

Config Watch #4

September 4, 2026 · MarketMania Research · Weekly series

WEEKLY · CONFIG WATCH · Issue #4 · two windows, and the ceiling falls on one of them · Issue date September 4, 2026 · Language English · Methodology v1.1 + Config Watch amendments (approved 12.08, hash e66c7e8c864a2233)
Search windows: **WK** Aug 24-31, 2026 and **M30** Aug 3-31, 2026 -- both frozen and Monday-aligned (WK = Mon Aug 24 00:00 -> Sun Aug 30 23:59 UTC; M30 = Mon Aug 3 00:00 -> Sun Aug 30 23:59 UTC, 4 full weeks), sharing the cut-off edge Mon Aug 31 00:00 UTC, so slots dated Aug 31 fall outside both. 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

16,447 simulations, 0 errors, in two frozen collects	WK Aug 24-31, 2026	M30 Aug 3-31, 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)	criteria frozen in writing before launch (w200 plan)
walk-forward: 26 past cards + 4 live presets replayed verbatim on both windows	7 of 26 past cards Survive (net-positive on both windows)	new boards: Smoothness , quality composite; stability from the night run	models: 7 (grok axis = grok-4.6 (incl. 4.5-era))
best WK council +\$399.4 (+39.9%)	best M30 council +\$1,676.0 (+167.6%)	zero-trade cells: 1,021 (WK) / 1,188 (night run)	margin-skip alerts on leaders: 0 (WK) / 8 (M30) rows

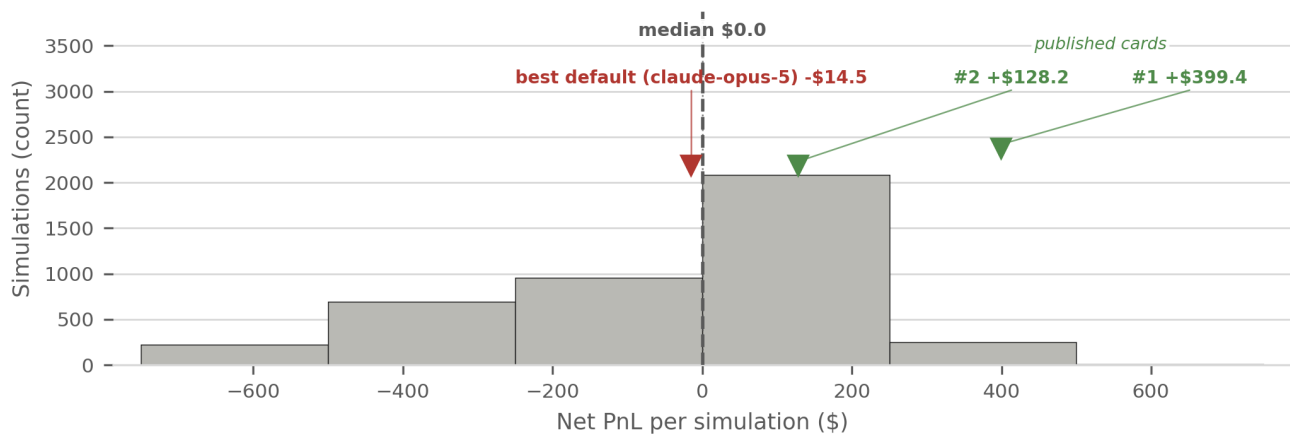
KEY FINDING · OBSERVATION (the week falls, the month holds)

The ceiling of the fresh search fell from +\$2,175.1 to +\$399.4 on the week while the four-week window ending at the same edge still pays +\$1,676.0, and under the standing two-window rule 7 of the 26 cards this series has published Survive: 19 of them clear the month, 7 clear the week.

TL;DR

- OBSERVATION -- The walk-forward verdict split by window.** Issue #3 reported 19 of 20 past cards net-positive on its fresh week; this issue replays every past card on two fresh windows and **7 of the 26 cards published so far Survive** the standing rule (net-positive on both). On the week alone 7 of 26 are positive, on the four-week window 19 of 26, and every row that clears the week also clears the month -- the week is what removes cards this issue. Issue #3's own week champion, replayed verbatim, makes -\$599.6 on the week and +\$1,263.4 on the month.
- The search ceiling fell on the week and held on the month.** The best mature council on WK makes +\$399.4 (+39.9% of a \$1,000 deposit, 95.5% win rate, 0.1% maxDD, 22 closed) against issue #3's +\$2,175.1; on M30 it makes +\$1,676.0 (+167.6%, 80.6% win rate, 53.6% maxDD, 31 closed) against issue #3's +\$2,137.6. The grids behind them differ the same way: median \$0.0 with 31.0% positive across 4,208 WK cells (only 6 of them above +\$500), median \$0.0 with 47.7% positive across 4,112 M30 cells (514 above +\$500).
- The two windows disagree about calibration and about the universe.** Across the whole stage-2 grid calibration improves 61.0% of the 1,468 matched WK pairs (mean +\$123.3) and 29.1% of the 1,414 matched M30 pairs (mean -\$111.0); pooled, 45.4% of 2,882. Both published WK councils are **calibrate=ON** cells and neither M30 card is. The mature net top-10 is a single universe on each window and not the same one: BTC+ETH+SOL+BNB+XRP on WK (10 of 10 slots), SOL+XRP on M30 (10 of 10).
- Two new selection views, at zero extra simulations** -- a smoothness board (R2 of the daily closed-equity line, ulcer as tie-break; 364 rows in the WK pool, 1,527 in the M30 pool) and a rank-average quality composite -- plus a standing measurement of the live hand-tuned anchor against the whole mature field on each window: 187 of 1,774 mature WK configs and 19 of 3,184 mature M30 configs beat it on win rate, maxDD and smoothness at the same time. Net PnL stays the headline board (continuity with issues #1-#3).

Stage-2 systematic grid -- net PnL distribution -- WK window (Aug 24-31, 2026) -- n=4,208



Net PnL across all 4,208 stage-2 grid sims, WK window (Aug 24-31, 2026), universe-aware stake throughout. Grid median **\$0.0**, 31.0% positive -- the hardest grid this series has published (issue #3's week: median +\$123.6, 64.7% positive; its month: \$0.0, 49.6% positive). The spike at \$0 is mostly cells whose entry filters produced no trades (1,017 zero-trade sims on this window). Range -\$706.8 to +\$623.1; only 6 cells finished above +\$500. Red marker = best solo default at table stake; green = the published cards. The same board definition one week earlier put its top-5 at +\$2,175.1 down to +\$1,889.3; this week's runs +\$399.4 down to +\$372.2 -- a property of the week, not evidence about either issue's cards. The M30 histogram appears after the M30 cards. Source: grid_hist_wk.json.

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 (brutal), issue #3 the second (the opposite). Issue #4 delivers the third on two windows at once, and the two windows answer differently: on the fresh week the replays lose, on the fresh four weeks most of them win, and the standing rule -- positive on both -- keeps the smaller number. The other job of this issue is to widen the selection question -- net PnL alone ranked the boards for three issues; a win rate, a shallow drawdown and a straight equity line are a different ask, and this issue measures them as their own boards and against a hand-tuned configuration that is actually running.

How we searched

Two frozen collects, run to the same plan with the criteria frozen in writing *before* launch (w200 plan) and printed into the meta of every JSON deliverable: the weekly collect over the WK window, and the night collect, which sweeps the M30 and the calendar-month windows together in one pass. Each stage below prints **weekly + night**; the calendar-month boards the night collect also produced ship with Monthly Config Watch #1, so the night numbers are printed whole and never split between the two issues.

