

Config Watch #1

August 13, 2026 · MarketMania Research · Weekly series

WEEKLY · CONFIG WATCH · Issue #1 · Issue date August 13, 2026 · Language English · Methodology v1.1 + Config Watch amendments (owner-approved 12.08), hash e66c7e8c864a2233

Search windows: **7d** Aug 5-12, 2026 and **30d** Jul 13-Aug 12, 2026 (frozen date ranges). Both windows are frozen date ranges, not the Mon-Mon reporting window: issue #1 ran ad-hoc. From issue #2 onward, windows align to Mondays.

RESEARCH SNAPSHOT			
15,416 sims total (12,199 council + 3,217 solo)	7d branch 7,307 sims (Aug 5-12, 2026)	30d branch 4,892 sims (Jul 13-Aug 12, 2026)	3-stage search: scan -> grid -> waves
engine: \$1,000 deposit, 10x leverage	TOP-3 rule: ≥ 2 coins, distinct universes	methodology v1.1 + Config Watch amendments	Issue #1 -- new series

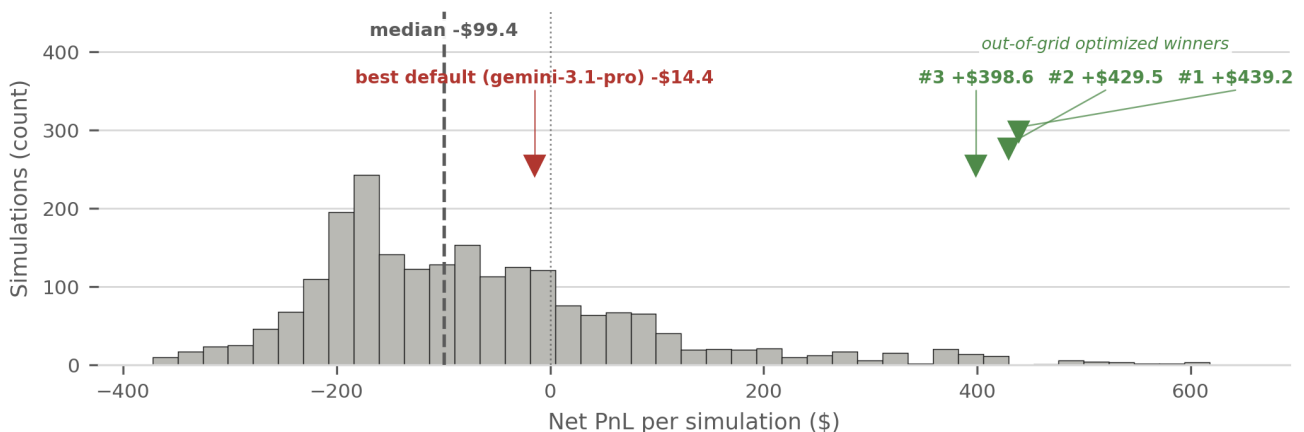
KEY FINDING · OBSERVATION (one weekly window)

Search found configurations that materially outperformed the default setup in-sample; whether that edge survives starts next week.

TL;DR

- OBSERVATION -- Search found configurations that materially outperformed the default setup in-sample on both windows; whether that edge survives out-of-sample starts next week.
- Defaults lost on both windows: 7d grid median $-\$99.4$ (24.7% positive); 30d grid median $-\$221.0$ (2.4% positive). All 7 solo-model defaults and the council-of-7 default were negative on both windows.
- Top result per window beat the best hand-tuned pre-search config: $+\$439.2$ (7d) and $+\$868.4$ (30d) -- search validation passed against independent, pre-search baselines.
- Cross-window robustness check: most tuned configs are fragile across windows; one family (fable-5+deepseek-v4-pro+opus-5 core, 4h/1h, ladder+BE, conf ≥ 60 , farthest source) showed a cross-window robustness hint -- this is not independent validation. Real out-of-sample testing starts next week.

Stage-1b systematic grid -- net PnL distribution -- 7d window (Aug 5-12, 2026) -- n=2,160



Net PnL across all 2,160 stage-1b grid sims, 7d window (Aug 5-12, 2026). Grid median $-\$99.4$, 24.7% positive.

Red marker = best solo default; green = the three TOP-3 configs detailed below (waves stage, which searches beyond this grid).

