

Config Watch #3

August 26, 2026 · MarketMania Research · Weekly series

WEEKLY · CONFIG WATCH · Issue #3 · the OOS verdict flips · Issue date August 26, 2026 · Language English · Methodology v1.1 + Config Watch amendments (approved 12.08, hash e66c7e8c864a2233)

Search windows: **WK** Aug 17-24, 2026 and **M30** Jul 27-Aug 24, 2026 -- both frozen and Monday-aligned (WK = Mon Aug 17 00:00 -> Sun Aug 23 23:59 UTC; M30 = Mon Jul 27 -> Sun Aug 23 UTC (4 full weeks)). Slots fall inside the window; trades settle past its right edge at their horizon (the same 2-3 day luft the engine has used since issue #1 and that the live presets run on).

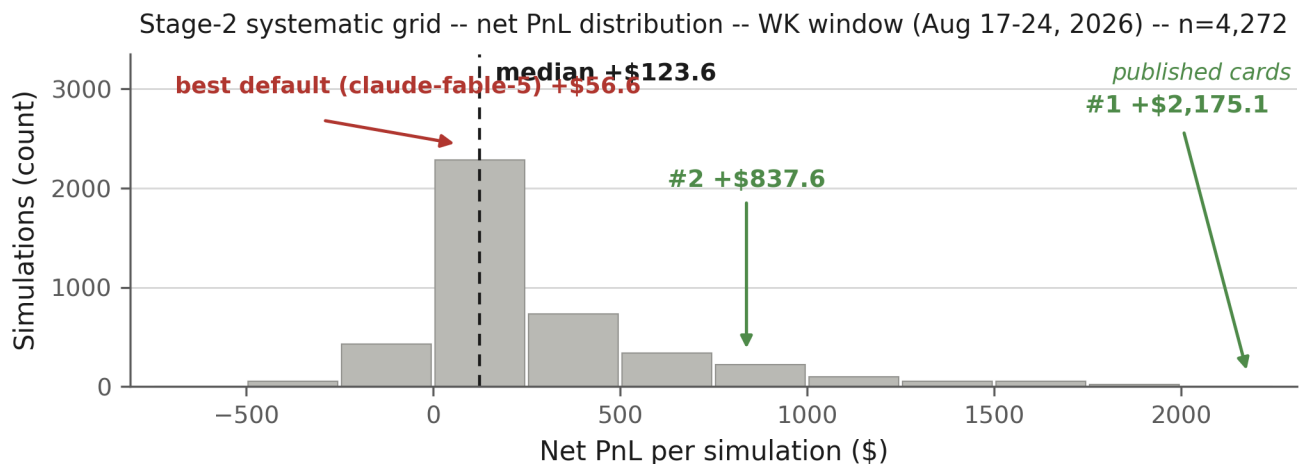
RESEARCH SNAPSHOT			
11,136 simulations, 0 errors	WK Aug 17-24, 2026	M30 Jul 27-Aug 24, 2026	7-stage search: scan -> grid -> refine -> stake sweep -> per-ticker -> OOS replays
engine 1.1; \$1,000 deposit, 10x leverage	stake is universe-aware from sim #1 (table below)	maturity gate: >=15 closed (WK) / >=20 (M30)	21 configs stable in both windows' mature top-40
walk-forward: 20 past cards + 2 live presets replayed verbatim	19 of 20 past cards net-positive on the fresh week	new boards: Calmar-like, WR, stability	models: 7 (grok axis = grok-4.6 incl. 4.5-era)
best WK council +\$2,175.1 (+217.5%)	best M30 council +\$2,137.6 (+213.8%)	zero-trade cells: 1,794	margin-skip alerts on leaders: 26 rows

KEY FINDING · OBSERVATION (the OOS verdict flips)

On a market that came back to life, out-of-sample decay stopped: all 8 cards Config Watch #2 published are net-positive on the fresh week -- its 7d champion made +\$1,861.1 at 83.3% win rate and 0.0% drawdown -- and the fresh search still beat every one of them (+\$2,175.1).

TL;DR

- OBSERVATION -- The walk-forward verdict inverted.** Issue #2 reported 10 of 12 cards fading; on this week's fresh window **19 of the 20 cards published so far are net-positive** (the exception: issue-1's gemini solo, -\$332.8 on 81 closed). Issue #2's 7d champion, replayed verbatim, made **+\$1,861.1** (WR 83.3%, maxDD 0.0%, 36 closed, 0 margin skips) -- and still finished **below** this week's fresh council (**+\$2,175.1**, WR 88.6%). One good week is not evidence of durability: the honest read is that regime, not config quality, moved most of this.
- The window is the same week the weekly wave #3 reports call 'the market came alive' -- field accuracy 43.8% -> 54.1% and mean absolute 1d move 0.73% -> 4.28% week over week (see Related research; their tables are not duplicated here). Every number in this issue is conditioned on that one regime.
- Universe collapse, again, and harder.** The net board's top-10 is a single universe -- SOL+XRP -- on BOTH windows, so only one council card per window survives the pairwise-different-universes rule; the second card on each window is the best non-SOL+XRP config on the new mature quality boards (WK +\$837.6 on BTC+ETH at 96.2% WR; M30 +\$412.3 on BTC+ETH, calibrate ON). Read the cards as two universes deep, not three.
- Two methodology changes ship this issue** (Watch amendments below): stake is universe-aware from the first simulation (a \$1,000 deposit funds \$450 on 1-2 coins but only \$200 on 5), and windows are calendar Monday-to-Sunday. The first change is visible immediately: two of issue #2's own 30d champions are **concentrated** under the new table and carried margin skips at their origin -- their printed economics were never fully funded. Issue #2 is not restated; the amendment note is the correction.