- **1. Scan (s1)** -- 476 + 952 sims across council compositions and coarse settings.
- **2. Systematic grid (s2)** -- 4,208 + 8,224 sims sweeping the declared axes (archetype, FH/TF, entry filters, universe, council size, membership, TP/SL source, ladder, break-even, TP/SL shift, calibrate and stake) -- the denominator for every distribution claim below.
- **3. Refinement: ladder + stake sweep (s2b+s3b)** -- 122 + 243 sims around the grid winners.
- **4. Preset neighbourhood grid (s2p)** -- 53 + 106 sims: one-knob neighbours of the live presets that have a local grid this issue.
- **5. Per-ticker probes (s5t)** -- 70 + 130 sims: best finalists split onto single coins.
- **6. Dedicated per-ticker (s6t+s6tc)** -- 585 + 1,160 sims: a compacted independent per-coin grid plus calibrate twins (the standing branch from issue #2).
- **7. OOS replays (s4cw3+s4cw3o+s4cw2+s4cw2o+s4cw1+s4pre)** -- 44 + 74 sims: every published card and all four live presets, verbatim, on every fresh window, plus origin re-runs of the issue-2 and issue-3 cards for the drift baseline.

Total 16,447 simulations (5,558 weekly + 10,889 night), 0 errors, 2,209 zero-trade cells (1,021 + 1,188); stage sums reconcile exactly in each collect (counts.by_stage). The axis grid was NOT widened relative to issue #3 (anti-overfit rule): the selection layer grew by two boards, the search space did not. Equity series, grid histograms, baselines and the calibrate pack are produced inside the same two collects -- there is no separate same-day extras pass this issue.

Stake: 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. Stake is a function of how many instruments the config trades, recomputed on every universe change and swept as its own axis; 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. Margin-skip alert rows on leaders: 0 on WK, 8 on M30 (the night collect raised 18 in total, the rest on the calendar-month board). Source: criteria in both collects.

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; 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 24-31, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	opus, deepseek, gemini, qwen	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$399.4	+39.9%	0.1%	22	95.5%
#2	fable, deepseek, qwen	BNB+XRP	4h/1h	+\$128.2	+12.8%	0.0%	18	83.3%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	3/0	reenter	\$200
#2	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	2/0	reenter	\$450

Against the 4,208-cell stage-2 grid on this window: card #1 lands in the +\$250 to +\$500 bucket (248 of 4,208 cells), and only 6 cells in the whole grid finished above +\$500 (grid median \$0.0). Card percentiles exported this issue cover the raw net-board leaders and the replay rows, not the mature cards -- no percentile is claimed for the two cards themselves. Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine). Entry = max_sideways / max_diff_side.

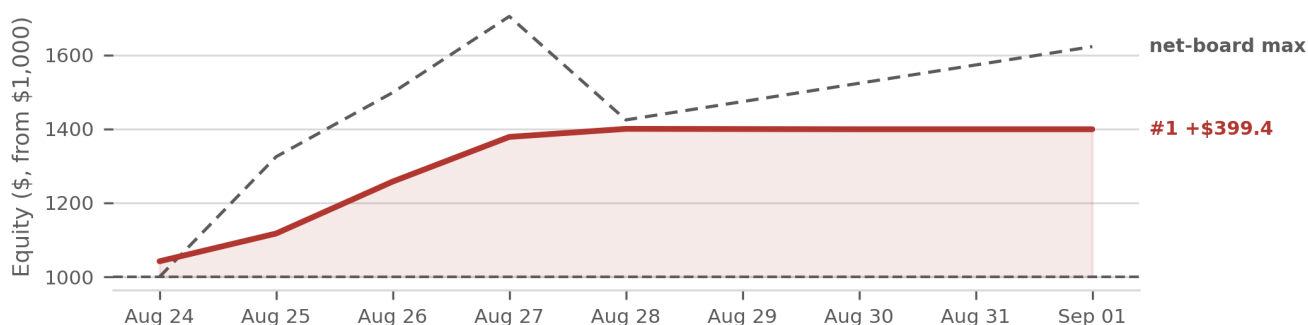
Only two councils are published on this window. All ten slots of the mature net board are the same 5-coin universe (BTC+ETH+SOL+BNB+XRP), so the pairwise-different-universes rule admits exactly one; #2 is the best mature multi-coin council outside that universe on the quality boards -- it comes off the new **Smoothness** board (R2 0.951, ulcer 0.00, 0.0% maxDD, 0 unfunded entries) -- disclosed rather than padded. Read the cards as two universes deep, not three.

Both cards are calibrate=ON cells -- the first issue where every published council is a calibrated cell (issue #3 published one, on its month window). Card #1 also runs the 5-coin table stake of \$200, so its +\$399.4 is earned on the smallest stake in the table; the same knobs at \$150 (the sweep value) are on the quality boards at +\$299.6. Neither card carries an unfunded entry.

OPEN IN SANDBOX marketmania.ai/s/cw4-wk1 marketmania.ai/s/cw4-wk2

Each link opens this exact frozen window; results visible without sign-in (embedded share signature). The codes are minted at publish time and are NOT live in this draft; every card's full frozen query signature is in the open-data JSON next to its simulation index.

Published cards equity -- WK (Aug 24-31, 2026) -- \$1,000 at 10x -- 31 settled trades



Daily settled equity for card #1 above (\$1,000 start, 10x leverage, frozen window; #1 red -- 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 luft); each curve's last point reconciles to its config's net to the cent. The muted dashed line is the net-board maximum, drawn for contrast only: at 9 closed trades it sits below the ≥ 15 maturity gate and is not a card. Card #2 has no daily series in this issue's equity pack, and the solo and smoothness series stay in the open data (equity.cards) -- the PDF carries one equity chart per window. Source: equity_wk.json.

TOP councils -- M30 window (Aug 3-31, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	gpt, deepseek, gemini	SOL+XRP	1d/1d	+\$1,676.0	+167.6%	53.6%	31	80.6%
#2	fable, gpt, deepseek, gemini	BNB+XRP	1d/1d	+\$1,155.2	+115.5%	11.3%	28	75.0%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	farthest	2/0	reenter	\$450
#2	50:40,100:60	on	TP -10 / SL +75	≥ 60 (each)	farthest	3/0	reenter	\$450

Against the 4,112-cell stage-2 grid on this window: card #1 is the same simulation as the raw net-board leader ($i=1684$), so its exported percentile is the card's -- above **99.9%** of the grid; card #2 lands in the +\$1,000 to +\$1,250 bucket (63 cells) with 97.3% of the grid finished below it. Grid median \$0.0, 514 of the 4,112 cells above +\$500. Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine). Entry = max_sideways / max_diff_side.

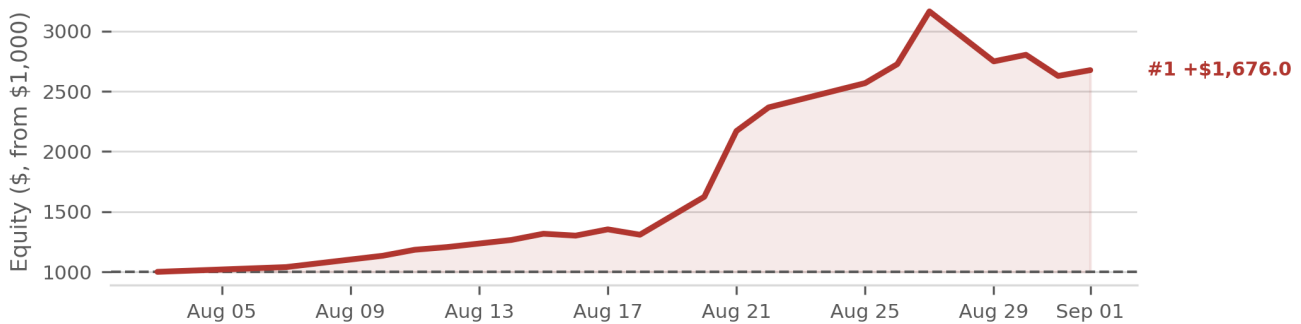
Only two councils are published on this window too, and the collapse is to a different universe. All ten slots of the M30 mature net board are SOL+XRP -- issue #3's universe, not this week's -- so the pairwise-different-universes rule admits exactly one; #2 is the best mature multi-coin council outside that universe on the quality boards (off the **Calmar** board: +\$1,155.2 on BNB+XRP, R2 0.852, 11.3% maxDD, 0 unfunded entries). Read the cards as two universes deep, not three.