Why it matters

MarketMania publishes default LLM-council trading configs. This series asks a narrower, more useful 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 establishes the in-sample (IS) baseline: search scale, what wins, and a first, honest look at whether winners on one window still win on another. It is not yet an out-of-sample (OOS) verdict -- that begins next week.

How we searched

Three stages per window, run independently for 7d and 30d:

- **1. Council scan (stage-1a)** -- 600 sims/branch across council compositions and coarse settings.
- **2. Skeleton grid (stage-1b)** -- a systematic grid (2,160 configs on 7d; 2,172 on 30d) sweeping declared axes: archetype, forecast horizon vs. timeframe, entry (max-sideways / max-diff-side), universe size, council size, model membership. This grid is the denominator for every percentile claim in this report.
- **3. Coordinate-ascent waves** -- local refinement beyond the grid (4,547 sims on 7d; 2,120 on 30d), producing the published TOP-3 and the solo-coin side note.

Stage design: owner. Fixed economics (deposit, leverage, stake table, fees, liquidation modeling) are held constant across every sim in every stage.

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-4.5.

TOP-3 -- 7d window (Aug 5-12, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	fable-5, deepseek-v4-pro, opus-5, gpt-5.6-sol	SOL+XRP	4h/1h	+\$439.2	+43.9%	5.0%	32	68%
#2	fable-5, qwen-3.8-max, gemini-3.1-pro, opus-5	XRP+BNB	4h/4h	+\$429.5	+43.0%	8.0%	19	76%
#3	gpt-5.6-sol, opus-5	BNB+XRP+SOL	4h/1h	+\$398.6	+39.9%	0.4%	21	85%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:40, 100:60	on	TP -10 / SL +50	>=60 (each)	farthest	4/0	reenter	\$450
#2	100:100	off	TP -10 / SL +40	>=50 (each)	median	4/2	keep	\$450
#3	80:50, 100:50	off	TP +10 / SL +20	>=60 (each)	nearest	1/1	keep	\$300

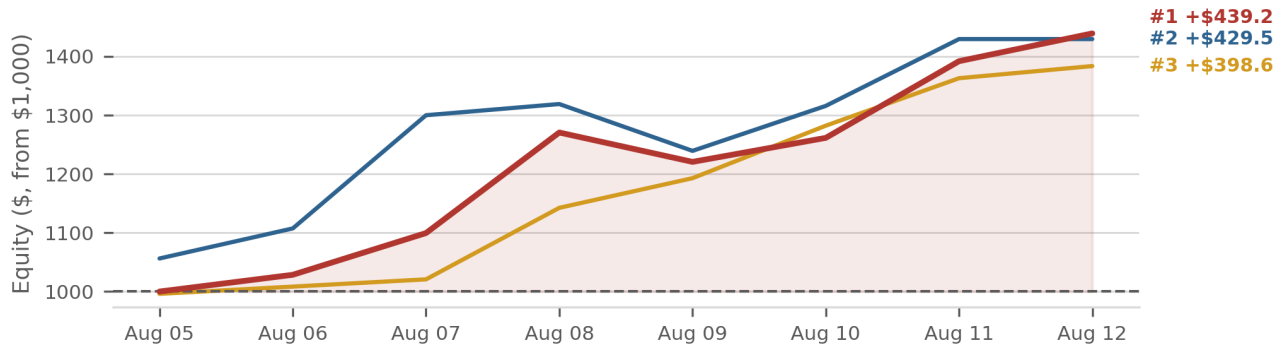
All three are waves-stage results -- best of 7,307 sims in the 7d branch; better than 99.0% / 99.0% / 98.4% of the 2,160-config stage-1b grid (#1/#2/#3; the waves stage searches beyond that grid). Grid median -\$99.4.

Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine).

OPEN IN SANDBOX marketmania.ai/s/cw1-7d1 marketmania.ai/s/cw1-7d2 marketmania.ai/s/cw1-7d3

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

TOP-3 equity -- 7d (Aug 5-12, 2026) -- \$1,000 at 10x -- 69 settled trades



Daily settled equity for the three councils above (\$1,000 start, 10x, frozen window Aug 5-12, 2026); ranks match the table (#1 red, #2 blue, #3 amber -- fixed rank colors across all issues). Sandbox-widget style; a curve past the window end is an open-at-cutoff trade settling at its horizon.