Net PnL across all 4,272 stage-2 grid sims, WK window (Aug 17-24, 2026), universe-aware stake throughout. Grid median **+\$123.6**, 64.7% positive -- the friendliest grid this series has seen (issue #2's 7d grid: median \$0.0, 31.4% positive). The spike at \$0 is mostly cells whose entry filters produced no trades (1,026 zero-trade sims on this window). Red marker = best solo default; green = the published cards: **#1 is the grid maximum itself**, #2 sits above 92.2% of the grid. The M30 histogram appears after the M30 cards. Source: grid_hist.json (w167 extras).

Why it matters

MarketMania publishes default LLM-council trading configs. This series asks one narrower question every week: how much apparent performance can large-scale in-sample search extract -- and how much of it survives out-of-sample? Issue #1 set the in-sample baseline; issue #2 delivered the first walk-forward verdict, and it was brutal. Issue #3 delivers the second, on a week where the market moved: same replay machinery, same frozen rules, an opposite answer -- which is exactly why the walk-forward table is a standing section and not a one-off. The other job of this issue is corrective: the stake model changes, and the change reaches back into how issue #2's champions should be read.

How we searched

One overnight pipeline, run per window (WK and M30 branches, council and solo together), with the criteria frozen in writing *before* launch (w157 plan) and printed into the meta of both JSON deliverables:

- **1. Scan (s1)** -- 952 sims across council compositions and coarse settings.
- **2. Systematic grid (s2)** -- 8,544 sims (4,272 per window), sweeping the declared axes (archetype, FH/TF, entry filters, universe, council size, membership, TP/SL source, ladder, break-even, TP/SL shift, calibrate and, new, stake) -- the denominator for every percentile claim.
- **3. Refinement: ladder + stake sweep (s2b+s3b)** -- 240 sims around the grid winners.
- **4. Preset neighbourhood grid (s2p)** -- 78 sims: one-knob neighbours of both live presets, on both windows.
- **5. Per-ticker probes (s5t)** -- 120 sims: best finalists split onto single coins.
- **6. Dedicated per-ticker (s6t+s6tc)** -- 1,150 sims: a compacted independent per-coin grid + calibrate twins (the standing branch from issue #2).
- **7. OOS replays (s4cw2+s4cw2o+s4cw1+s4pre)** -- 52 sims: every published card and both live presets, verbatim, on both fresh windows.

Total 11,136 simulations, 0 errors, 1,794 zero-trade cells; stage sums reconcile exactly (summary.json counts.by_stage). The axis grid was NOT widened relative to issue #2 (anti-overfit rule): the selection layer grew, the search space did not.

Stake, the new axis: universe-aware from simulation #1 (\$1,000 deposit)

Instruments in universe	Stake cap	Sweep values also simulated
1 coin	\$450	\$300 / \$375
2 coins	\$450	\$300 / \$375
3 coins	\$300	\$225 / \$375
4 coins	\$250	\$200 / \$300
5 coins	\$200	\$150 / \$250

A \$1,000 deposit cannot fund five concurrent \$450 positions. From the first simulation this issue the stake is a function of how many instruments the config trades, recomputed on every universe change. Any simulated stake above the cap for its universe size is marked **concentrated** and excluded from the headline boards (it stays in the open data). Every card also prints **skipped_nofunds** -- entries the engine could not fund; 26 such alert rows were raised this issue, flagged with * below. Source: summary.json criteria + skipped_alerts.

Models under test (exact engine IDs): claude-fable-5, claude-opus-5, gpt-5.6-sol, deepseek-v4-pro, gemini-3.1-pro, qwen-3.8-max, grok-4.6. The grok axis runs as **grok-4.6 (incl. 4.5-era)** -- one lineage, 4.5 before the Aug 24 cutover and 4.6 after; published grok-4.5 configs replayed out-of-sample get the same 4.5 -> 4.6 lineage remap ([remap] rows).

TOP councils -- WK window (Aug 17-24, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	opus-5, deepseek-v4-pro, gemini-3.1-pro	SOL+XRP	4h/1h	+\$2,175.1	+217.5%	0.0%	35	88.6%
#2	opus-5, deepseek-v4-pro, gemini-3.1-pro, grok-4.6	BTC+ETH	4h/1h	+\$837.6	+83.8%	1.1%	26	96.2%

#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	1/0	reenter	\$450
#2	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	1/0	reenter	\$450

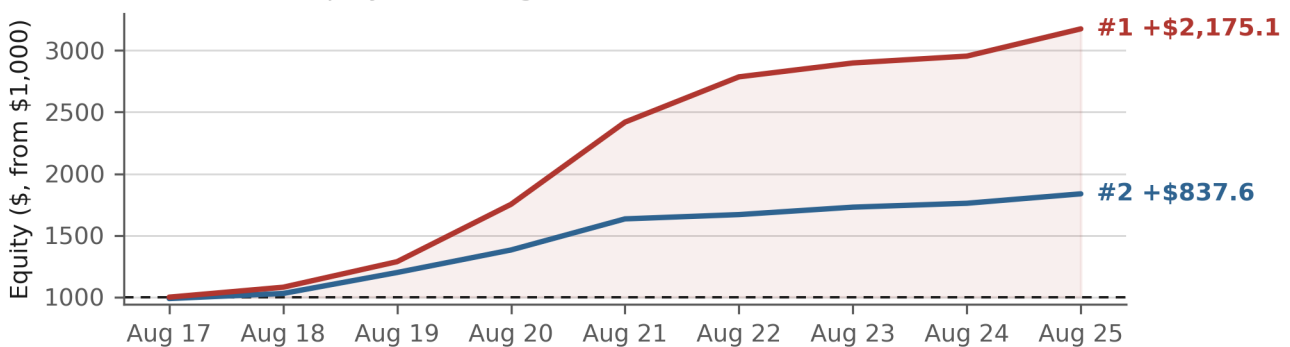
Against the 4,272-config stage-2 grid on this window: #1 is the grid maximum itself, #2 sits above 92.2% of it (histogram above; grid median +\$123.6). Net %% = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine). * on Trades = non-zero skipped_nofunds (entries the engine could not fund under the universe-aware stake) -- the printed economics are then not the economics that ran. Entry = max_sideways / max_diff_side.