Neither M30 card is a calibrated cell, where both WK cards are -- the same axis, opposite answers on the two windows (Calibrate axis below). Both M30 cards run the \$450 two-coin table stake and carry no unfunded entry; card #1 pays for its +\$1,676.0 with a **53.6% drawdown** on 31 closed trades, card #2 with 11.3% on 28. The M30 mature pool is 3,184 configs against 1,774 on WK, because four weeks clear the ≥ 20 -trade gate that one week does not.

OPEN IN SANDBOX marketmania.ai/s/cw4-m301 marketmania.ai/s/cw4-m302

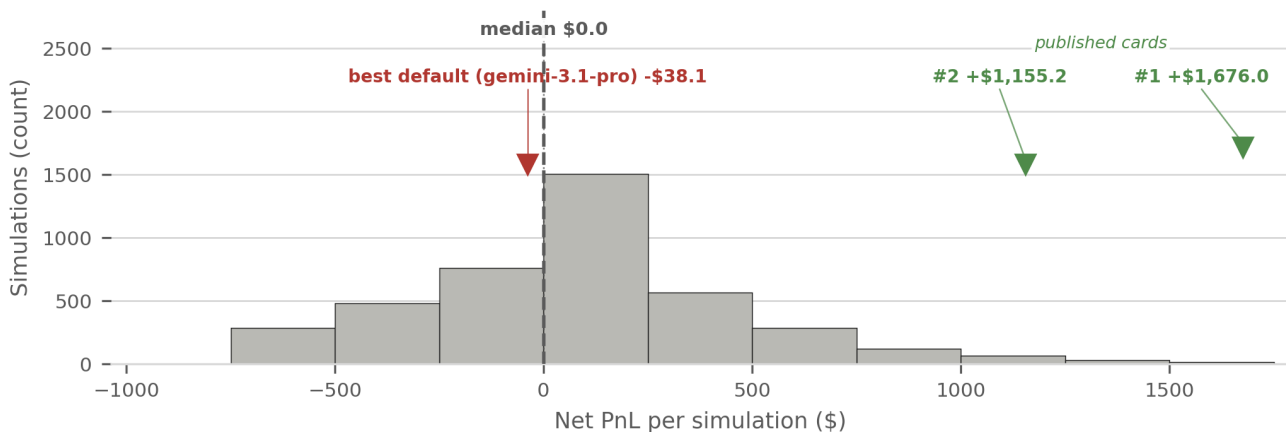
Each link opens this exact frozen window; results visible without sign-in (embedded share signature). The codes are minted at publish time and are NOT live in this draft; every card's full frozen query signature is in the open-data JSON next to its simulation index.

Published councils equity -- M30 (Aug 3-31, 2026) -- \$1,000 at 10x -- 31 settled trades



Daily settled equity for card #1 above (\$1,000 start, 10x leverage, frozen window; #1 red -- 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 luft); its last point reconciles to the card's net to the cent. The line is flat until Aug 17 and makes +\$1,323.0 of its +\$1,676.0 after it. Card #2 has no daily series in this issue's equity pack, and the solo, Calmar, WR and smoothness series stay in the open data (equity.cards) -- the PDF carries one equity chart per window. Source: equity_night.json.

Stage-2 systematic grid -- net PnL distribution -- M30 window (Aug 3-31, 2026) -- n=4,112



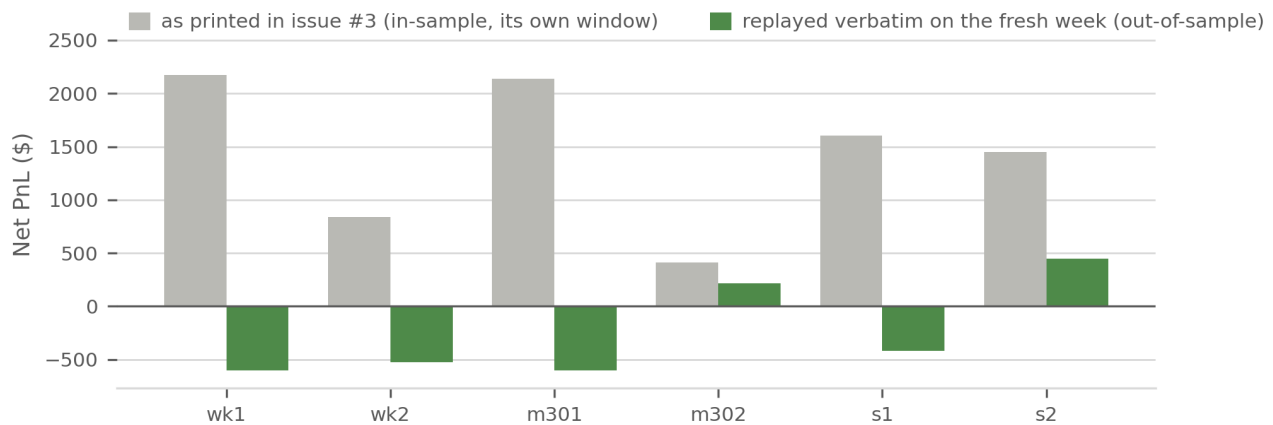
Net PnL across all 4,112 stage-2 grid sims, M30 window (Aug 3-31, 2026), universe-aware stake throughout. Grid median \$0.0, 47.7% positive against 31.0% on the week -- four weeks of the same engine on the same coins are the friendlier window this issue (issue #3's month: median \$0.0, 49.6% positive). The spike at \$0 is mostly cells whose entry filters produced no trades (624 zero-trade sims on this window). Range -\$796.0 to +\$1,741.7; 514 cells finished above +\$500, against 6 on the week. Red marker = best solo default at table stake; green = the published cards. Source: grid_hist_night.json.

Walk-forward: the standing OOS verdict

Every config this series has ever published -- all 26 cards from issues #1, #2 and #3 -- replayed VERBATIM (same knobs, same stake as printed, no re-tuning) on both fresh windows, plus the four hand-tuned presets running live on the platform, listed first. The replays and their origin re-runs cost 44 simulations in the weekly collect and 74 in the night collect, which replays every row on M30 and the calendar month together.

Verdict rule, unchanged since issue #2. Survived = net-positive on BOTH fresh windows; **Faded** = net-negative on at least one. The live presets are judged by the same rule. New WK is pure out-of-sample for every row -- the week begins exactly at issue #3's right edge (Aug 24). New M30 shares 21 of its 28 days with issue #3's own M30 window, so that column is decay context rather than pure out-of-sample: the same asymmetry issue #3 disclosed for its own two columns.

Issue #3's six published cards: home number vs fresh-week replay



Issue #3's six published cards: the number its own issue printed (grey, in-sample on its own window) against the same configuration replayed verbatim on this week (green, out-of-sample). Four of six are net-negative on the fresh week; wk1 and m301 are the same configuration published on both of issue #3's windows, so their replays are identical by construction. Source: oos_cw3_wk.json + issue #3 open data.

