, and tied peer splits are excluded rather than forced either way. LOO peers are genuinely external to the model being scored, so the comparison below is not comparing a model to itself. No model line is lineage-spliced this month (the prior window carries the grok and qwen splices), and a spliced line never counts as two peers in a cell. The month is one pull at one cutoff (Thu Oct 1, 2026, 16:00 UTC); the four weekly blocks quoted alongside it are the published issues at their own frozen cutoffs.

Pooled result: agree vs. disagree with the LOO consensus

Group	n	Hit rate	Wilson 95% CI	Lift, pp
Agree with LOO peers	17,488	45.1%	[44.3%, 45.8%]	-9.7 (vs. disagree)
Disagree with LOO peers	137	54.7%	[46.4%, 62.8%]	ref.

The disagree bucket is n=137 over the whole month; the two Wilson intervals do not overlap. Prior window Aug 1-31: agree 45.5% / disagree 43.6%, lift +1.9pp, on n=15,878 agree and n=211 disagree. These are dependent observations pooled across the whole month: two calm weeks, two volatile weeks and 3 days outside them.

By forecast horizon / timeframe

The pooled result above mixes 5 different FH/TF cells. Broken out:

FH / TF	Agree n	Agree hit	Dis. n	Dis. hit	Lift, pp	Excluded
1h / 1h	10,928	46.9%	106	54.7%	-7.8	1,974
4h / 4h	2,683	44.4%	11	63.6%	-19.2	445
4h / 1h	2,826	42.4%	15	53.3%	-10.9	481
1d / 1d	558	34.1%	5	40.0%	-5.9	102
1d / 4h	493	35.3%	0	n/a	n/a	55

"Excluded" = thin (fewer than 3 directional peers) or tied peer splits in that cell; not folded into agree or disagree. "Dis." = disagree. Cells whose disagree bucket is under N=10 are insufficient and are never used to rank anything.

Margin: does more agreement mean a safer call?

For every LOO-scored forecast that sided with its peers' majority (the agree bucket, n=17,488) we also count how many peers were on that majority side -- from the bare minimum of 3 up to the full peer set agreeing (6 of 6, since 7 model lines were active and LOO always drops one); the 137 disagreeing calls sit in no tier, so the tiers sum to the agree bucket, not to the 17,625 scoreable calls. The month-over-month chart is on page 1; the full per-tier numbers: 3 peers 45.9% (n=1,712), 4 peers 49.7% (n=2,400), 5 peers 43.8% (n=3,972), 6 peers 44.3% (n=9,317). Against the 45.5% field base: 2 tiers above, 2 tiers below; in the prior window the same tiers read 45.9% / 45.1% / 45.8% / 45.1%. The week-by-week table further down prints each tier for each of the four weeks. The prior 6-peer tier is the published Monthly #1's six-peer bucket (n=6,664), the same strict leave-one-out definition as this month's; the pack's prior block carries the wider unanimity rule instead (every directional call of a cell with 6 or more models on one side, n=10,984, 45.4%), which is the Unanimity column of the week-by-week table, not a margin tier.

Daily-horizon herds: the thinnest cells in the report

The two daily cells stay the report's thinnest even at monthly n. 1d/1d unanimity hit 32.1% (n=371) and 1d/4h unanimity 37.0% (n=273) against the 45.5% field base; the whole 1d/1d agree bucket ran 34.1% (n=558) -- not the highest of the five FH/TF cells -- while 1d/4h ran 35.3% (n=493). Both daily buckets are an order of magnitude thinner than the hourly one (n=10,928), so they are reported, not ranked.

Herd failures: the 5 worst unanimous misses

When the whole field agreed, and was wrong. This month 1,965 slot / symbol / FH / TF cells had 6 or more directional models sitting on the same side.

Slot (UTC)	Symbol	FH / TF	Side	Models	Mean conf	Hit rate
Sep 22, 00:01	SOL	1d / 1d	long	7	70.4	0.0%
Sep 26, 00:01	SOL	1d / 1d	long	7	70.4	0.0%
Sep 18, 20:01	XRP	4h / 4h	long	7	70.0	0.0%
Sep 21, 01:01	SOL	1h / 1h	long	7	69.7	14.3%
Sep 16, 09:01	BNB	1h / 1h	short	7	69.6	0.0%

The 5 worst cells were long/short swarms at 69.6-70.4 mean confidence and resolved 0.0%-14.3%. 'Sideways' was 48.0% of mature forecasts (19,124 of 39,806) this month.

Per-model: agree vs. disagree

Per-model disagree counts run 4 to 72 over the month. 3 of the 7 lines clear the N=10 floor (deepseek-v4-pro 30, gemini-3.1-pro 72, grok-4.6 14); every other per-model contrarian cell below is insufficient.

Model	Agree n	Agree hit	Disagree n	Disagree hit
claude-fable-5	2,587	45.5%	6	16.7%
claude-opus-5	2,347	45.3%	4	25.0%
deepseek-v4-pro	2,623	44.7%	30	56.7%
gemini-3.1-pro	2,522	47.2%	72	62.5%
gpt-5.6-sol	2,660	45.1%	7	57.1%
grok-4.6	2,097	43.5%	14	35.7%
qwen-3.8-max	2,652	43.9%	4	50.0%

Prior window Aug 1-31: per-model agree-side hit ran 44.5% to 47.3% on n=2,065 to 2,459 per line. At these counts a handful of flipped calls moves a per-model lift by tens of points, so models are not ranked by contrarian lift here. 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).

