

Quarterly Model Watch #1

October 8, 2026 · MarketMania Research · Quarterly series · window Jul 15-Sep 30, 2026 UTC

RESEARCH SNAPSHOT			
59,606 directional calls scored	7 models (lineage spliced: opus, grok, qwen)	of 122,770 mature forecasts	window Q3 2026 · Jul 15-Sep 30 · partial
hit = direction rule: tp1/tp2->hit, sl->miss, expiry->sign of gross pnl	FH gate 1h / 4h / 1d	base field hit 45.6%	cutoff Thu Oct 1, 2026, 16:00 UTC
market: BTC net +28.56% (prior n/a)	ann vol 38.9% (was n/a)	TOP5 volume \$173.94B, n/a q/q	pairwise corr 0.77 (was n/a)

KEY FINDING · OBSERVATION (one quarterly window)

Across the three months the field hit 45.6% (Jul 15-31 45.0%, Aug 1-31 46.3%, Sep 1-30 45.5%), 3 distinct model lines led the three months, and no quarterly title was awarded.

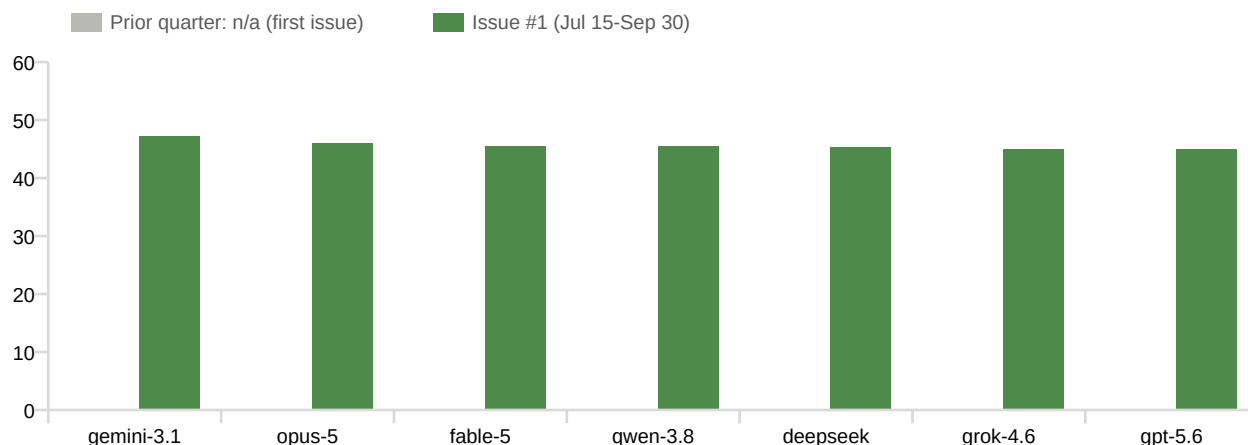
MONTHLY EVIDENCE -> QUARTERLY VERDICT (rule below)

Metric	M1 Jul 15-31	M2 Aug 1-31	M3 Sep 1-30	Quarter	Verdict
Leader (hit-rate)	qwen-3.8	opus-5	gemini-3.1	gemini-3.1	Rejected (1/3)
Most readable coin	ETH	BNB	XRP	BNB	Rejected (1/3)
Field hit below 50%	45.0%	46.3%	45.5%	45.6%	Confirmed (3/3)

Verdict rule: each row states one reading of the quarter; a month agrees when it shows the same reading (same sign, or the same name). Confirmed = 2 or 3 of the 3 months agree with the quarter; Rejected = 0 or 1. M1 holds Jul 15-31 only: the platform's first forecast slot is Jul 15 18:01 UTC. Monthly cells come from the frozen monthly packs (7-line roster, lineage spliced); the quarter is the pooled Jul 15-Sep 30 pull.