TOP-3 -- 30d window (Jul 13-Aug 12, 2026)

#	Council	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	fable-5, deepseek-v4-pro, opus-5, qwen-3.8-max	ETH+SOL	4h/1h	+\$868.4	+86.8%	9.4%	42	80%
#2	gemini-3.1-pro, gpt-5.6-sol, deepseek-v4-pro	SOL+XRP+ETH	1d/1d	+\$590.6	+59.1%	11.0%	24	83%
#3	fable-5, deepseek-v4-pro, opus-5, qwen-3.8-max	BTC+ETH+SOL	4h/1h	+\$580.5	+58.1%	12.8%	57	78%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	50:80, 100:20	on	TP +10 / SL +35	>=60 (each)	farthest	1/1	update	\$450
#2	40:80, 100:20	on	TP +30 / SL +20	>=60 (each)	nearest	2/0	update	\$300
#3	50:80, 100:20	on	TP +10 / SL +35	>=60 (each)	farthest	1/1	update	\$300

All three are waves-stage results -- best of 4,892 sims in the 30d branch; better than 99.9% / 99.6% / 99.6% of the 2,172-config stage-1b grid (#1/#2/#3; the waves stage searches beyond that grid). Grid median -\$221.0.

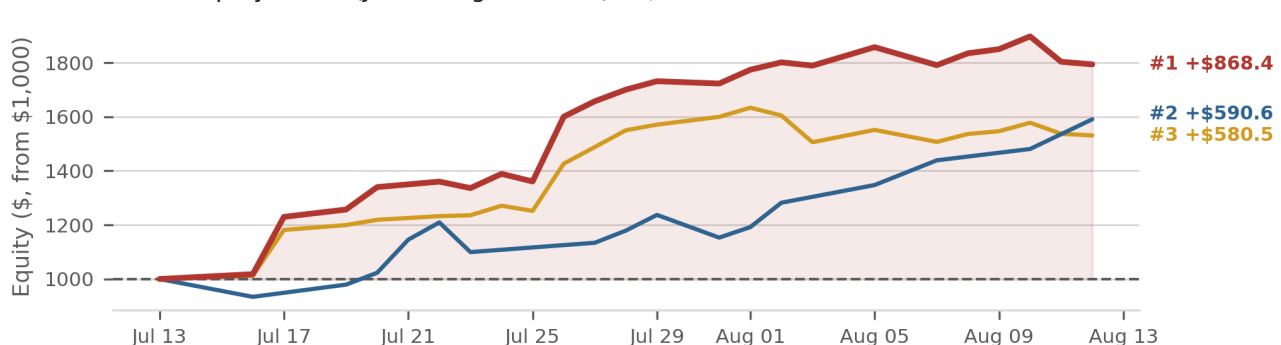
Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine).

#2: time_stop = 3x (only non-default extra on this config).

OPEN IN SANDBOX marketmania.ai/s/cw1-30d1 marketmania.ai/s/cw1-30d2 marketmania.ai/s/cw1-30d3

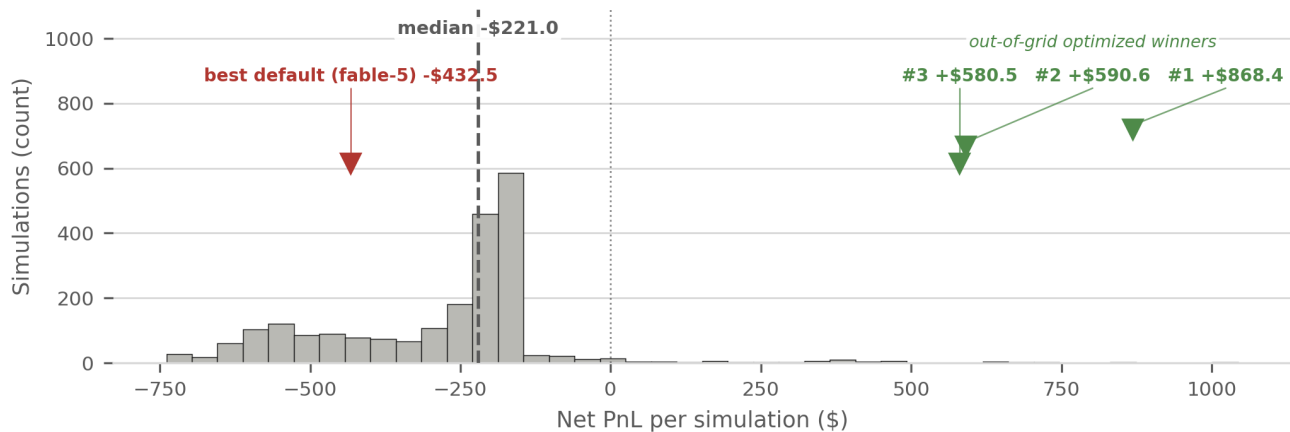
Each link opens this exact frozen window; results visible without sign-in (embedded share signature).