Config	Family	Coins	Home	New WK (n)	New M30 (n)	Verdict
preset 4h4	fable, qwen, deepseek, opus	ETH+SOL	-- live	-\$376.3 (16)	+\$274.0 (44)	Faded
preset bc2	gemini, gpt, deepseek	SOL+XRP	-- live	+\$369.3 (9) *	+\$1,235.9 (21)	Survived
preset sol	deepseek, fable, opus	SOL [conc]	-- live	+\$486.8 (16)	+\$2,588.3 (39)	Survived
preset xrp	gemini, gpt, deepseek	XRP [conc]	-- live	-\$118.9 (3) * +	+\$1,155.2 (10)	Faded
cw3-m301	opus, deepseek, gemini	SOL+XRP	+\$2,137.6	-\$599.6 (26) +	+\$1,263.4 (81) +	Faded
cw3-m302	fable, opus, deepseek, gemini, qwen	BTC+ETH	+\$412.3	+\$218.3 (6) *	+\$229.8 (20)	Survived
cw3-s1	gemini	SOL+XRP	+\$1,604.3	-\$417.4 (38) +	-\$387.9 (121) +	Faded
cw3-s2	gemini	SOL+XRP	+\$1,449.0	+\$446.1 (14)	+\$1,741.7 (40) +	Survived
cw3-wk1	opus, deepseek, gemini	SOL+XRP	+\$2,175.1	-\$599.6 (26) +	+\$1,263.4 (81) +	Faded
cw3-wk2	opus, deepseek, gemini, grok	BTC+ETH	+\$837.6	-\$524.1 (11) +	-\$550.4 (52) +	Faded
cw2-30d1	opus, gpt, deepseek, gemini, qwen	BTC+ETH+ SOL+BNB+XRP [conc]	+\$744.7	+\$636.3 (10) +	+\$1,424.0 (31) +	Survived
cw2-30d2	opus, deepseek, gemini	ETH+SOL+ BNB+XRP [conc]	+\$655.1	+\$127.4 (16) +	+\$1,356.2 (48) +	Survived
cw2-30d3	opus, deepseek, gemini	BNB+XRP	+\$489.1	-\$81.7 (10)	+\$907.1 (30)	Faded
cw2-30ds1	gemini	BNB+XRP	+\$383.8	-\$82.1 (9) *	+\$749.2 (37)	Faded
cw2-7d1	fable, opus, deepseek, gemini	SOL+XRP	+\$299.1	-\$595.1 (25) +	+\$834.2 (79) +	Faded
cw2-7d2	gpt, deepseek, gemini, qwen	BNB+XRP	+\$223.7	+\$177.3 (5) *	+\$528.3 (25) +	Survived
cw2-pt1	fable, deepseek, grok [remap]	SOL	+\$698.3	+\$201.3 (17)	+\$1,277.2 (53)	Survived
cw2-pt2	opus	XRP	+\$543.7	-\$424.8 (2) *	+\$54.3 (13)	Faded
cw1-30d1	fable, deepseek, opus, qwen	ETH+SOL	+\$868.4	-\$349.8 (17)	-\$77.1 (46) +	Faded
cw1-30d2	gemini, gpt, deepseek	SOL+XRP+ETH	+\$590.6	+\$437.2 (13)	+\$1,114.7 (29)	Survived

Config	Family	Coins	Home	New WK (n)	New M30 (n)	Verdict
cw1-30d3	fable, deepseek, opus, qwen	BTC+ETH+SOL	+\$580.5	-\$366.8 (26)	-\$187.2 (66) +	Faded
cw1-30ds1	opus	SOL+XRP	+\$442.8	-\$122.8 (33) +	+\$326.7 (80) +	Faded
cw1-30ds2	grok [remap]	SOL+XRP	+\$262.9	-\$251.5 (24) +	+\$275.7 (69) +	Faded
cw1-30ds3	deepseek	XRP+BNB	+\$119.0	-\$354.7 (23) +	+\$5.1 (84) +	Faded
cw1-7d1	fable, deepseek, opus, gpt	SOL+XRP	+\$439.2	-\$97.8 (36)	+\$1,065.0 (112) +	Faded
cw1-7d2	fable, qwen, gemini, opus	XRP+BNB	+\$429.5	-\$621.7 (15) +	-\$300.8 (88) +	Faded
cw1-7d3	gpt, opus	BNB+XRP+SOL	+\$398.6	-\$462.0 (37) +	+\$416.6 (131) +	Faded
cw1-7ds1	gemini	XRP+BNB	+\$417.8	-\$569.2 (23) +	-\$565.4 (175) +	Faded
cw1-7ds2	gpt	XRP+BNB	+\$366.3	-\$153.2 (27) +	+\$378.4 (115)	Faded
cw1-7ds3	qwen	XRP+BNB	+\$267.0	-\$591.7 (20) +	-\$233.5 (103) +	Faded

n in parentheses; * = fewer than 10 closed trades (insufficient cell); + = non-zero skipped_nofunds (entries the engine could not fund -- the printed economics are then not the economics that ran). Home = the number the card's origin issue printed; the live presets have no home issue. [conc] = stake above this issue's cap for its universe size; [remap] = grok-4.5 replayed on the grok-4.6 lineage. Verdict rule: Survived = net-positive on both fresh windows, Faded = net-negative on at least one. Survivors are tinted green.

Counted: 7 of 26 past cards **Survive** both fresh windows (issue #3 on its two windows: 18 of 20) -- by cohort 2 of 6 issue-3 cards, 4 of 8 issue-2 cards, 1 of 12 issue-1 cards. Taken a window at a time: 7 of 26 are positive on the week and 19 of 26 on the four weeks, and every row that clears the week clears the month too, so the week decides every verdict in the table. The survival curve by age is the one ordering that holds on both: the oldest cohort is the emptiest (1 of 12 issue-1 cards on the week, 7 of 12 on the month). Live presets: 2 of 4 Survive, 2 of 4 positive on the week, 4 of 4 on the month.

The four live presets lead the table by standing rule. All four are net-positive on the four-week window (sol +\$2,588.3 on 39 closed, bc2 +\$1,235.9 on 21, xrp +\$1,155.2 on 10, 4h4 +\$274.0 on 44) and two of them are negative on the week (4h4 -\$376.3, xrp -\$118.9), so two Survive. **sol** and **xrp** are marked [conc]: both run a \$900 stake on a single coin, twice the \$450 cap the stake table sets for a 1-coin universe, so their printed economics assume funding this issue's own rule would not grant. Passports for all four are in the open data (presets_recon_wk.json and presets_recon_night.json).

Origin re-runs (drift baseline). Re-simulating the issue-3 cards on their OWN origin windows today lands within -\$251.1..\$0.0 of the printed issue-3 numbers: five of the six reproduce to the cent, and the exception is cw3-m302 (+\$412.3 printed, +\$161.2 today), the calibrated month card -- calibration is the one axis whose multipliers are learned from a forecast history that keeps growing. The issue-2 cards drift -\$177.7..+\$88.6, the same range issue #3 measured. Snapshot principle: no past issue is restated; the full drift table is in the open data (oos_walk_forward.drift_baseline).

Honest read. The two columns of this table disagree, and the disagreement is the finding. On the fresh week 19 of 26 past cards are net-negative, where seven days earlier 19 of 20 were positive; on the fresh four weeks 19 of 26 are net-positive. Same replay machinery, same frozen knobs, same rows -- only the window changed, which is what makes the week, not the configs, the most likely cause. Three things keep the verdict from being a statement about search. The fresh search moved the same way as the replays: its week ceiling fell to +\$399.4 from +\$2,175.1, while its month ceiling stayed at +\$1,676.0 -- when the ceiling and the replays move together, the common factor is the market. The month column is not clean evidence either: it shares 21 of its 28 days with issue #3's own M30 window, so a card that was tuned there is partly being scored on its own data. And the rows that lost most on the week are the ones with the largest in-sample numbers to give back -- cw3-wk1/m301 at -\$599.6 on the week against +\$2,175.1 printed at home. One bad OOS week, one good one, one bad one, and now a week and a month that point opposite ways. That is a series of weathers, not a strategy.

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

Matched pairs -- identical knobs, calibration OFF vs ON -- measured on net PnL. The table is the **whole stage-2 grid** of each window, and the **Both** row pools the two windows' pairs and nothing else. The board-leader slice is a second, separate measurement (21 WK pairs and 15 M30 pairs whose OFF or ON leg reached the top of a board); it is reported in the paragraph below and in the open data, and the two are never pooled together.