Only two councils are published on this window. All ten slots of the mature net board are SOL+XRP, so the pairwise-different-universes rule admits exactly one; #2 is the best non-SOL+XRP config on the mature boards (WR board, 0 skips) -- disclosed rather than padded. Stake sweep, live: #1 re-simulated at \$375 instead of its \$450 cap returns +\$1,812.6 on the identical trade sequence -- the stake axis scales the result, not the decisions.

OPEN IN SANDBOX marketmania.ai/s/cw3-wk1 marketmania.ai/s/cw3-wk2

Each link opens this exact frozen window; results visible without sign-in (embedded share signature).

Published councils equity -- WK (Aug 17-24, 2026) -- \$1,000 at 10x -- 61 settled trades



Daily settled equity for the councils above (\$1,000 start, 10x leverage, frozen window; ranks match the table -- #1 red, #2 blue, fixed rank colors across all issues). A curve running past the window end is an open-at-cutoff trade settling at its horizon (the 2-3 day left). Each curve's last point reconciles to its card's net to the cent. Source: equity_cw3.json (w167 extras).

TOP councils -- M30 window (Jul 27-Aug 24, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	opus-5, deepseek-v4-pro, gemini-3.1-pro	SOL+XRP	4h/1h	+\$2,137.6	+213.8%	41.5%	71 *	78.9%
#2	fable-5, opus-5, deepseek-v4-pro, gemini-3.1-pro, qwen-3.8-max	BTC+ETH	1d/4h	+\$412.3	+41.2%	2.9%	22	95.5%

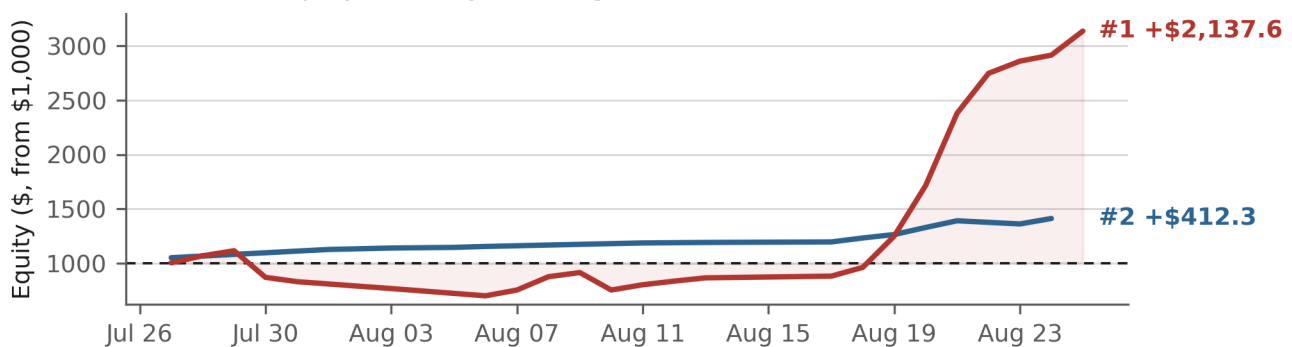
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	1/0	reenter	\$450
#2	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	3/1	reenter	\$450

Read #1 with its flag: 71 closed but 15 entries unfunded (* = non-zero skipped_nofunds; 41.5% maxDD) -- a margin-constrained run, printed as-is because the board is the board; the best fully-funded M30 council (0 skips), +\$1,891.9 on 72 closed, is in the open data. #2 is the only non-SOL+XRP multi-coin config on the M30 mature boards and a **calibrate=ON** cell -- the first calibrated champion this series has published (same two-card universe-collapse disclosure as WK). Against the stage-2 grid: #1 is the grid maximum, #2 above 83.9% of it (histogram below).

OPEN IN SANDBOX marketmania.ai/s/cw3-m301 marketmania.ai/s/cw3-m302

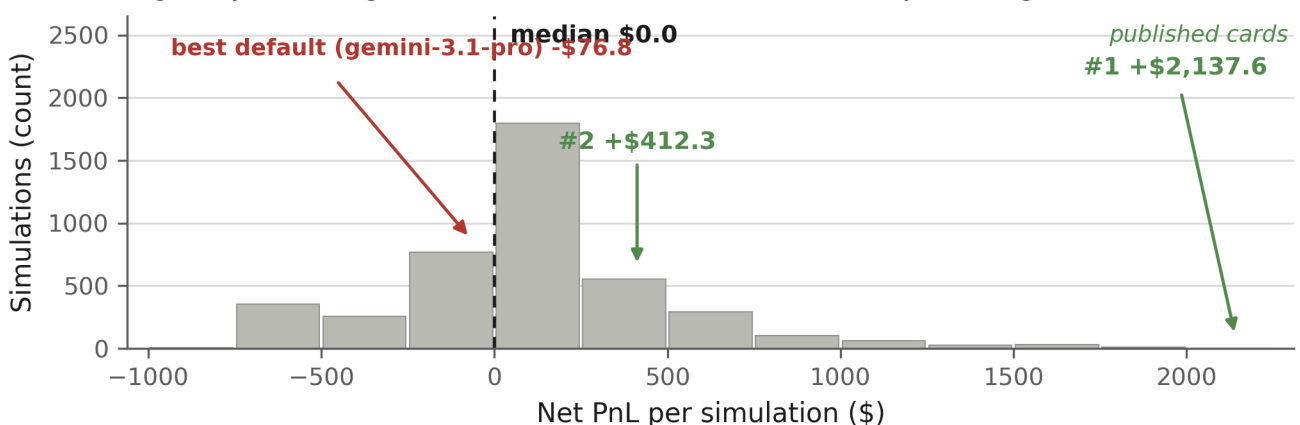
Each link opens this exact frozen window; results visible without sign-in (embedded share signature).