TL;DR

- **OBSERVATION -- No quarterly title.** gemini-3.1-pro leads the quarter descriptively at 47.2% [46.1%, 48.2%]; the gap to #2 (claude-opus-5 (spliced), 45.9%) is 1.3pp. There is no prior quarter; the monthly top spot went to qwen-3.8-max (spliced) in Jul 15-31, claude-opus-5 (spliced) in Aug 1-31 and gemini-3.1-pro in Sep 1-30, so 3 distinct model lines held it across the three months.
- **No ticker cleared 51%.** BNB is the quarter's most readable coin (46.3%, n=10,571) and SOL the hardest at 44.9%; 0 of the 5 assets cleared 51% over the quarter. Pair of the quarter: gemini-3.1-pro x XRP 48.0% (n=1,746); worst: deepseek-v4-pro x SOL 43.7% (n=1,509).
- **Mean monthly rank ran 2.00 to 5.67 across the three months.** grok-4.6 (spliced) swung across 5 rank positions and deepseek-v4-pro across 2. Consecutive-month Spearman came in at M1-M2 -0.43, M2-M3 0.82.
- **Calm and volatile months read the same.** Split by the pack's own volatility rule the field hit 45.0% (n=19,541) in the one calm month against 45.9% (n=40,065) in the two volatile months, a difference of +0.9pp. Cross-TF/FH lifts ran -3.2pp to -0.4pp. On trend days, 4h TF-agreement hit 57.2% (n=2,303) against 36.0% (n=3,626) on flat days.



Directional hit-rate (%) by model over the whole quarter, ranked by the quarter'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 45.6%; by month 45.0% / 46.3% / 45.5% (Jul 15-31 / Aug 1-31 / Sep 1-30). First quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026), so the chart carries no gray series and every prior cell in this report reads n/a; a month inside the quarter is not a prior window.

Why it matters

MarketMania scores 7 model lines against the same market, hour after hour. Quarterly Model Watch asks the monthly questions at quarterly n -- who led, which tickers were readable, who saw the turn first, does self-agreement help -- and then asks the three questions a single month cannot answer: which coin the field reads best over a quarter, how far a monthly rank moves inside one quarter, and how the field's accuracy differs between the calm and volatile months 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 quarter at one cutoff (Thu Oct 1, 2026, 16:00 UTC). The quarterly title is awarded only on a ≥ 5 pp gap AND non-overlapping 95% Wilson CIs with $N \geq 10$ per side -- a descriptive lead is not a title. Reversals use the 4h grid with a $\geq 2.0\%$ 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-month blocks come from the three frozen monthly packs at their own published cutoffs. 3 model lines are lineage splices this quarter (claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5)).

Leaderboard of the quarter

Model	N	Cov.	Hit rate	95% CI	Q/q pp
gemin-3.1-pro	9,130	51.4%	47.2%	46.1%-48.2%	n/a
claude-opus-5 (spliced)	8,261	46.3%	45.9%	44.8%-46.9%	n/a
claude-fable-5	8,447	47.3%	45.5%	44.5%-46.6%	n/a
qwen-3.8-max (spliced)	9,537	53.7%	45.4%	44.4%-46.4%	n/a
deepseek-v4-pro	7,799	43.7%	45.3%	44.2%-46.4%	n/a
grok-4.6 (spliced)	7,244	45.9%	45.0%	43.9%-46.2%	n/a
gpt-5.6-sol	9,188	51.4%	44.9%	43.9%-45.9%	n/a

No quarterly title is awarded this issue. Title rule: ≥ 5 pp gap AND non-overlapping 95% CI, $N \geq 10$ per side. gemini-3.1-pro's lead over claude-opus-5 (spliced) is 1.3pp and their CIs overlap, so neither condition is met. The Q/q pp column reads n/a in every row because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026). claude-opus-5 (spliced) = claude-opus-5 (incl. 4.8-era, Jul 15-30). grok-4.6 (spliced) = grok-4.6 (incl. 4.5-era, Jul 15-Aug 24). qwen-3.8-max (spliced) = qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5).

Ticker of the quarter

Symbol	N	Hit rate	95% CI	Prior
BNB	10,571	46.3%	45.3%-47.2%	n/a
XRP	11,196	46.3%	45.3%-47.2%	n/a
BTC	12,224	45.5%	44.6%-46.4%	n/a
ETH	12,232	45.5%	44.6%-46.4%	n/a
SOL	11,791	44.9%	44.0%-45.8%	n/a

BNB was the field's most readable ticker (46.3%) and SOL the hardest (44.9%); 0 of 5 assets cleared 51% this quarter. The Prior column reads n/a in every row because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026). The quarter-by-month ticker history is in the asset-league section below. This issue's ordering is BNB > XRP > BTC > ETH > SOL; only the five current coins are ranked (the retired coins of Jul 15-31 are in Limitations).