Window	Pairs	Calibrate wins	Mean delta	Best delta	Worst delta
WK	1,468	61.0%	+\$123.3	+\$777.1	-\$389.3
M30	1,414	29.1%	-\$111.0	+\$663.4	-\$977.2
Both	2,882	45.4%	+\$8.3	+\$777.1	-\$977.2

The two windows answer the axis differently, and this is the first issue in which they do. On the week calibration improves 61.0% of the 1,468 matched pairs (mean +\$123.3, median +\$59.6) and 88.3% of the 409 mature pairs (mean +\$229.0); on the four weeks it improves 29.1% of 1,414 (mean -\$111.0, median -\$82.9) and 35.3% of the 842 mature pairs (mean -\$119.3). Pooled over both windows the axis improves 45.4% of 2,882 pairs with a mean of +\$8.3 -- a number that describes neither window. The board-leader slices are separate and disagree with the population on the month: 52.4% of the 21 WK slice pairs improve (mean -\$13.8, median +\$52.1, OFF legs already at a median of +\$441.6), and 0.0% of the 15 M30 slice pairs (mean -\$619.2, OFF legs at +\$1,582.4). Issue #3 measured a 400-pair top slice and found calibration losing 397 of 400. Where the axis reached the cards is visible above: both published WK councils are calibrate=ON cells and 26 of the 40 WK finalists run it, while neither M30 card is and 35 of the 62 M30 finalists run it.

The open-data pack also carries a 400-pair export (calibrate_effect) per collect. Its delta distribution is not the population's (median +\$352.7 against the WK grid's +\$59.6), so no claim in this section is computed from it; it is published for inspection only.

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 (>=15 closed on WK, >=20 on M30), one config per model.

Solo top -- WK (Aug 24-31, 2026)

#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	deepseek	BTC+ETH+SOL+BNB+XRP	1d/1d,4h	+\$325.0	+32.5%	12.3%	23	91.3%
#2	gemini	BTC+ETH+SOL+BNB+XRP	1d/1d,4h	+\$227.2	+22.7%	34.4%	27	81.5%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	median	1/0	reenter	\$200
#2	50:40,100:60	on	TP -10 / SL +75	>=60 (each)	farthest	1/0	reenter	\$200

Both qualifying solos are the 5-coin universe at the \$200 table stake, the same collapse the council board shows. The two single-coin rows that outrank them on the mature net board (qwen on SOL at +\$317.2, opus on SOL at +\$300.4) are excluded by the >=2-coin rule and appear under Per-ticker bests; neither solo card carries an unfunded entry. Source: summary_wk.json top_mature_by_window_branch.

Solo top -- M30 (Aug 3-31, 2026)

#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	gemini	SOL+XRP	1d/1d,4h	+\$1,741.7	+174.2%	54.9%	40 *	75.0%
#	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

One solo card on M30, not two. Every one of the ten rows on the M30 mature solo board is the same model (gemini) on the same universe (SOL+XRP), so the one-config-per-model rule admits exactly one; the card carries 2 unfunded entries (flagged * on Trades) and its clean twin, 0 skips, makes +\$1,414.9 on 42 closed. Source: top_mature_by_window_branch in the night collect.

OPEN IN SANDBOX marketmania.ai/s/cw4-s1

Each link opens this exact frozen window; results visible without sign-in (embedded share signature). The codes are minted at publish time and are NOT live in this draft; every card's full frozen query signature is in the open-data JSON next to its simulation index.

Council vs solo. On the week the best solo sits below the best council (+\$325.0 vs +\$399.4 on the mature board), as it did on both of issue #3's windows. On the four weeks it does not: the solo card makes +\$1,741.7 against the council card's +\$1,676.0, on 40 closed trades against 31, with a 54.9% drawdown against 53.6% and 2 unfunded entries against 0. The ensemble edge is a week-scale result this issue, not a month-scale one. Same in-sample caveats apply to both sides.

Quality boards -- Smoothness joins this issue

Net PnL stays the primary board (continuity with issues #1-#3); four selection views sit on top of it at zero additional simulations. **Calmar-like** = net / max(maxDD, 1.0), positive net only; **WR board** = win rate among mature configs (>=15 closed on WK, >=20 on M30); **Smoothness** (new) = R2 of a linear fit through the daily closed-equity series, mature rows with net > 0 and at least 5 days, ties broken by ulcer index; **Quality composite** (new) = mean rank over net, Calmar-like, win rate and R2, lower is better. Top-3 of each board on each window below; all 15 rows of all eight boards are in the open-data JSON.

Board	Config	Coins	FH/TF	Net	WR	maxDD	CIs	R2
WK Calmar	opus, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$399.4	95.5%	0.1%	22	0.898
WK Calmar	gpt, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$374.7	95.2%	0.1%	21	0.898
WK Calmar	deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$373.4	95.5%	0.1%	22	0.897
WK WR	opus, gpt, deepseek, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$308.8	95.8%	0.1%	24	0.898
WK WR	opus, gpt, deepseek, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$231.6	95.8%	0.1%	24	0.898
WK WR	opus, gpt, deepseek, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$306.6	95.7%	0.1%	23	0.898
WK Smoothness	fable, deepseek [cal]	BNB	4h/4h	+\$99.3	90.5%	0.0%	21	0.975
WK Smoothness	fable, deepseek, qwen [cal]	BNB+XRP	4h/1h	+\$128.2	83.3%	0.0%	18	0.951
WK Smoothness	fable, deepseek, qwen [cal]	BNB+XRP	4h/1h	+\$128.2	83.3%	0.0%	18	0.951
WK Quality	opus, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$399.4	95.5%	0.1%	22	0.898
WK Quality	gpt, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$374.7	95.2%	0.1%	21	0.898
WK Quality	deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	+\$373.4	95.5%	0.1%	22	0.897
M30 Calmar	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 Calmar	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 Calmar	fable, opus, deepseek [cal]	SOL	4h/1h	+\$866.2	94.2%	0.0%	52	0.978
M30 WR	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 WR	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 WR	gemini, gpt, deepseek [cal]	SOL+XRP	1d/1d	+\$969.4	96.2%	13.6%	26	0.964
M30 Smoothness	fable, opus, deepseek [cal]	SOL	4h/1h	+\$866.2	94.2%	0.0%	52	0.978
M30 Smoothness	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977

Board	Config	Coins	FH/TF	Net	WR	maxDD	Cls	R2
M30 Smoothness	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 Quality	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 Quality	gemini [cal]	SOL	1d/1d,4h	+\$941.7	100.0%	0.0%	20	0.977
M30 Quality	fable, opus, deepseek [cal]	SOL	4h/1h	+\$866.2	94.2%	0.0%	52	0.978

[cal] = calibrate=ON cell; the board column names the window. Both Calmar boards are degenerate by construction -- their top rows sit at 0.1% (WK) and 0.0% (M30) maxDD, so the score collapses onto net PnL; read them as 'net among configs that barely drew down'. The two Smoothness boards are the ones that leave each window's dominant universe: on WK its top rows are BNB and BNB+XRP configs with nearly straight equity lines (R2 0.975 at the top) and small nets (+\$99.3); on M30 they are single-coin SOL configs (R2 0.978, +\$866.2). That trade -- straightness bought with size -- is exactly what the board exists to expose. The M30 boards are drawn from a mature pool of 3,184 configs against 1,774 on WK, because four weeks clear a >=20-trade gate that one week does not.

Stability, and what it is across this issue. The board is defined as presence in the mature net top-40 of BOTH windows of the run that computes it. The run that produced it here is the night collect, whose two windows are M30 and the calendar month -- so this issue's stability board reads **stable across M30 and the calendar month**, not across WK and M30, and it is not comparable with issue #3's WK-and-M30 board. 25 configs qualify, all of them SOL+XRP, 19 of them councils and 14 clean (zero unfunded entries on both windows). The top three by rank sum are solo gemini (+\$1,741.7); solo gemini (+\$1,741.7); solo gemini (+\$1,608.1). The WK window is not part of this measurement: the weekly collect has one window and a cross-window board cannot be built inside it. Full list in the open data (stability).

Hand-tuned baseline vs the grid

New standing section. One hand-tuned configuration is treated as the reference the search has to beat -- not on net PnL, where any large search wins by construction, but on the three properties a configuration is actually tuned for: **win rate, shallow drawdown and a straight equity line**. The reference (referred to below as the **hand-tuned anchor**) is the live preset 1d_3_llm_BC2_v02, a 3-model 1d council on SOL+XRP at the \$450 two-coin stake; it is distinct from the pre-search hand-tuned set A/B/C carried in Baselines below. Its replay row is in the walk-forward table above; here it is the yardstick.