Published councils equity -- M30 (Jul 27-Aug 24, 2026) -- \$1,000 at 10x -- 93 settled trades



Daily settled equity for the councils above (\$1,000 start, 10x leverage, frozen window; #1 red, #2 blue -- fixed rank colors). The month tells the regime story in one line: #1 spends three weeks underwater (trough -\$299.5 on Aug 6) and makes everything in the final week; #2, the calibrated BTC+ETH cell, grinds a flat monotone path. Source: equity_cw3.json (w167 extras).

Stage-2 systematic grid -- net PnL distribution -- M30 window (Jul 27-Aug 24, 2026) -- n=4,272



Net PnL across all 4,272 stage-2 grid sims, M30 window (Jul 27-Aug 24, 2026). Grid median **\$0.0**, 49.6% positive -- the harder window again, though far friendlier than issue #2's month (median -\$159.0, 13.7% positive). The \$0 spike is 764 zero-trade cells. Red marker = best solo default; green = the published cards: #1 is the grid maximum, #2 above 83.9% of the grid. Source: grid_hist.json (w167 extras).

Walk-forward: the standing OOS verdict

Every config this series has ever published -- all 20 cards from issues #1 and #2 -- replayed VERBATIM (same knobs, same stake as printed, no re-tuning) on both fresh windows, plus the two live hand-tuned presets running on the platform, listed first. 52 of this issue's 11,136 simulations went into the replays.

Config	Family	Coins	Issue #1/#2 (home)	New 7d	New 30d	Verdict
preset sol_4h4	fable, qwen, deepseek, opus	ETH+SOL	-- live	+\$320.8 (35)	-\$561.8 (41)	Faded
preset xrp_bc2	gemini, gpt, deepseek	SOL+XRP	-- live	+\$431.8 (8) *	+\$543.8 (13)	Survived
cw2-30d1	opus, gpt, deepseek, gemini, qwen	BTC+ETH+SOL+BNB+XRP [conc]	+\$744.7	+\$1,024.1 (12)	+\$1,793.7 (30)	Survived
cw2-30d2	opus, deepseek, gemini	ETH+SOL+BNB+XRP [conc]	+\$655.1	+\$1,053.6 (13)	+\$1,393.5 (32)	Survived
cw2-30d3	opus, deepseek, gemini	BNB+XRP	+\$489.1	+\$391.9 (4) *	+\$1,006.6 (25)	Survived
cw2-30ds1	gemini	BNB+XRP	+\$383.8	+\$315.2 (10)	+\$890.0 (34)	Survived
cw2-7d1	fable, opus, deepseek, gemini	SOL+XRP	+\$299.1	+\$1,861.1 (36)	+\$1,835.4 (72)	Survived
cw2-7d2	gpt, deepseek, gemini, qwen	BNB+XRP	+\$223.7	+\$391.9 (4) *	+\$820.3 (26)	Survived
cw2-pt1	fable, deepseek, grok [remap]	SOL	+\$698.3	+\$772.1 (20)	+\$1,313.8 (49)	Survived
cw2-pt2	opus	XRP	+\$543.7	+\$250.3 (4) *	+\$601.4 (13)	Survived
cw1-30d1	fable, deepseek, opus, qwen	ETH+SOL	+\$868.4	+\$523.0 (18)	+\$348.6 (38)	Survived
cw1-30d2	gemini, gpt, deepseek	SOL+XRP+ETH	+\$590.6	+\$566.3 (12)	+\$1,076.7 (28)	Survived
cw1-30d3	fable, deepseek, opus, qwen	BTC+ETH+SOL	+\$580.5	+\$473.8 (27)	+\$234.3 (63)	Survived
cw1-30ds1	opus	SOL+XRP	+\$442.8	+\$660.9 (33)	+\$329.6 (66)	Survived
cw1-30ds2	grok [remap]	SOL+XRP	+\$262.9	+\$296.3 (25)	+\$474.8 (58)	Survived
cw1-30ds3	deepseek	XRP+BNB	+\$119.0	+\$769.4 (41)	+\$422.9 (63)	Survived
cw1-7d1	fable, deepseek, opus, gpt	SOL+XRP	+\$439.2	+\$1,366.7 (43)	+\$1,218.4 (95)	Survived
cw1-7d2	fable, qwen, gemini, opus	XRP+BNB	+\$429.5	+\$217.5 (39)	-\$563.5 (12)	Faded
cw1-7d3	gpt, opus	BNB+XRP+SOL	+\$398.6	+\$435.2 (53)	+\$142.5 (94)	Survived
cw1-7ds1	gemini	XRP+BNB	+\$417.8	-\$332.8 (81)	-\$555.6 (23)	Faded
cw1-7ds2	gpt	XRP+BNB	+\$366.3	+\$355.0 (37)	+\$360.6 (97)	Survived
cw1-7ds3	qwen	XRP+BNB	+\$267.0	+\$300.8 (33)	+\$163.0 (86)	Survived

n in parentheses; * = fewer than 10 trades (insufficient cell). Issue #1/#2 (home) = the number the card's origin issue printed (issue #2 for cw2-*, issue #1 for cw1-*); the presets are live platform configs with no home issue. Verdict rule as in issue #2 -- **Faded** = negative on at least one fresh window, **Survived** = positive on both (green rows); presets judged by the same rule. New 7d is pure out-of-sample for every row (the week starts at issue #2's right edge, Aug 17); New 30d shares 21 of its 28 days with issue #2's 30d window -- decay context, not pure OOS. [conc] = stake above this issue's cap; [remap] = grok-4.5 replayed on the grok-4.6 lineage. Origin re-runs (drift baseline): re-simulating the eight issue-2 cards on their OWN origin windows today lands within **-\$177.7...+\$88.6** of the printed #2 numbers (cw2-pt1 reproduces to the cent; the two [conc] cards carry 12-13 unfunded entries at origin). Snapshot principle -- issue #2 is not restated; the full drift table is in the open-data JSON (oos_walk_forward.drift_baseline).