Pair of the quarter

Top 6 of the 35 model x ticker pairs tracked this quarter (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 on one of the five current coins.

Model (top 6)	Symbol	N	Hit rate
gemini-3.1-pro	XRP	1,746	48.0%
gemini-3.1-pro	BNB	1,630	47.8%
gemini-3.1-pro	BTC	1,887	47.3%
gemini-3.1-pro	SOL	1,817	46.8%
claude-opus-5 (spliced)	ETH	1,727	46.8%
deepseek-v4-pro	BNB	1,412	46.6%

Model (5 worst)	Symbol	N	Hit rate
gpt-5.6-sol	ETH	1,887	44.4%
deepseek-v4-pro	ETH	1,659	44.2%
gpt-5.6-sol	SOL	1,778	43.8%
grok-4.6 (spliced)	SOL	1,473	43.7%
deepseek-v4-pro	SOL	1,509	43.7%

gemini-3.1-pro x XRP (48.0%, n=1,746) is the quarter's best pair and deepseek-v4-pro x SOL (43.7%, n=1,509) the weakest. Pair history accumulates across issues -- read this as the first quarterly 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	Mon Jul 20, 08:00	Long	+2.32%	23	claude-opus-5 (spliced)
BTC	Mon Jul 27, 20:00	Short	-2.48%	3	grok-4.6 (spliced)
BTC	Mon Aug 3, 08:00	Long	+2.10%	9	qwen-3.8-max (spliced)
BTC	Fri Aug 28, 12:00	Short	-2.49%	8	gpt-5.6-sol
BTC	Wed Sep 23, 08:00	Short	-2.90%	1	gemini-3.1-pro
ETH	Mon Jul 20, 08:00	Long	+2.38%	32	qwen-3.8-max (spliced)
ETH	Mon Jul 27, 16:00	Short	-2.27%	0	NOBODY
ETH	Tue Jul 28, 12:00	Long	+2.23%	0	NOBODY
ETH	Sat Aug 1, 20:00	Long	+2.27%	0	NOBODY
ETH	Mon Aug 10, 08:00	Short	-2.65%	0	NOBODY
ETH	Sat Aug 22, 00:00	Short	-3.46%	1	gemini-3.1-pro
ETH	Sun Aug 23, 08:00	Long	+2.14%	17	claude-fable-5
ETH	Wed Aug 26, 16:00	Long	+2.31%	13	gpt-5.6-sol
ETH	Fri Aug 28, 12:00	Short	-2.80%	0	NOBODY
ETH	Sun Aug 30, 16:00	Short	-2.52%	1	deepseek-v4-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 long turn drew 23 callers, first claude-opus-5 (spliced) on a 4h/4h call at Jul 19, 20:01 with confidence 58; the BTC short turn drew 3 callers, first grok-4.6 (spliced) on a 4h/1h call at Jul 27, 08:01 with confidence 57; the BTC long turn drew 9 callers, first qwen-3.8-max (spliced) on a 4h/1h call at Aug 2, 20:01 with confidence 68; the BTC short turn drew 8 callers, first gpt-5.6-sol on a 4h/1h call at Aug 28, 04:01 with confidence 62; 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 32 callers, first qwen-3.8-max (spliced) on a 4h/1h call at Jul 19, 20:01 with confidence 55; the ETH short turn drew NOBODY; the ETH long turn drew NOBODY; the ETH long turn drew NOBODY; the ETH short turn drew NOBODY; the ETH short turn drew 1 caller, first gemini-3.1-pro on a 4h/1h call at Aug 22, 00:01 with confidence 60; the ETH long turn drew 17 callers, first claude-fable-5 on a 1d/1d call at Aug 23, 00:01 with confidence 62; the ETH long turn drew 13 callers, first gpt-5.6-sol on a 1d/4h call at Aug 27, 00:01 with confidence 69; the ETH short turn drew NOBODY; the ETH short turn drew 1 caller, first deepseek-v4-pro on a 4h/4h call at Aug 30, 04:01 with confidence 56; 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; 21 events in this window, against 5 / 9 / 7 in the three frozen monthly packs (M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30), and 13 of the 21 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 8 cases (BTC Jul 20 by 11h 59min; BTC Jul 27 by 11h 59min; BTC Aug 3 by 11h 59min; BTC Aug 28 by 7h 59min; ETH Jul 20 by 11h 59min; ETH Aug 23 by 7h 59min; ETH Aug 30 by 11h 59min; ETH Sep 20 by 11h 59min) and after it in 5 cases (BTC Sep 23 by 16h 01min; ETH Aug 22 by 1 min; ETH Aug 26 by 8h 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	5,937	44.3%	314	45.5%	44.7% (8,798)	-0.4
cross-TF	1d	TF 1d vs 4h	1,102	38.8%	75	52.0%	40.9% (1,677)	-2.2
cross-FH	4h	by FH 1h	5,079	43.7%	321	45.8%	44.7% (8,798)	-1.0
cross-FH	1d	by FH 4h	938	37.7%	84	54.8%	40.9% (1,677)	-3.2