Beat rule (frozen with the criteria): a config beats the anchor only if it does so on all three at once -- win rate higher, maxDD lower and smoothness R2 higher -- among mature, non-concentrated search rows. Net PnL is not part of the rule; it is reported next to the winners so the cost of the improvement is visible.

The anchor on each window

Window	Config	Coins	FH/TF	Net	WR	maxDD	R2	Ulcer	Cls
WK	Hand-tuned anchor (1d_3_llm_BC2_v02)	SOL+XRP	1d/1d	+\$369.3	77.8%	31.9%	0.308	10.24	9
M30	Hand-tuned anchor (1d_3_llm_BC2_v02)	SOL+XRP	1d/1d	+\$1,235.9	85.7%	31.9%	0.914	4.24	21

187 of the 1,774 mature WK configs (10.5%) clear all three bars at once, and 19 of the 3,184 mature M30 configs (0.6%). The anchor's own two windows are different rows: on the week +\$369.3 net at 77.8% win rate with a 31.9% drawdown, an R2 of 0.308 and 9 closed trades; on the four weeks +\$1,235.9 at 85.7%, the same 31.9% drawdown, an R2 of 0.914 and 21 closed. The week row is thin against the >=15 gate the search rows must pass (the anchor is exempt from that gate by construction: it is the reference, not a candidate). Of the ten best WK winners by net, 6 also beat the anchor on net PnL; of the ten best M30 winners, 0 do -- on the month the anchor's own +\$1,235.9 is above every config that beats it on the three quality bars, which is the cleanest statement this section has produced so far: the search can buy a straighter, shallower, more accurate line on that window only by giving up money. The share itself moves with the window (10.5% of the WK field, 0.6% of the M30 field), so read it as one field at a time, not as a property of the anchor.

Top-10 by net among the 187 WK configs that beat the anchor on win rate, maxDD and smoothness at once

#	Config	Coins	FH/TF	Stake	Net	WR	maxDD	R2
#1	opus, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$399.4	95.5%	0.1%	0.898
#2	gpt, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$374.7	95.2%	0.1%	0.898
#3	gpt, deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/4h	\$200	+\$373.9	95.0%	1.1%	0.819
#4	deepseek, gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$373.4	95.5%	0.1%	0.897
#5	gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$372.2	94.7%	3.4%	0.808
#6	gemini, qwen [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$372.2	94.7%	3.4%	0.808
#7	deepseek, gemini, qwen, grok [cal]	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$368.9	95.2%	0.1%	0.896
#8	opus, deepseek, gemini, qwen	BTC+ETH+SOL+BNB+XRP	1d/4h	\$200	+\$349.8	80.0%	10.5%	0.507
#9	opus, deepseek, gemini, qwen	BTC+ETH+SOL+BNB+XRP	1d/4h	\$200	+\$349.8	80.0%	10.5%	0.507
#10	gemini, gpt, deepseek	BTC+ETH+SOL+BNB+XRP	1d/1d	\$200	+\$339.8	85.0%	13.0%	0.791

Top-5 by net among the 19 M30 configs that beat the anchor on win rate, maxDD and smoothness at once

#	Config	Coins	FH/TF	Stake	Net	WR	maxDD	R2
#1	gemini [cal]	SOL+XRP	1d/1d,4h	\$450	+\$1,101.4	87.2%	14.3%	0.937
#2	gemini [cal]	SOL+XRP	1d/1d,4h	\$450	+\$1,101.4	87.2%	14.3%	0.937
#3	gemini, gpt, deepseek	SOL+XRP+ETH	1d/1d	\$300	+\$1,040.0	86.2%	17.6%	0.942
#4	gemini, gpt, deepseek [cal]	SOL+XRP	1d/1d	\$450	+\$969.4	96.2%	13.6%	0.964
#5	gemini [cal]	SOL	1d/1d,4h	\$450	+\$941.7	100.0%	0.0%	0.977

Anchor passport (verbatim, as stored): members=gemini-3.1-pro,gpt-5.6-sol,deepseek-v4-pro tokens=SOL,XRP fh=1d tf=1d conf_mode=each min_conf=60 max_sideways=2 max_diff_side=0 tp_sl_source=nearest trade_mode=position same_side=update opposite=reverse no_signal=hold time_stop=3x steps=40:80,100:20 steps_on=1 be=1 tp_shift=30 sl_shift=50 deposit=1000 lev=10 refill=0 stake=450 calibrate=- variant=v0

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 585 + 1,160 sims -- a compacted grid, one best per coin per window; the branch tag in brackets says which kind won.

WK (Aug 24-31, 2026)

Coin	Best config (branch)	FH/TF	Net (trades)
BTC	gemini, gpt, deepseek [council]	1d/1d	+\$265.7 (5) *
ETH	deepseek [solo]	1d/1d	+\$331.5 (6) *
SOL	gemini, gpt, deepseek [council]	1d/1d	+\$530.3 (8) *
BNB	deepseek [solo]	1d/1d,4h	+\$121.3 (5) *
XRP	fable, gemini [council]	1d/1d	+\$222.9 (3) *

M30 (Aug 3-31, 2026)

Coin	Best config (branch)	FH/TF	Net (trades)
BTC	gpt [solo]	1d/1d,4h	+\$550.2 (22)
ETH	gpt, deepseek, gemini, qwen [council]	1d/1d	+\$511.3 (11)
SOL	fable, opus, deepseek [council]	4h/1h	+\$1,334.1 (57)
BNB	gemini, qwen [council]	1d/1d	+\$715.7 (14)
XRP	opus, deepseek, gemini [council]	4h/1h	+\$782.0 (71)

n in parentheses; * = fewer than 10 closed trades (thin -- direction only). All ten winners run calibrate OFF; 3 of the five WK winners and 4 of the five M30 winners are councils.

The reads. Every coin is nominally positive on both windows. On the week SOL leads at +\$530.3 but **all five rows are thin** -- 3 to 8 closed trades against a 15-trade gate, the first time in this series that not one per-coin winner reaches it -- so the WK half of this table carries less evidence than any other table in the issue. On the four weeks nothing is thin: 11 to 71 closed, SOL leading at +\$1,334.1, and the same coin leads both windows. The 1d horizon wins every WK coin and 3 of the five M30 coins. **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 both of this issue's frozen windows inside the same two collects (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) -- with the best and worst of the seven solo defaults on each window; all eleven rows per window and both stake modes are in the open data.

WK (Aug 24-31, 2026)

Baseline	Net PnL	Trades	WR	maxDD
Best solo default (claude-opus-5)	-\$14.5 *	211	63%	3.8%
Worst solo default (deepseek-v4-pro)	-\$52.9 *	320	58%	5.3%
Council-of-7 default	-\$35.1 *	133	56%	3.5%
Hand-tuned config A (pre-search)	-\$32.8 *	11	73%	50.1%
Hand-tuned config B (pre-search)	-\$412.8 *	42	55%	41.3%
Hand-tuned config C (pre-search)	-\$349.8	17	47%	45.8%

M30 (Aug 3-31, 2026)

Baseline	Net PnL	Trades	WR	maxDD
Best solo default (gemini-3.1-pro)	-\$38.1 *	447	56%	8.9%
Worst solo default (deepseek-v4-pro)	-\$153.8 *	954	55%	15.4%
Council-of-7 default	-\$63.2 *	400	57%	6.3%
Hand-tuned config A (pre-search)	+\$558.2	26	81%	50.1%
Hand-tuned config B (pre-search)	+\$319.2 *	139	66%	68.1%
Hand-tuned config C (pre-search)	-\$77.1 *	46	59%	45.8%

* = non-zero skipped_nofunds at table stake. On the 5-coin defaults those counts are the stake table doing its job: at \$200 a \$1,000 deposit funds at most five concurrent positions while the defaults fire hundreds of entries per window (101 to 336 skipped each on the week, 350 to 1,125 on the four weeks). The other five solo defaults run between the two printed rows on each window, and both stake modes carry the same sign for every row -- **the conclusions do not change between modes**, with one exception, printed as it stands: hand-tuned C on M30 is +\$60.8 at the config's own stake and -\$77.1 at the table stake. Hand-tuned A and B are already at their \$300 3-coin cap, so their two modes are identical on both windows. Source: baselines_wk.json (22 sims) and baselines_night.json (44 sims, of which 22 are the M30 half), inside this issue's two collects.