Honest read. The direction of this table is the opposite of issue #2's, and the most likely cause is not that the configs got better -- it is that the week did. On a week where field accuracy jumped ten points and daily moves grew six-fold, almost any long-biased, confidence-gated, break-even-stopped ladder made money: 19 of 20 past cards are net-positive on the fresh week, **18 of 20 Survive** the standing verdict (issue #2's read: 2 of 12). Two facts

keep this from being a victory lap: the fresh search still beat every replayed card (+\$2,175.1 vs +\$1,861.1), which is what in-sample search is supposed to do and says nothing about next week; and decay did not disappear -- cw1-7ds1 lost both fresh windows, cw1-7d2 and the live preset sol_4h4 lost the month. One bad OOS week and one good one. Neither is a strategy.

Calibrate axis: does TP/SL calibration help a config?

Matched pairs -- identical knobs, calibration OFF vs ON -- measured on net PnL. The search produced **2,988** matched pairs; the open-data pack exports a **400-pair top-of-board sample**, and every number below is computed from that sample only.

Window	Pairs	Calibrate wins	Mean delta	Best delta	Worst delta
WK	279	0%	-\$1,045.3	-\$730.3	-\$1,728.5
M30	121	2.5%	-\$1,071.8	+\$863.6	-\$2,166.3
Both	400	0.8%	-\$1,053.3	+\$863.6	-\$2,166.3

This is a tail measurement, not a population estimate: the sample is drawn from the top of the boards, and 99.2% of its calibrate-OFF legs are already net-positive (median +\$1,162.0; median delta -\$976.6 overall, -\$969.5 / -\$1,029.6 by window). At the tail calibration lost 397 of 400 pairs this week, while issue #2 measured the middle of its population and found it winning 84% of 396 -- both sharpen the same standing claim: **calibration lifts the middle of the population and clips the extreme tail** (the calibrated legs are not broken -- 84.5% still net-positive, median +\$193.3 -- just capped below their raw twins on a week that rewarded raw exposure). One live counter-example sits in the cards above: this issue's M30 #2 IS a calibrate=ON cell, the first calibrated champion the series has published.

Secondary analysis -- solo-model top

Sidebar to the council narrative. Solo configs run inside the same pipeline and face the same gates: ≥ 2 coins, the maturity gate for the window, one config per model.

Solo top -- WK (Aug 17-24, 2026)

#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	gemini-3.1-pro	SOL+XRP	4h/4h	+\$1,604.3	+160.4%	8.3%	57 *	63.2%
#2	fable-5	SOL+XRP	4h/4h	+\$1,440.4	+144.0%	15.1%	39	79.5%

#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	median	0/0	reenter	\$450
#2	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	median	0/0	reenter	\$450

Solo top -- M30 (Jul 27-Aug 24, 2026)

#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	gemini-3.1-pro	SOL+XRP	1d/1d,4h	+\$1,449.0	+144.9%	29.7%	34 *	73.5%
#2	opus-5	SOL+XRP	1d/1d	+\$1,221.8	+122.2%	5.2%	22	81.8%

#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	farthest	1/0	reenter	\$450
#2	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	median	0/0	reenter	\$450

Both window leaders are gemini-3.1-pro and both carry margin skips (1 and 4 skipped_nofunds, flagged *); the clean twin of the WK leader (0 skips) made +\$1,514.9 on 26 closed at 80.8% WR. Both sit above 98.8% of their stage-2 grids; solo equity series are in the open-data JSON (equity.series). Source: summary.json top_mature_by_window_branch.

OPEN IN SANDBOX marketmania.ai/s/cw3-s1 marketmania.ai/s/cw3-s2

Each link opens this exact frozen window; results visible without sign-in (embedded share signature).

Council vs solo. The best solo sits below the best council on both windows (+\$1,604.3 vs +\$2,175.1 on WK; +\$1,449.0 vs +\$2,137.6 on M30): the ensemble edge held at the top for the third issue running. Same in-sample caveats apply to both sides.

Quality boards -- new in issue #3

Net PnL stays the primary board (continuity with issues #1 and #2); three selection views are added on top at zero additional sims. **Calmar-like** = net / max(maxDD, 1.0), positive net only; **WR board** = win rate among mature configs (>=15 closed WK / >=20 M30); **stability** = present in the mature net top-40 of BOTH windows. Top-3 of each board below; all 15 rows of all four boards are in the open-data JSON.

Board	Config	Coins	FH/TF	Net	WR	maxDD	Cls	Calmar
WK Calmar	opus, deepseek, gemini	SOL+XRP	4h/1h	+\$2,175.1	88.6%	0.0%	35	2,175.1
WK Calmar	fable, opus, deepseek, gemini	SOL+XRP	4h/1h	+\$1,923.6	84.8%	0.0%	33	1,923.6
WK Calmar	fable, opus, deepseek, gemini	SOL+XRP	4h/1h	+\$1,861.1	83.3%	0.0%	36	1,861.1
WK WR	fable, opus, gpt, deepseek, gemini	BTC+ETH	4h/4h,1h	+\$703.7	100.0%	0.0%	16	703.7
WK WR	fable, opus, deepseek, gemini	BTC+ETH	4h/4h,1h	+\$688.0	100.0%	0.0%	15	688.0
WK WR	fable, opus, gpt, deepseek, grok	BTC+ETH	4h/4h,1h	+\$605.0	100.0%	0.0%	15	605.0
M30 Calmar	fable, opus, deepseek, gemini	SOL+XRP	1d/1d	+\$1,480.5	85.0%	2.0%	20	740.2
M30 Calmar	fable, opus, deepseek, gemini	SOL+XRP	1d/1d	+\$1,480.5	85.0%	2.0%	20	740.2
M30 Calmar	fable, opus, deepseek, gemini	SOL+XRP	1d/1d	+\$1,233.7	85.0%	1.7%	20	725.7
M30 WR	fable, opus, deepseek, gemini, qwen [cal]	BTC+ETH	1d/4h	+\$412.3	95.5%	2.9%	22	142.2
M30 WR	fable, opus, deepseek, gemini, qwen [cal]	BTC+ETH	1d/4h	+\$343.6	95.5%	2.4%	22	143.2
M30 WR	fable, opus, deepseek, gemini, qwen [cal]	BTC+ETH	1d/4h	+\$274.9	95.5%	1.9%	22	144.7