Week by week: how often the field agreed, and what it paid

The same leave-one-out scoring run on each of the four Mon-Sun weeks from Aug 31 to Sep 27 gives four weekly lifts; the pooled monthly number also counts the 3 days outside those weeks. Herd share ran 98.9% to 99.3% and the lift ran -30.4pp to +12.1pp across the four weeks against -9.7pp for the month as a whole.

Week	Scoreable	Herd share	Agree hit	Disagree hit	Lift pp	Unanimity n	Unanimity hit
W1 Sep 1-6	3,191	99.3%	43.3%	69.6%	-26.3	2,214	40.3%
W2 Sep 7-13	3,788	99.3%	38.8%	69.2%	-30.4	2,761	36.6%
W3 Sep 14-20	4,877	98.9%	50.6%	38.5%	+12.1	3,617	49.6%
W4 Sep 21-27	4,245	99.3%	46.0%	60.0%	-14.0	3,393	47.7%
Month	17,625	99.2%	45.1%	54.7%	-9.7	13,121	44.2%

Source: the month pull split by slot week at the single monthly cutoff (Thu Oct 1, 2026, 16:00 UTC), so the four week rows and the Month row are one consistent pool. The frozen-cutoff cross-check is the published weekly issues at their own cutoffs: the recomputed lift reproduces the published lift for 3 of 4 weeks (W1 -26.3pp vs -27.6pp, W2 -30.4pp vs -30.4pp, W3 +12.1pp vs +12.1pp, W4 -14.0pp vs -14.0pp). W1 holds Sep 1-6 here: the month pull starts on Sep 1. * = N<10 (insufficient) -- such cells are never ranked.

Margin	W1	W2	W3	W4	Month
3 peers	48.4% (384)	42.8% (400)	52.1% (424)	39.1% (304)	45.9% (1,712)
4 peers	53.3% (525)	46.9% (550)	55.5% (685)	39.0% (485)	49.7% (2,400)
5 peers	48.2% (732)	40.4% (984)	44.2% (990)	46.8% (876)	43.8% (3,972)
6 peers (unanimity)	36.5% (1,512)	34.2% (1,813)	51.6% (2,681)	47.8% (2,541)	44.3% (9,317)

Cells are hit rate with n in parentheses; * = N<10 (insufficient), "n/a (0)" = no call landed in that tier that week. Weeks are W1 Sep 1-6, W2 Sep 7-13, W3 Sep 14-20, W4 Sep 21-27. Same source and cutoff as the table above; the Month column is the pooled month, not the mean of the four weeks. Month-over-month columns compare back-to-back windows: Aug 1-31 vs Sep 1-30 (this issue). "Was" values are the pack's own prior-window block -- window A of the same pull, same engine, same FH gate and same hit rule, matured at the prior cutoff Tue Sep 1, 2026, 16:00 UTC. The prior month is the full August (31 days against 30 in September), and it holds 19,383 directional calls against 20,682 in September. Deltas are computed on unrounded rates.

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

- 2 of the 4 margin tiers sit above the 45.5% field base this month; the four tier rates span 43.8% to 49.7%, a spread of +5.9pp, and the four weeks set against the month ran lifts of -30.4pp to +12.1pp against a pooled monthly lift of -9.7pp.
- The pooled monthly lift is -9.7pp while the four weeks ran -30.4pp to +12.1pp and the prior window read +1.9pp.
- The field base moved -0.8pp from August to September (46.3% -> 45.5%), and every consensus tier is measured against that base.

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 are not independent: the same models watch overlapping symbol / FH / TF cells hour after hour for 30 days. Wilson intervals here are descriptive, not inferential, and pooling a month does not make them inferential.
- The pooled lift rests on 137 disagreeing calls; per-model counts run 4-72. 3,057 calls (14.8% of the 20,682-call mature directional universe) were thin or tied and are excluded from every consensus table. Correlational, not causal: LOO removes self-match bias, not the confound that an easy slot lifts agreement and hit together.
- This window ran 10 trend days of 30, and the pack's own volatility rule splits the four Mon-Sun weeks from Aug 31 to Sep 27 as W1 volatile, W2 calm, W3 volatile, W4 calm; every month-over-month comparison is a comparison across regimes as well as across windows. 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: 17,625 LOO-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 · 39 published reports.

Issue #2. Monthly Consensus Watch is a living monthly comparison; each issue appends one more calendar month of leave-one-out agreement-vs-hit data. This second issue reads September against the full August: herd share 99.2%, pooled lift -9.7pp (agree 45.1% vs disagree 54.7% on n=137), four weekly lifts set against the month running -30.4pp to +12.1pp, and the prior window Aug 1-31 at herd share 98.7% and lift +1.9pp. 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 Calibration #2	https://marketmania.ai/research/reports/calibration-monthly-2026-09.pdf
Monthly Model Watch #2	https://marketmania.ai/research/reports/model-watch-monthly-2026-09.pdf
Consensus Watch #8	https://marketmania.ai/research/reports/consensus-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_consensus_monthly_2026m09,
  title = {Monthly Consensus Watch #2: does agreeing with the crowd make an LLM's market call safer?},
  author = {{MarketMania Research}},
  year = {2026}, month = {October}, day = {2},
  url = {https://marketmania.ai/research/reports/consensus-watch-monthly-2026-09.pdf},
  note = {Methodology v1.1; window Sep 1-30, 2026 UTC; source monthly_metrics_2026-09.json}
}
```