* = 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): 2527 / 499 / 2767 / 655; primaries with no partner call stored at that slot, in no bucket: 20 / 1 / 631 / 0 -- so agree + disagree + mixed + unpaired = the base n of each family. Pooled across all four cell families the lifts run -3.2pp to -0.4pp this quarter.

Regime control. Methodology calls for a trend-vs-flat split on this table. This window has 24 trend days of 77 (a day is a trend day when its absolute BTC 1d move is 1.5% or more, else a flat day), the days of Jul 1-15 carrying no BTC daily move, and the three months label as M1 calm, M2 volatile, M3 volatile under the pack's rule (a month is volatile when its mean absolute BTC 1d move is at or above the median of the three months, else calm; median mean |BTC 1d move| 1.317%). On trend days, 4h TF-agreement hit 57.2% (n=2,303) against 36.0% (n=3,626) on flat days.

Side-mix and the quarter's failure

Model	N	Long	Short	Sideways	L:S
claude-fable-5	17,887	32.5%	14.8%	52.7%	2.19
claude-opus-5 (spliced)	17,890	30.9%	15.5%	53.6%	2.00
deepseek-v4-pro	17,870	23.7%	20.1%	56.3%	1.18
gemini-3.1-pro	17,769	29.7%	21.7%	48.5%	1.37
gpt-5.6-sol	17,897	26.5%	24.9%	48.5%	1.06
grok-4.6 (spliced)	15,815	21.3%	24.6%	54.0%	0.86
qwen-3.8-max (spliced)	17,791	31.1%	22.6%	46.3%	1.38

L:S = each model's own long-share divided by its short-share; the field average of the per-model ratios is 1.44 this quarter.

'Sideways' was 51.4% of mature forecasts (63,164 of 122,770). Consensus skew: 6,758 symbol/FH/TF cells had 5 or more models on the same side and hit 45.1% (n=44,406) against the 45.6% field base.

Model	Symbol	Side	Conf	FH / TF	Slot (UTC)	Exit	Net PnL
gemini-3.1-pro	XRP	Long	85	4h / 1h	Aug 21, 16:01	SL	-\$3.10

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

Calibration bridge

The hit-rate leader this quarter is gemini-3.1-pro (47.2%); the lowest-Brier line is grok-4.6 (spliced) (Brier 0.2716, gap +15.0pp). Per-model ok-rate this window ran 88.14% to 99.30%. The full confidence-bucket analysis and the month-by-month gap drift live in Quarterly Calibration #1.

Asset league: which coin the field reads best

A published month gives each coin 2,720 to 4,564 calls; the quarter gives 10,571 to 12,232 per coin. Across the quarter the field's hit rate ran from 46.3% on BNB down to 44.9% on SOL, a spread of +1.4pp against a 45.6% field base. The ordering below is the quarter pull; the third table shows how often that ordering held from month to month.