[cal] = calibrate=ON cell. The WK Calmar board is degenerate by construction -- its top rows all have maxDD 0.0%, so the score collapses onto net PnL: read it as 'net among configs that never drew down', not as a risk-adjusted ranking. **Stability:** 21 configs sit in the mature net top-40 of BOTH windows -- all 21 SOL+XRP councils, only 2 clean (zero unfunded entries) on both windows; the top row by rank sum is WK card #1 itself. Full list in the open data.

Per-ticker bests

Which coin was extractable this window, and by what. Single-coin results are excluded from the council cards by rule (idiosyncratic-coin risk) and reported here, never ranked against the cards. The dedicated per-coin branch contributed 1,150 sims -- a compacted grid, one best per coin; the branch tag in brackets says which kind won (issue #2's independent best-council / best-solo columns are not run this issue).

WK (Aug 17-24, 2026)

Coin	Best config (branch)	FH/TF	Net (trades)
BTC	gemini, qwen [council]	4h/4h	+\$728.2 (29)
ETH	opus, deepseek, gemini [council]	4h/1h	+\$697.2 (21)
SOL	opus, gemini [council]	4h/4h	+\$1,224.0 (32)
BNB	fable, gemini [council]	1d/1d	+\$427.0 (5) *
XRP	fable, gemini, grok [council]	4h/4h	+\$1,045.1 (14)

M30 (Jul 27-Aug 24, 2026)

Coin	Best config (branch)	FH/TF	Net (trades)
BTC	gemini [solo]	4h/4h,1h	+\$545.3 (39)
ETH	fable, deepseek, gemini [council]	1d/1d	+\$276.2 (9) *
SOL	fable, opus, deepseek, gemini, grok [council]	4h/1h	+\$1,329.1 (40)
BNB	fable, opus, gemini [council]	1d/1d	+\$663.4 (15)
XRP	fable, gemini, grok [council]	4h/4h	+\$1,172.7 (50)

n in parentheses; * = fewer than 10 closed trades (thin -- direction only). All ten winners run calibrate OFF; every winner is a council except BTC on the month.

The reads. Every coin was extractable this week -- including BTC, which no config could make pay in issue #2 (best of its 448-cell month grid there: -\$2.7; here +\$728.2 / +\$545.3). SOL and XRP lead both windows -- the same two coins that own every card universe above: the mechanism behind the collapse, not a coincidence. **Winner's curse applies to this whole table:** each row is the maximum of a per-coin grid, biased high by selection alone, and none has an out-of-sample read until next issue.

Baselines

Three independent reference points, none of them search output, re-simulated fresh on this issue's frozen windows (w167 extras pass, engine 1.1): the engine solo default for each of the 7 models, the default council-of-7, and the pre-search hand-tuned set A/B/C carried since issue #1. The tables show the universe-aware **table-stake** run -- this issue's canonical mode (\$200 for the 5-coin defaults; A and B already at their \$300 3-coin cap; C moves \$400 -> \$450 on 2 coins).

WK (Aug 17-24, 2026)

Baseline	Net PnL	Trades	WR	maxDD
Best solo default (claude-fable-5)	+\$56.6 *	304	67%	2.6%
Worst solo default (deepseek-v4-pro)	-\$45.2 *	350	61%	4.5%
Council-of-7 default [remap]	+\$16.4 *	140	68%	2.1%
Hand-tuned config A (pre-search)	+\$442.3	11	82%	4.2%
Hand-tuned config B (pre-search)	+\$333.0 *	52	62%	43.4%
Hand-tuned config C (pre-search)	+\$523.0	18	78%	15.0%

M30 (Jul 27-Aug 24, 2026)

Baseline	Net PnL	Trades	WR	maxDD
Best solo default (gemini-3.1-pro)	-\$76.8 *	376	53%	8.4%
Worst solo default (deepseek-v4-pro)	-\$167.6 *	823	53%	16.8%
Council-of-7 default [remap]	-\$51.7 *	350	57%	6.7%
Hand-tuned config A (pre-search)	+\$795.2	19	89%	7.9%
Hand-tuned config B (pre-search)	+\$575.8 *	120	69%	43.4%
Hand-tuned config C (pre-search)	+\$348.6 *	38	68%	25.2%

* = non-zero skipped_nofunds at table stake. On the defaults that is the stake table doing its job: at \$200 a \$1,000 deposit funds at most five concurrent positions, the defaults fire hundreds of entries per window, and most of the queue goes unfunded. The other five solo defaults and both stake modes (original + table) with full stats are in the open-data JSON; **the conclusions do not change between modes** -- every default is negative on the month in both, A/B/C positive in both. [remap] = grok-4.6 lineage. The PF column of issue #2 is not computed this issue (absent from the collect; returns next issue). Source: baselines_cw3.json (w167 extras).