TOP-3 equity -- 30d (Jul 13-Aug 12, 2026) -- \$1,000 at 10x -- 120 settled trades



Daily settled equity for the three councils above (\$1,000 start, 10x, frozen window Jul 13-Aug 12, 2026); ranks match the table (#1 red, #2 blue, #3 amber -- fixed rank colors across all issues). Sandbox-widget style; a curve past the window end is an open-at-cutoff trade settling at its horizon.

Stage-1b systematic grid -- net PnL distribution -- 30d window (Jul 13-Aug 12, 2026) -- n=2,172



Net PnL across all 2,172 stage-1b grid sims, 30d window (Jul 13-Aug 12, 2026). Grid median -\$221.0, only 2.4% positive -- a materially harder window than 7d.

Red marker = best solo default; green = TOP-3.

Cross-window robustness check

Each TOP-3 config, replayed on the *other* window's history with the exact same config values (no re-tuning). This is a same-history overlap check, **not** an out-of-sample (OOS) test -- both histories are in-sample and the 30d window contains the 7d window.

Config	Family	Home (net)	Other (net)	Trades	Verdict
7d#1	fable-5, deepseek-v4-pro, opus-5, gpt-5.6-sol	7d: +\$439.2	30d: +\$178.8	108	Held up
7d#2	fable-5, qwen-3.8-max, gemini-3.1-pro, opus-5	7d: +\$429.5	30d: -\$551.3	16	Collapsed
7d#3	gpt-5.6-sol, opus-5	7d: +\$398.6	30d: -\$76.3	79	Collapsed
30d#1	fable-5, deepseek-v4-pro, opus-5, qwen-3.8-max	30d: +\$868.4	7d: -\$21.8	13	Collapsed
30d#2	gemini-3.1-pro, gpt-5.6-sol, deepseek-v4-pro	30d: +\$590.6	7d: +\$223.1	5	Insufficient (n=5)
30d#3	fable-5, deepseek-v4-pro, opus-5, qwen-3.8-max	30d: +\$580.5	7d: -\$11.5	15	Collapsed

Honest read: most tuned configs are fragile across windows -- 7d#2 and 7d#3 collapsed on 30d; 30d#1 and 30d#3 went mildly negative on 7d. One family did not collapse in either direction: the

fable-5+deepseek-v4-pro+opus-5 core, 4h/1h, ladder+BE, conf>=60, farthest source (7d#1 approx-equals the 30d#1/#3 family). 7d#1 held up on the 30d window (+\$178.8, 108 trades); the 30d#1/#3 family only mildly underperformed on the 7d window (-\$21.8, 13 trades). 30d#2 looked positive on 7d (+\$223.1) but on only 5 trades -- flagged insufficient, not evidence either way. Real OOS testing starts next week.

Secondary analysis -- solo-model top (dedicated search)

Sidebar to the main council-robustness narrative. The council search never contained single-model configs (councils were always ≥ 2 models), so the solo question got its own dedicated run: the same three-stage design (840-sim systematic grid + coordinate-ascent waves per branch), same frozen windows, same fixed economics, same publication gates (≥ 2 coins; ≥ 15 trades on 7d, ≥ 20 on 30d). One config per model; the ranks below are the best three *models*. From issue #2 the solo search runs alongside the council search every week.