Symbol	N	Hit rate	95% CI	Rank
BNB	10,571	46.3%	45.3%-47.2%	1
XRP	11,196	46.3%	45.3%-47.2%	2
ETH	12,232	45.5%	44.6%-46.4%	3
BTC	12,224	45.5%	44.6%-46.4%	4
SOL	11,791	44.9%	44.0%-45.8%	5

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

Symbol	1h	4h	1d	Best FH
BNB	48.2% (6,741)	43.8% (3,249)	38.0% (581)	1h
XRP	46.8% (7,209)	45.9% (3,379)	41.4% (608)	1h
ETH	46.8% (8,121)	43.9% (3,468)	38.6% (643)	1h
BTC	47.1% (7,949)	43.6% (3,649)	37.4% (626)	1h
SOL	44.6% (7,460)	45.9% (3,635)	43.1% (696)	4h

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	M1 Jul 15-31	M2 Aug 1-31	M3 Sep 1-30	Quarter
BNB	45.5% (2)	48.1% (1)	45.1% (4)	46.3% (1)
XRP	44.5% (3)	46.7% (3)	47.0% (1)	46.3% (2)
ETH	47.8% (1)	42.9% (5)	45.4% (3)	45.5% (3)
BTC	44.2% (4)	46.5% (4)	46.0% (2)	45.5% (4)
SOL	43.2% (5)	47.3% (2)	44.0% (5)	44.9% (5)

Cells are hit rate with the month's rank in parentheses. The three monthly columns are the PUBLISHED monthly numbers -- each month at its own frozen cutoff; the Quarter column is the single quarter pull. SOL ranked first in 0 of 3 months and 5 for the quarter; best horizon 4h (45.9% on n=3,635).

Rank stability across the three months

Every model's rank in each of the three published months, its rank for the pooled quarter, its mean rank and the span between its best and worst month, with the hit rate next to every rank. Mean rank runs 2.00 to 5.67, the widest span is 5 rank positions and the narrowest 2, and 3 distinct model lines held the monthly top spot across the three months.

Model	M1	M2	M3	Quarter	Mean rank	Range
gemini-3.1-pro	3 (45.1%)	2 (47.6%)	1 (48.5%)	1 (47.2%)	2.00	2
claude-opus-5 (spliced)	5 (44.9%)	1 (47.8%)	2 (45.4%)	2 (45.9%)	2.67	4
claude-fable-5	6 (44.5%)	3 (46.7%)	3 (45.4%)	3 (45.5%)	4.00	3
qwen-3.8-max (spliced)	1 (45.7%)	5 (45.6%)	5 (44.7%)	4 (45.4%)	3.67	4
deepseek-v4-pro	4 (45.1%)	6 (45.5%)	4 (45.3%)	5 (45.3%)	4.67	2
grok-4.6 (spliced)	2 (45.3%)	7 (45.3%)	7 (44.4%)	6 (45.0%)	5.33	5
gpt-5.6-sol	7 (44.3%)	4 (45.9%)	6 (44.5%)	7 (44.9%)	5.67	3

Rank = position in that window's leaderboard, which the pack sorts by hit rate descending; hit rate in parentheses. Months are the published windows at their own frozen cutoffs (M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30); the Quarter column is the single quarter pull, so the quarter rank is not the mean of the three monthly ranks. Consecutive-month Spearman rho: M1-M2 -0.43, M2-M3 0.82 (ties get average ranks). 3 distinct model lines held the monthly top spot. Sorted by quarter rank.

Calm vs volatile months

The pack labels each of the three months from the BTC daily series: a month is volatile when its mean absolute BTC 1d move is at or above the median of the three months, else calm, against a median mean |BTC 1d move| of 1.317% across the three months; 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 1 month calm and 2 months volatile, which splits the three months into two regime cells.

Month	Mean BTC 1d move	Trend days	Flat days	Label
M1 Jul 15-31	1.167%	5	11	calm
M2 Aug 1-31	1.390%	9	22	volatile
M3 Sep 1-30	1.317%	10	20	volatile

Mean |BTC 1d move| = the mean absolute daily BTC move of that month, from the entry-price series the pack reconstructs; the label is the pack's rule against the three-month median (1.317%). BTC only: this is not the market check's "Mean |1d move| (5 assets)" row, which averages five coins over the whole window. M1 rests on the 16 days with a BTC daily move.