The month is the cleanest read this series has had. Every default is negative on M30 in BOTH stake modes (solo defaults -\$76.8 to -\$167.6 at table stake, council-of-7 -\$51.7) while all three hand-tuned configs are positive: +\$795.2 / +\$575.8 / +\$348.6 -- and the fresh-search cards beat both (+\$2,137.6). The week adds a wrinkle: defaults hover near zero and the stake table does not merely scale them -- gemini's solo default flips +\$156.4 (orig) -> -\$13.3 (table, 273 unfunded entries), because capping the stake changes WHICH entries get funded, not just their size. Hand-tuned A/B/C make +\$333.0 to +\$523.0 on the week; the fresh cards +\$2,175.1. The standing

ordering -- search > hand-tuned > defaults -- is back, in-sample as ever.

Live presets & preset neighbourhoods

Two hand-tuned configurations trading live on the platform (not search output, distinct from the pre-search A/B/C set): preset_xrp_bc2 = '1d_3_llm_BC2_v02' (in service since 2026-08-18), a tuned 2-coin variant of config A's 1d family; preset_sol_4h4 = '4h_4_llm_v02' (since 2026-08-24), the live retune of config C's family. Their replay rows lead the Walk-forward table; passports and window detail are in the open data (presets_recon.json). The neighbourhood grid asks: does a one-knob neighbour beat the preset as configured?

Live preset	Window	As configured	Best neighbour in local grid	Cells	Verdict
preset_xrp_bc2	WK	+\$431.8	+\$578.5 (1d/1d, SOL+XRP, 8 cls)	19	neighbour +\$146.7
preset_xrp_bc2	M30	+\$543.8	+\$653.1 (1d/1d, SOL+XRP, 14 cls)	19	neighbour +\$109.3
preset_sol_4h4	WK	+\$320.8	+\$641.6 (1d/1d, ETH+SOL, 10 cls)	20	neighbour +\$320.8
preset_sol_4h4	M30	-\$561.8	+\$442.3 (1d/1d, ETH+SOL, 12 cls)	20	neighbour +\$1,004.0

Search validation. The fresh search beat the entire baseline field on both windows (week: +\$2,175.1 vs +\$523.0 best baseline / +\$431.8 best preset; month: +\$2,137.6 vs +\$795.2 / +\$543.8) -- which is what an 11,136-sim in-sample search is supposed to do, and is not evidence the search output will trade better. The more useful result: in **all four** preset-window cells a one-knob neighbour beat the configuration as it is actually running, in one case by +\$1,004.0. The live presets are not at a local optimum -- a testable, low-risk edit, unlike adopting a search champion wholesale.

Patterns: what the finalists look like

Knob modes across each window's unique finalists (mature net top-10 per branch plus both quality boards, deduplicated by full signature; 38 unique configs on WK, 41 on M30).

Knob	WK finalists mode	M30 finalists mode
Forecast horizon	4h (37/38)	4h (29/41)
Ladder shape (steps)	50:40,100:60 (38/38)	50:40,100:60 (41/41)
Break-even stop (BE)	on (38/38)	on (41/41)
Min confidence	60 (36/38)	60 (38/41)
SL shift	75 (38/38)	75 (41/41)
TP shift	-10 (38/38)	-10 (41/41)
TP/SL source	farthest (34/38)	farthest (36/41)
Calibrate ON	0/38	15/41
Universe	SOL+XRP (26/38)	SOL+XRP (26/41)

The mechanical core is unchanged for the third issue running: 50:40,100:60 two-rung ladder, break-even on, conf 60, SL +75 / TP -10, farthest source. What DID change is the horizon: issue #2's finalists were modal 1d on both windows, this issue's are **4h** -- the shorter horizon won the week the market started moving. Membership concentrates (deepseek in 24 of 28 unique WK council finalists, gemini 27, opus 23, qwen 2). Standing caveats: refinement seeds from the same grid winners, so part of the convergence is search-design echo; and a knob core that produced 10-of-12 decay one issue ago and 18-of-20 survival now is describing the weather, not a recommendation.

Liquidation note

Liquidation is modeled by the engine at the fixed 10x leverage used throughout. Across the published cards, stop-loss placement stays inside the distance that would approach the liquidation threshold; maxDD is printed on every config so realized risk is visible directly. This issue it matters: the M30 net leader ran a 41.5% drawdown. A high headline net and a survivable path are different claims.

Watch amendments (this issue)

Amendment 1 -- universe-aware stake, from simulation #1. A \$1,000 deposit funds \$450 on a 1-2 coin universe but only \$200 on five. Issue #2 simulated every universe at \$450, which margin-capped its wide-universe configs: cw2-30d1 and cw2-30d2 -- issue #2's 30d #1 and #2 -- are **concentrated** under the new table, and their origin re-runs show 12-13 entries the engine could not fund. Issue #2 is NOT restated (snapshot principle); this note is the correction of record. From this issue on, stake is a function of universe size, recomputed on every universe change, swept as its own axis; above-cap values are excluded from the headline boards.

Amendment 2 -- calendar Monday-aligned windows. Both windows are now whole calendar weeks (WK = Mon Aug 17 00:00 -> Sun Aug 23 23:59 UTC; M30 = Mon Jul 27 -> Sun Aug 23 UTC, 4 full weeks) instead of a free 7d / 30d right edge. The WK window is exactly the week the sibling weekly reports cover, and each issue's OOS window begins precisely where the previous issue's search ended -- what makes the New 7d column above pure out-of-sample.

Both amendments are provisional and scoped to this series. **Formalization is slated for methodology v1.2**; until then this issue runs on v1.1 plus the Config Watch amendments approved 12.08 (hash e66c7e8c864a2233).