Solo top-3 -- 7d window (Aug 5-12, 2026)

#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	gemini-3.1-pro	XRP+BNB	1h/1h	+\$417.8	+41.8%	14.9%	39	54%
#2	gpt-5.6-sol	XRP+BNB	4h/1h	+\$366.3	+36.6%	0.0%	30	87%
#3	qwen-3.8-max	XRP+BNB	4h/4h	+\$267.0	+26.7%	2.3%	30	57%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	100:100	off	TP +10 / SL +20	≥ 0 (each)	median	2/1	keep	\$450
#2	30:40, 100:60	on	TP +10 / SL +50	≥ 0 (each)	median	2/1	update	\$450
#3	100:100	off	TP +10 / SL -10	≥ 0 (each)	median	2/1	update	\$450

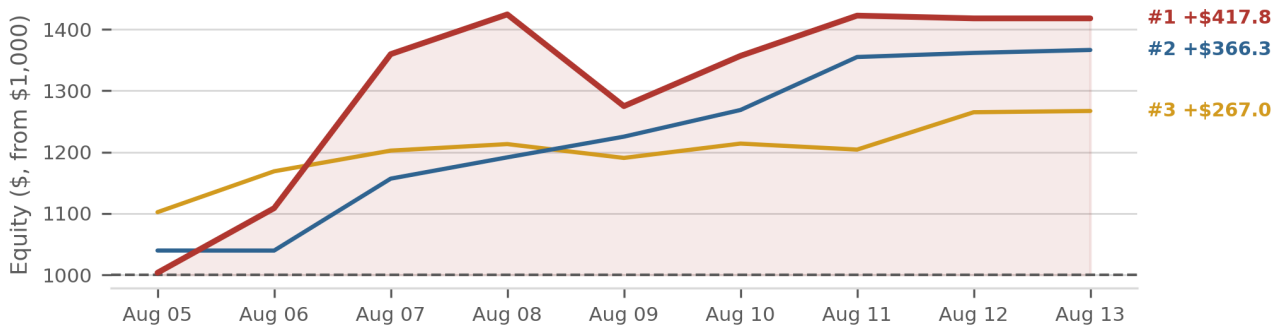
Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine).

#3: time_stop = 3x (only non-default extra on this config).

OPEN IN SANDBOX marketmania.ai/s/cw1-7ds1 marketmania.ai/s/cw1-7ds2 marketmania.ai/s/cw1-7ds3

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

Solo-model top equity -- 7d (Aug 5-12, 2026) -- \$1,000 at 10x -- 97 settled trades



Daily settled equity for the three solo models above (\$1,000 start, 10x, frozen window Aug 5-12, 2026); ranks match the table (#1 red, #2 blue, #3 amber -- fixed rank colors across all issues). Sandbox-widget style; a curve past the window end is an open-at-cutoff trade settling at its horizon.

Solo top-3 -- 30d window (Jul 13-Aug 12, 2026)

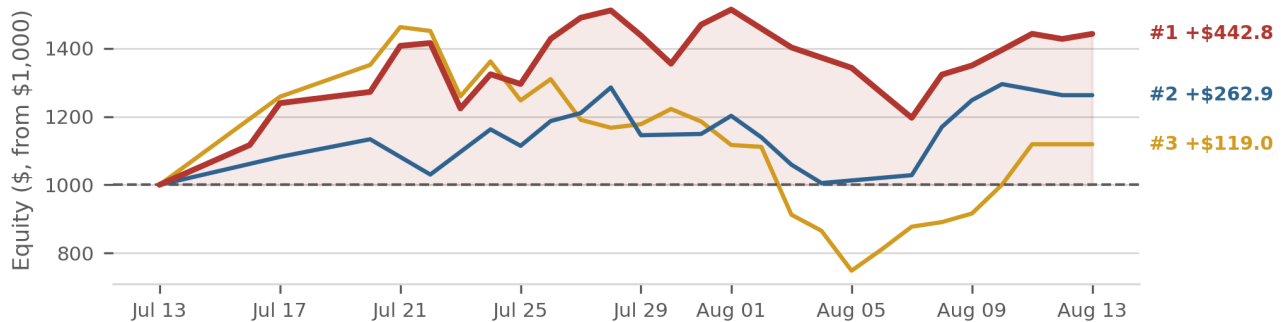
#	Model	Coins	FH/TF	Net	Net %	maxDD	Trades	WR
#1	opus-5	SOL+XRP	4h/1h	+\$442.8	+44.3%	31.7%	67	68%
#2	grok-4.5	SOL+XRP	4h/1h	+\$262.9	+26.3%	28.0%	47	70%
#3	deepseek-v4-pro	XRP+BNB	4h/1h	+\$119.0	+11.9%	71.4%	54	62%
#	Ladder	BE	Shifts	Conf filter	Source	Entry	Positional	Stake
#1	60:50, 100:50	on	TP -10 / SL +30	≥ 55 (each)	median	2/1	update	\$450
#2	50:50, 100:50	on	TP -10 / SL +0	≥ 60 (each)	median	2/1	update	\$450
#3	80:50, 100:50	on	TP +10 / SL +50	≥ 55 (each)	median	2/1	keep	\$450