Model	Calm n	Calm hit	Volatile n	Volatile hit	Diff pp
gemini-3.1-pro	2,813	45.1%	6,317	48.1%	+3.0
claude-opus-5 (spliced)	3,448	44.9%	4,813	46.5%	+1.6
claude-fable-5	2,690	44.5%	5,757	46.0%	+1.5
gpt-5.6-sol	3,085	44.3%	6,103	45.1%	+0.8
deepseek-v4-pro	1,869	45.1%	5,930	45.4%	+0.3
grok-4.6 (spliced)	2,265	45.3%	4,979	44.9%	-0.4
qwen-3.8-max (spliced)	3,371	45.7%	6,166	45.2%	-0.5
Field (all models)	19,541	45.0%	40,065	45.9%	+0.9

Definition: diff_pp = volatile hit_rate - calm hit_rate, in percentage points. Sorted by Diff pp descending, Field row last. * = N<10 (insufficient). No row falls outside the three months of the quarter, so the two columns sum to the quarter (19,541 + 40,065 = 59,606).

Market check: a calm July, then two volatile months

OBSERVATION -- mean |1d move| 1.54% / 1.98% / 1.91% by month (1.81% over the quarter), BTC realized vol 33.5% / 36.9% / 34.9% (35.1%; ann., hourly); field directional accuracy 45.0% / 46.3% / 45.5% (45.6%), sim win-rate 31.9% / 35.8% / 38.4% (35.4%), field sim PnL -\$3095.02 / -\$1054.79 / -\$2554.15 (-\$6703.95) (Jul 15-31 / Aug 1-31 / Sep 1-30, quarter in parentheses; descriptive, three windows, not a claim).

Robustness: raw price-sign accuracy 46.8% / 46.0% / 48.4% by month (47.1% over the quarter) against 45.0% / 46.3% / 45.5% (45.6%) under the trade-based hit rule; month to month the two moved in the same direction in 0 of 2 steps.

Measure	Jul 15-31	Aug 1-31	Sep 1-30	Quarter
Mean 1d move (5 assets)	1.54%	1.98%	1.91%	1.81%
BTC realized vol (ann., hourly)	33.5%	36.9%	34.9%	35.1%
Field directional accuracy	45.0%	46.3%	45.5%	45.6%
Models improving hit-rate	--	5 of 7	1 of 7	--
Raw price-sign accuracy	46.8%	46.0%	48.4%	47.1%
Field sim win-rate	31.9%	35.8%	38.4%	35.4%
Field sim net PnL	-\$3095.02	-\$1054.79	-\$2554.15	-\$6703.95

There is no prior quarter, so this slice prints the three months and the quarter side by side instead of a prior-against-current pair; each month is its own pack's window at its own frozen cutoff and the Quarter column is the quarter pull. Descriptive, three 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). "Models improving" compares a month with the month before it (Aug 1-31 against Jul 15-31; Sep 1-30 against Aug 1-31). 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 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5). 24,922 rows were remapped inside the quarter (19,532 / 5,390 / 0 in Jul 15-31 / Aug 1-31 / Sep 1-30).

Practical implications

- On this quarter's leaderboard no title was awarded, the top two CIs overlap, and 3 distinct model lines led the three months of the quarter.
- BNB ranked #1 for the quarter on $n=10,571$; the month-by-month league shows it ranked #1 in 1 of the three published months.
- The field differs by +0.9pp between the two regimes on $n=19,541$ calm vs $n=40,065$ volatile; the rule that draws the line is the pack's own three-month median, recomputed every quarter.