Limitations

- **One regime, and it is the whole story.** The WK window is a single exceptional week (field accuracy 43.8% -> 54.1%, mean |1d| move 0.73% -> 4.28%); the survival result, the 4h flip and the calibrate reading are all conditioned on it. Do not extrapolate.
- **Multiple testing.** 11,136 simulations; at this scale some winners are expected from chance alone. Antidotes: axes frozen before launch, the walk-forward table, independent baselines, in-sample labeling. No formal correction yet.
- **Winner's curse.** Every card, quality-board row and per-ticker best is the maximum of a search, biased high by selection alone; OOS reads arrive next issue. Read the in-sample numbers as an upper bound, not an expectation.
- **Extras provenance.** Grid histograms, equity curves and baselines come from a same-day read-only extras pass (w167) on the same frozen windows and engine; every equity curve reconciles to its card's net to the cent.
- **Calibrate sample is not the population.** The 400 exported pairs are a top-of-board slice of 2,988; that section describes the tail only. The full population was not exported this issue.
- **Margin skips on leaders.** 26 alert rows: several board leaders, including the M30 council and both solo leaders, could not fund every entry -- printed economics are not the economics that ran. Flagged with *, never silently pooled.
- **Universe collapse.** SOL+XRP holds every slot of both net boards and all 21 stability rows: two council cards per window instead of three, the second sourced from a quality board.
- **Thin cells.** Several rows run on 4-14 closed trades (the BNB week best on 5; four walk-forward week cells under 10) -- flagged with *, reported for completeness.
- **Research-to-date counter is pinned.** The counter below is pinned from wave 3 onward; issue #2 printed 23,458 under an earlier, unpinned definition -- not comparable, and issue #2 is not restated.
- Research output, not financial advice.

Counters & lineage

Stage	Sims	Stage	Sims
Scan (s1)	952	Per-ticker probes (s5t)	120
Systematic grid (s2)	8,544	Dedicated per-ticker (s6t+s6tc)	1,150
Refinement: ladder + stake (s2b+s3b)	240	OOS replays (s4cw2+s4cw2o+ s4cw1+s4pre)	52
Preset neighbourhood grid (s2p)	78	Errors / zero-trade	0 / 1,794
Total	11,136	Engine	1.1

Scale note (vs issue #2). Issue #2 ran 14,861 sims (9,901 overnight pipeline + 4,960 dedicated per-ticker added same-day); issue #3 runs 11,136 in ONE pipeline: main grid marginally larger (8,544 vs 8,480), refinement comparable (318 vs 269), 52 sims to walk-forward replays. The whole reduction falls on the per-coin branch (1,150

vs 4,960 -- a compacted grid, not an independent 448-cell grid per coin-window); the stake sweep buys real sims, the three quality boards cost zero.

Extras pass (w167, same day, read-only). 56 verification sims outside the pipeline count: 12 equity replays (reconciling to the cent) + 44 baseline runs (11 configs x 2 windows x 2 stake modes); the grid histograms cost zero extra sims.

Generated at: 2026-08-26T12:53:31Z · Methodology v1.1 + Config Watch amendments (approved 12.08), hash e66c7e8c864a2233.

Snapshot principle: this report is a frozen snapshot of the search taken at the generated-at timestamp above; numbers are not updated retroactively. Each issue re-runs the pipeline fresh over that week's windows.

Research to date

RESEARCH SNAPSHOT

THIS REPORT -- search effort: **11,136 simulations** (0 errors), of which 52 walk-forward replays · 56 verification sims (equity re-runs + baselines, same-day extras pass)

MARKETMANIA RESEARCH TO DATE (as of cutoff): 33,809 directional forecasts resolved since Jul 11 · 11 models tracked (7 current + 4 archived legacy) · 5 assets · 5 horizons · hourly · 14 published reports

Counter pinned from wave 3 onward; issue #2 printed 23,458 under an earlier, unpinned definition -- the two are not comparable and issue #2 is not restated.

Issue #3. The walk-forward table is now a standing section with two opposite data points in it -- that is the point of running it every week. Windows are calendar Monday-aligned; stake is universe-aware from simulation #1; three quality boards (Calmar-like, WR, stability) join net PnL as selection views, at zero extra simulation cost. Formalization of both amendments is slated for methodology v1.2.

What we're testing next

- Next issue: the OOS fate of THIS issue's cards, including first OOS reads on all ten per-ticker bests and both calibrated champions -- the third data point is the first worth plotting.
- The 4h-vs-1d horizon flip: a property of a moving market, or was issue #2's 1d finding a flat-market artifact? Testable by replaying both issues' finalists across both regimes.
- Fold the extras pass into the overnight pipeline and export the full 2,988-pair calibrate population: equity, histograms, baselines and pairs belong in the frozen pack, not a same-day pass or a top-slice.
- Break the universe collapse (the search keeps landing on SOL+XRP): a per-universe board ranking BTC+ETH and BNB-bearing configs against their own field.

Related research

Consensus Watch #3 -- <https://marketmania.ai/research/reports/consensus-watch-2026-08-17.pdf>

Weekly Calibration #3 -- <https://marketmania.ai/research/reports/weekly-calibration-2026-08-17.pdf>

Weekly Model Watch #3 -- <https://marketmania.ai/research/reports/model-watch-2026-08-17.pdf>

Config Watch #2 -- <https://marketmania.ai/research/reports/config-watch-2026-08-12.pdf>

The three weekly reports publish together as one issue each week; Config Watch follows its own cycle. The market-regime figures in this issue come from that wave, not re-tabulated here.

CITE THIS REPORT

```
@techreport{mm_configwatch_2026w35,
  title      = {Config Watch #3},
  author     = {{MarketMania Research}},
  institution = {MarketMania},
  year       = {2026},
  month      = aug,
  day        = {26},
  type       = {Weekly Research Report},
  series     = {Config Watch},
  number     = {3},
  note       = {Methodology v1.1 + Config Watch amendments, hash e66c7e8c864a2233},
  url        = {https://marketmania.ai/research/reports/config-watch-2026-08-19.pdf}
}
```