Net % = net PnL as a percentage of the fixed \$1,000 starting deposit; every simulation runs at **10x leverage** (liquidation modeled by the engine).

OPEN IN SANDBOX marketmania.ai/s/cw1-30ds1 marketmania.ai/s/cw1-30ds2
marketmania.ai/s/cw1-30ds3

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

Solo-model top equity -- 30d (Jul 13-Aug 12, 2026) -- \$1,000 at 10x -- 164 settled trades



Daily settled equity for the three solo models above (\$1,000 start, 10x, frozen window Jul 13-Aug 12, 2026); ranks match the table (#1 red, #2 blue, #3 amber -- fixed rank colors across all issues). Sandbox-widget style; a curve past the window end is an open-at-cutoff trade settling at its horizon.

Council vs. solo: on 7d the best solo (+\$417.8) beats the council #3 but stays under the council #1 (+\$439.2); on 30d every solo finalist sits below the council TOP-3 -- the council keeps the #1 spot on both windows, but a well-tuned solo model is competitive with the council mid-ranks. Same in-sample caveats apply.

Baselines

Solo-model and council-of-7 defaults, plus three hand-tuned configs saved *before* this search began (independent baselines, not derived from the search).

7d (Aug 5-12, 2026)

Baseline	Net PnL	Trades	WR	maxDD	PF
Best solo default (gemini-3.1-pro)	-\$14.4	20	40%	2.2%	0.93
Worst solo default (grok-4.5)	-\$329.1	32	14%	33.4%	0.27
Council-of-7 default	-\$164.9	34	26%	17.3%	0.59
Hand-tuned config A (pre-search)	+\$137.3	7	80% (N=7)	0.0%	51.07 (N=7)
Hand-tuned config B (pre-search)	+\$295.9	27	84%	1.6%	2.84
Hand-tuned config C (pre-search)	-\$21.8	13	67%	10.3%	0.92

30d (Jul 13-Aug 12, 2026)

Baseline	Net PnL	Trades	WR	maxDD	PF
Best solo default (fable-5)	-\$432.5	99	30%	43.6%	0.58
Worst solo default (grok-4.5)	-\$770.1	65	14%	77.0%	0.26
Council-of-7 default	-\$693.5	63	16%	69.3%	0.26
Hand-tuned config A (pre-search)	+\$240.1	16	79%	9.3%	2.13
Hand-tuned config B (pre-search)	+\$112.5	92	71%	36.1%	1.10
Hand-tuned config C (pre-search)	+\$793.7	42	78%	10.3%	2.41

WR and PF greyed with (N=X) where fewer than 10 trades closed in the window -- insufficient sample; shown for completeness, not comparison.

Search validation: the TOP-3 headline beat the best hand-tuned pre-search config on both windows -- +\$439.2 vs +\$295.9 (7d); +\$868.4 vs +\$793.7 (30d).

Patterns: what the top decile looks like

Top decile vs. the full stage-1b pool (7d: 43 of 434; 30d: 62 of 623). Value shown per axis is the one most overrepresented in the top decile vs. the full pool (descriptive only, not a causal claim).