The week takes every reference point down; the month puts two of them back. On WK the seven solo defaults run -\$14.5 to -\$52.9 at table stake, the council-of-7 -\$35.1, and the three hand-tuned configs -\$32.8 /

-\$412.8 / -\$349.8 -- everything negative, in both stake modes, which inverts the lower half of the standing ordering (search > hand-tuned > defaults): hand-tuned B alone gives back -\$412.8, more than any default. On M30 the ordering is back: every default is negative (-\$38.1 to -\$153.8 solo, council-of-7 -\$63.2) while hand-tuned A and B are positive (+\$558.2 / +\$319.2) and C is -\$77.1. The upper half holds on both windows, in-sample as ever -- the fresh search makes +\$399.4 on the week against a best baseline of -\$14.5, and +\$1,676.0 on the four weeks against +\$558.2. The live presets are a separate row of evidence: 2 of the four are net-positive on the week and 4 of four on the month (walk-forward table above), and two of them run stakes above this issue's cap.

Live presets & preset neighbourhoods

Four hand-tuned configurations trade live on the platform (not search output, distinct from the pre-search A/B/C set): preset bc2 = '1d_3_llm_BC2_v02'; preset 4h4 = '4h_4_llm_v02'; preset sol = 'Sol'; preset xrp = 'XRP'. Their replay rows lead the walk-forward table; passports and window detail are in the open data (presets_recon_wk.json and presets_recon_night.json). The neighbourhood grid asks a narrower question: does a one-knob neighbour beat the preset as configured? Two of the four have a local grid this issue (53 cells per window); the two single-coin presets do not (0 cells -- the local grid is not built for them, so no neighbour claim is made about either).

Live preset	Window	As configured	Best neighbour in local grid	Cells	Verdict
preset bc2	WK	+\$369.3	+\$541.3 (1d/1d, SOL+XRP, 12 cls)	34	neighbour +\$172.0
preset bc2	M30	+\$1,235.9	+\$1,568.7 (1d/1d, SOL+XRP, 32 cls)	34	neighbour +\$332.8
preset 4h4	WK	-\$376.3	+\$258.4 (1d/4h, ETH+SOL, 7 cls)	19	neighbour +\$634.7
preset 4h4	M30	+\$274.0	+\$574.9 (4h/1h, ETH+SOL, 50 cls)	19	neighbour +\$300.9

In **all four** preset-window cells with a grid, a one-knob neighbour beat the configuration as it is actually running: bc2 by +\$172.0 on the week and +\$332.8 on the four weeks, 4h4 by +\$634.7 and +\$300.9. The largest single edit is 4h4's horizon move (4h/1h -> 1d/4h), which turns -\$376.3 into +\$258.4 on the week. The presets are not at a local optimum -- a testable, low-risk edit, unlike adopting a search champion wholesale. Cell counts are small (34 and 19 per window) and every neighbour is an in-sample maximum of its own little grid.

Patterns: what the finalists look like

Knob modes across each window's unique finalists (mature net top-10 per branch plus all four quality boards, deduplicated by full signature; 40 unique configs on WK, 30 of them councils; 62 on M30, 48 councils).

Knob	WK finalists mode	M30 finalists mode
Forecast horizon	1d (31/40)	1d (49/62)
Ladder shape (steps)	50:40,100:60 (39/40)	50:40,100:60 (56/62)
Break-even stop (BE)	on (40/40)	on (62/62)
Min confidence	60 (40/40)	60 (62/62)
SL shift	75 (39/40)	75 (56/62)
TP shift	-10 (39/40)	-10 (56/62)
TP/SL source	farthest (34/40)	farthest (50/62)
Calibrate ON	26/40	35/62
Universe	BTC+ETH+SOL+BNB+XRP (31/40)	SOL+XRP (36/62)

The mechanical core is unchanged for the fourth issue running and it is the same on both windows: 50:40,100:60 two-rung ladder, break-even on, conf 60, SL +75 / TP -10, farthest source. The horizon flipped back to **1d** on both (31/40 on WK, 49/62 on M30) after issue #3's 4h week -- 4h won the week the market moved, 1d wins both windows that end in a flat one. Two things the windows do NOT share. **Calibrate=ON** runs in 26 of 40 WK finalists against 35 of 62 on M30, where issue #3 measured 0/38 on its week and 15/41 on its month. And the universe: BTC+ETH+SOL+BNB+XRP takes 31 of the 40 WK finalists while SOL+XRP takes 36 of the 62 M30 finalists. Membership concentrates differently too -- qwen leads the WK council finalists (27 of 30) and gemini the M30 ones (42 of 48), with deepseek second on both (25 and 38). Standing caveats: refinement seeds from the same grid winners, so part of the convergence is search-design echo; and a knob core that produced 18-of-20 survival one issue ago and 7-of-26 this issue 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 the published cards are unusually shallow on the week (0.1% and 0.0% maxDD) and not on the month (53.6% and 11.3%), and the replay table is deeper still: 15 replayed rows drew down more than 40% on the week and 19 on the month, 9 and 12 of them past 55%. A high headline net and a survivable path are different claims, and so are a shallow in-sample drawdown and a shallow one next week.

Watch amendments (this issue)

Amendment 3 -- smoothness and a quality composite join the boards. Every simulation now stores its daily closed-equity series, and from it three numbers per row: `n_days`, `smooth_r2` (R2 of a linear fit, undefined below 4 days) and `ulcer` (root-mean-square drawdown across the daily points). The smoothness board ranks mature rows with `net > 0` and at least 5 days by R2, ties broken by ulcer; the quality composite ranks mature rows by the mean of their net, Calmar-like, win-rate and R2 ranks. Both cost zero additional simulations. The headline board stays net PnL.

Amendment 4 -- the monthly split. From this issue the weekly Config Watch keeps two windows -- the calendar week (WK) and the four calendar weeks that end at the same edge (M30, Aug 3-31, 2026) -- and the **calendar-month** board moves to a separate **Monthly Config Watch** issue produced by the same engine, the same criteria and the same collect code. The two windows of this issue come from two collects: WK from the weekly run, M30 from the night run that also computes the calendar month. One consequence is stated wherever it appears: the **stability** board is built inside the run that computes it, so this issue's is defined across M30 and the calendar month, not across WK and M30, and is not comparable with issue #3's. The walk-forward verdict rule is unchanged and is evaluated on both fresh windows (see Walk-forward above).

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

- **Two windows, one regime, and they overlap.** M30 (Aug 3-31, 2026) contains WK (Aug 24-31, 2026) and shares 21 of its 28 days with issue #3's own M30 window. The two columns of this issue are therefore not two independent tests: agreement between them is partly arithmetic, and the M30 replay column is decay context rather than pure out-of-sample. Do not read them as a cross-validation.
- **The two windows come from two collects.** WK is the weekly run (generated 2026-09-02T08:54:43Z), M30 the night run (2026-09-03T12:07:22Z), which computed the calendar-month boards in the same pass. Same engine, same criteria, same collect code, but not the same execution -- a number is comparable across the two windows only as far as that is.
- **Stability is defined inside the run that computes it.** This issue's stability board comes from the night run and therefore reads across M30 and the calendar month, not across WK and M30. It answers a different question from issue #3's board of the same name and the two counts are not comparable.
- **Model lineage is a splice.** The grok axis runs as grok-4.6 (incl. 4.5-era) -- one lineage across a mid-series model cutover -- and issue-1 and issue-2 configs that named grok-4.5 are replayed on it ([remap] rows). A remapped replay is not the same simulation the origin issue ran.
- **Concentrated rows are published, not hidden.** Two of the four live presets (sol, xrp) and two issue-2 cards run stakes above this issue's cap for their universe size and are marked [conc]; their printed economics assume funding the current rule would not grant. They stay in the table because removing them would flatter the preset row.
- **The stake rule changes what runs, not only its size.** Capping the stake changes WHICH entries get funded: the seven table-stake solo defaults skip 101 to 336 entries each, and 16 replay rows carry unfunded entries (up to 34 on cw1-7ds1). Where `skipped_nofunds` is non-zero the printed economics are not the economics that ran; the rows are flagged, never silently pooled.
- **Snapshot principle.** This report is a frozen snapshot taken at the generated-at timestamp; numbers are not updated retroactively and past issues are not restated. The drift baseline exists precisely because today's engine and a longer forecast history do not reproduce every past number (cw3-m302: +\$412.3 printed, +\$161.2 today).