Limitations

- No prior quarter: this is the first quarterly issue (platform data begin Jul 15, 2026), so every "prior" cell reads n/a and the page-2 chart carries no gray series; a month inside the quarter is not a prior window. Partial quarter: the window is the calendar quarter Jul 1-Sep 30; the platform's first forecast slot is Jul 15 18:01 UTC, so 78 of 92 days carry data. Asset universe: the quarter's rows also carry 1,592 directional calls on AVAX (414), DOGE (347), LINK (436), PEPE (394) and TON (1), all in Jul 15-31; they are counted in the field totals but not ranked in the asset tables.
- Two cutoffs live in this issue. The quarter is frozen at Thu Oct 1, 2026, 16:00 UTC; the three monthly blocks are frozen at their own published cutoffs (M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30). A forecast still pending at its month's cutoff is immature in the monthly pack and mature in the quarter pull, so the quarter need not equal the sum of the months; here the two agree (59,606 vs 59,606 directional; no day of the quarter falls outside the three months): the quarter pull split by slot month gives 19,541 / 19,383 / 20,682, the same as the frozen monthly packs.
- 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; 21 events in this window still cannot characterise detector performance. Cross-TF and cross-FH disagreement cells are small ($n=314$, 75, 321, 84); cells under $N=10$ are insufficient and never used to rank models.
- Model lines in this window: claude-fable-5, claude-opus-5 (spliced), deepseek-v4-pro, gemini-3.1-pro, gpt-5.6-sol, grok-4.6 (spliced), qwen-3.8-max (spliced). Series density and lineage notes (quarterly #1): (1) rolling-1w forecasts emit daily since Aug 22, 2026 and rolling-1M daily since Aug 23, 2026, so across this 92-day window daily coverage is PARTIAL as measured -- the 1w series has slots on 55 of 92 days and the 1M series on 41 of 92 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 4,001, 1M 2,787). (2) 3 model lines are lineage splices inside this window: claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5). (3) Every per-model aggregate in this issue sees the merged head id; the raw id is kept in the pack, and 24,922 rows were remapped inside the quarter (19,532 / 5,390 / 0 in Jul 15-31 / Aug 1-31 / Sep 1-30).
- 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 quarter block covers 78 days (Jul 15-Sep 30) of the 92-day calendar quarter, and there is no prior-quarter block, so the TOP5 volume figure is a 78-day sum with no quarter-over-quarter 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 quarterly window (not applicable: there is no prior quarter, and the prior-cutoff counts are 0 by construction), so the control this issue is the monthly reproduction: the M3 block of this pack (Sep 1-30) reproduces the published Monthly #2 on 7 of 7 checks, the M2 block (Aug 1-31, cutoff Tue Sep 1, 2026, 16:00 UTC) reproduces the published Monthly #1 on 7 of 7 checks, and the M1 block (Jul 15-31, cutoff Sat Aug 1, 2026, 16:00 UTC) reproduces the prior window of Monthly #1 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	133,039	Mature (scored pool)	122,770
OK in gate	122,919	-- of them directional	59,606
Out of gate (1w / 1M)	4,001 / 2,787	-- of them sideways	63,164
Invalid	3,332	Pending (next issue)	149
Late closes	0	Uptime, all grids	2,929 / 3,496 slots
Source file	quarterly_metrics_2026-q3.json	Report cutoff	Thu Oct 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-10-07 11:01 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 59,606 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 · 49 published reports.

Issue #1. Quarterly Model Watch is a living quarterly comparison; each issue appends one more calendar quarter of leaderboard, asset-league, rank-stability and regime data. This first issue sets the baseline on a partial quarter (Jul 15-Sep 30): gemini-3.1-pro at the top (47.2%), 13 of 21 qualifying turns found a caller, 0 of 5 tickers above 51%, 3 distinct model lines leading the three months and a calm-vs-volatile field difference of +0.9pp; there is no prior quarter. 3 model lines are lineage splices this quarter (claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5)). Engine 1.1 has powered the sandbox since Aug 18, i.e. inside this window; no cross-engine PnL comparisons are claimed.

What we're testing next

- Quarterly #2 (Q4 2026): the same three series over Oct 1-Dec 31, cut on Jan 1 plus the maturity lag, so its quarter-over-quarter comparison runs against this partial Q3.
- Quarterly Config Watch #1: the same Q3 window through the config search engine (w309), published as the fourth quarterly report of this issue.
- Quarterly Stability Index: Run 3, the next repeat-prompt stability run, is the first cross-quarter point on model stability.

Related research

Report	Direct PDF link
Quarterly Consensus Watch #1	https://marketmania.ai/research/reports/consensus-watch-quarterly-2026-q3.pdf
Quarterly Calibration #1	https://marketmania.ai/research/reports/calibration-quarterly-2026-q3.pdf
Monthly Model Watch #2	https://marketmania.ai/research/reports/model-watch-monthly-2026-09.pdf

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

CITE THIS REPORT

```
@misc{mm_modelwatch_quarterly_2026q3,
  title = {Quarterly Model Watch #1: quarterly leaderboard, asset league, rank stability and regime split, Jul 15-Sep 30 2026},
  author = {{MarketMania Research}},
  year  = {2026}, month = {October}, day = {8},
  url   = {https://marketmania.ai/research/reports/model-watch-quarterly-2026-q3.pdf},
  note  = {Methodology v1.1 (2026-08-10), hash e66c7e8c864a2233; window Jul 15-Sep 30, 2026 UTC; source quarterly_metrics_2026-q3.json}
}
```