- **Research-to-date counter is pinned, and its model line changes definition.** The counter below is pinned from wave 3 onward; issue #3 printed 11 models tracked (7 current + 4 archived legacy) and this issue prints 7 (current line-up; earlier versions folded into their successors' lineage). The two model counts are not comparable and issue #3 is not restated.
- **The monthly issue follows.** Any claim about calendar August -- the calendar-month board, its cards, its per-ticker bests -- belongs to Monthly Config Watch #1 and is not made here. This issue carries the four-week M30 window; it does not carry the calendar month.
- **Multiple testing.** 16,447 simulations across the two collects; 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 and thin cells.** Every card, board row and per-ticker best is the maximum of a search, biased high by selection alone. On the week the thinness is unusual: all five per-ticker winners closed 3-8 trades, the net-board maximum (+\$623.1) closed 9, and 6 of the 30 replay rows are under 10 closed. On the four weeks no replay row is thin and no per-ticker winner is. Flagged with *, reported for completeness.
- Research output, not financial advice.

Counters & lineage

Weekly collect (WK window)

Stage	Sims	Stage	Sims
Scan (s1)	476	Dedicated per-ticker (s6t+s6tc)	585
Systematic grid (s2)	4,208	OOS replays (s4cw3+s4cw3o+s4cw2+s4cw2o+s4cw1+s4pre)	44
Refinement: ladder + stake sweep (s2b+s3b)	122	Total	5,558
Preset neighbourhood grid (s2p)	53	Errors / zero-trade	0 / 1,021
Per-ticker probes (s5t)	70	Engine	1.1

Night collect (M30 and calendar-month windows together; the calendar-month boards ship with Monthly Config Watch #1)

Stage	Sims	Stage	Sims
Scan (s1)	952	Dedicated per-ticker (s6t+s6tc)	1,160
Systematic grid (s2)	8,224	OOS replays (s4cw3+s4cw3o+s4cw2+s4cw2o+s4cw1+s4pre)	74
Refinement: ladder + stake sweep (s2b+s3b)	243	Total	10,889
Preset neighbourhood grid (s2p)	106	Errors / zero-trade	0 / 1,188
Per-ticker probes (s5t)	130	Engine	1.1

Scale note (vs issue #3). Issue #3 ran 11,136 sims over two windows in one pipeline; issue #4 runs 16,447 over three windows in two -- 5,558 in the weekly collect (WK) and 10,889 in the night collect (M30 and the calendar month together, the second of which ships with Monthly Config Watch #1). The per-window grid is close to unchanged: 4,208 WK cells and 4,112 M30 cells here against 4,272 per window there. The per-coin branch is the reduction (585 and 1,160 against 1,150); the replay branch grew in rows (30 configs against 22) and in sims (44 and 74 against 52), because every row is replayed on every window of its collect. The two new boards cost zero simulations.

No extras pass in either collect. Equity series, the grid histograms (0 extra sims), the baselines and the calibrate packs are produced by the same two collects that wrote the boards -- issue #3's separate same-day extras pass (w167) is gone. The baseline runs are verification, outside both search totals: 22 sims on WK (11 configs x 2 stake modes) and 44 in the night collect, of which the 22 M30 rows belong to this issue and the 22 calendar-month rows to Monthly Config Watch #1. 0 equity replays were needed.

Generated at: 2026-09-02T08:54:43Z (weekly collect) and 2026-09-03T12:07:22Z (night collect) · 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 timestamps above; numbers are not updated retroactively. Each issue re-runs the pipeline fresh over that issue's windows.

Research to date

RESEARCH SNAPSHOT

THIS REPORT -- search effort: 16,447 simulations (0 errors) across two frozen collects, of which 118 walk-forward replays and origin re-runs · 66 verification sims (baseline runs inside the same collects; 0 equity replays needed)

MARKETMANIA RESEARCH TO DATE (as of cutoff): 38,385 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 18 published reports

Counter pinned from wave 3 onward. The model line changes definition this issue: issue #3 printed 11 models tracked (7 current + 4 archived legacy); from this issue the line-up counts 7, with earlier versions folded into their successors' lineage. The two are not comparable and issue #3 is not restated. Config Watch #4 is the 18th published report (Consensus Watch #4, Weekly Calibration #4 and Weekly Model Watch #4 are 15, 16 and 17).

Issue #4. Two windows again -- the calendar week and the four calendar weeks ending at the same edge -- and they disagree: the walk-forward table keeps 7 of 26 past cards where the week alone keeps 7 and the month alone 19. The calendar-month board moves to Monthly Config Watch #1. Two selection views join net PnL at zero simulation cost -- a smoothness board (R2 of the daily closed-equity line, ulcer as tie-break) and a rank-average quality composite -- and a new standing section measures a live hand-tuned configuration against the whole mature field of each window on win rate, drawdown and smoothness at once. Formalization of the amendments is still slated for methodology v1.2.

What we're testing next

- **Monthly Config Watch #1** -- the calendar-month window (2026-08-01 to 2026-09-01) from the same night collect: same engine, same criteria, same collect code, published as its own issue. It carries the calendar-month cards, its own per-ticker bests and its own baselines; this issue carries none of them.
- **Config Watch #5** -- the next Monday-aligned pair of windows, with the first out-of-sample read on THIS issue's four cards: a calibrated 5-coin council at the \$200 table stake, a smoothness-board pick, and two uncalibrated SOL+XRP and BNB+XRP councils from the month.
- Whether the calibrate split is a regime effect or a window-length effect: the same axis improves 61.0% of the WK pairs and 29.1% of the M30 pairs in this one collect pair. The test is to replay both issues' matched pairs across both window lengths.
- The universe collapse moved rather than broke -- SOL+XRP owns the month board, the 5-coin universe owns the week. A per-universe board ranking each field against its own peers is the next thing to try.
- **The quarterly** -- the first cross-issue aggregate of the walk-forward table: four issues of verdicts on one axis, and the survival curve by cohort age plotted rather than counted.

Related research

Consensus Watch #4 -- <https://marketmania.ai/research/reports/consensus-watch-2026-08-24.pdf>

Weekly Calibration #4 -- <https://marketmania.ai/research/reports/weekly-calibration-2026-08-24.pdf>

Weekly Model Watch #4 -- <https://marketmania.ai/research/reports/model-watch-2026-08-24.pdf>

Config Watch #3 -- <https://marketmania.ai/research/reports/config-watch-2026-08-19.pdf>

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

Config Watch #1 -- <https://marketmania.ai/research/reports/config-watch-2026-08-05.pdf>

The three weekly reports publish together as one issue each week; Config Watch follows its own cycle. Market-regime figures, where this issue refers to them, come from that wave and are not re-tabulated here.

CITE THIS REPORT

```
@techreport{mm_configwatch_2026w36,
  title      = {Config Watch #4},
  author     = {{MarketMania Research}},
  institution = {MarketMania},
  year       = {2026},
  month      = sep,
  day        = {4},
  type       = {Weekly Research Report},
  series     = {Config Watch},
  number     = {4},
  note       = {Methodology v1.1 + Config Watch amendments, hash e66c7e8c864a2233},
  url        = {https://marketmania.ai/research/reports/config-watch-2026-08-26.pdf}
}
```