Axis	7d value	7d top-decile vs grid	30d value	30d top-decile vs grid
Archetype	ladder	58.1% (all 38.9%)	ladder	51.6% (all 38.0%)
Forecast horizon / timeframe	1d/4h	46.5% (all 7.8%)	4h/1h	98.4% (all 59.9%)
Entry (max-sideways / max-diff-side)	4/2	44.2% (all 40.3%)	0/0	38.7% (all 24.4%)
Universe size (coins considered)	5	81.4% (all 59.0%)	1	93.5% (all 20.9%)
Council size (models in ensemble)	2	46.5% (all 22.1%)	5	51.6% (all 30.3%)
Model membership	qwen-3.8-max	79.1% (all 50.7%)	qwen-3.8-max	62.9% (all 36.8%)

Winning knobs differ by window (waves top-20 per branch):

Knob	7d top-20 mode	30d top-20 mode
Ladder shape (steps)	100:100 (20/20)	50:40,100:60 (18/20)
Break-even stop (BE)	0 (19/20)	1 (20/20)
Min confidence	50 (18/20)	0 (8/20)
SL shift	-10 (18/20)	50 (16/20)
TP/SL source	nearest (15/20)	median (20/20)

7d top-20 favors a single rung (100:100), min-conf 50, SL shift -10, nearest source. 30d top-20 favors a 2-rung ladder, break-even on, SL shift +40..+50, median source. Honest observation against a one-size-fits-all config -- descriptive, not a recommendation to run either knob set live.

Solo-coin side note (excluded from TOP-3 by the ≥ 2 -coin rule)

- **7d:** fable-5+qwen-3.8-max+gemini-3.1-pro on XRP alone scored +\$618.4 (16 trades, WR 67%, maxDD 13.5%) -- higher than every published 7d TOP-3 config, excluded for failing the ≥ 2 -coin / pairwise-different-universe rule.
- **30d:** opus-5+fable-5+gemini-3.1-pro+qwen-3.8-max+grok-4.5 on SOL alone scored +\$1,225.3 (22 trades, WR 90%, maxDD 27.9%) -- the single highest score in the entire 30d search, excluded for the same reason.

Both flagged as **idiosyncratic-coin risk**: a single-coin edge (or lack of one) can be dominated by that one coin's path during the window -- a weaker generalization claim than a multi-coin, multi-universe result.

Liquidation note

Liquidation is modeled by the simulation engine at the fixed 10x leverage used throughout this search. Across the published configs, observed stop-loss placements stay well inside the distance that would approach the liquidation threshold. Max drawdown (maxDD) is reported on every config in this issue so realized risk is visible directly, rather than relying on this note alone.

Walk-forward

This is an **in-sample (IS)** issue: every number above comes from the same historical windows the search optimized against. The out-of-sample (OOS) fate of all six TOP-3 configs (three per window) becomes a **standing section starting issue #2**. Tracking a decay series over 5-10 weeks -- does performance hold, fade, or reverse -- is the core research question of this series: how robust (or overfit) are LLM-ensemble configs found by large-scale in-sample search?

Limitations

- **Multiple testing:** ~15,416 simulations were searched (council + dedicated solo); with a search this large, some winners are expected by chance alone. Antidotes in this issue: the full stage-1b histogram (not just winners), declared search axes, independent pre-search baselines, IS labeling throughout, and a walk-forward plan starting next issue. No formal multiple-testing correction is applied in issue #1 (see What we're testing next).
- **Single regime:** both windows sit inside one market regime; no regime-change evidence yet.
- **Small trade counts:** several cross-window reads run on as few as 5-16 trades -- directional, not conclusive.
- Research output, not financial advice.

Counters & lineage

Branch	Council scan	Skeleton grid	Waves	Total
7d (Aug 5-12, 2026)	600	2,160	4,547	7,307
30d (Jul 13-Aug 12, 2026)	600	2,172	2,120	4,892
Total (council search)				12,199
7d solo search (dedicated)	--	840	812	1,652
30d solo search (dedicated)	--	840	725	1,565

Solo-search sims are counted separately from the council-search total above.

Generated at: 2026-08-13T04:32:27Z · Methodology v1.1 + Config Watch amendments (owner-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 week's issue re-runs the full 3-stage pipeline fresh over that week's windows.

RESEARCH TO DATE

THIS REPORT -- council search: 12,199 sims · solo search: 3,217 sims · total search effort: 15,416 simulations

MARKETMANIA RESEARCH TO DATE (as of August 13, 2026, 06:57 UTC): 59,253 resolved forecasts since Jul 11 · 27,341 closed simulated trades · 9 models tracked · 5 assets · 5 horizons · hourly cadence · 5 published research reports

Resolved = horizon expired. Counting rule tightened in this issue; not comparable with the 14,151+ figure quoted in earlier issues.
Source: platform_counters_2026-08-13

NEW SERIES · Issue #1: Config Watch joins MarketMania's published research line.

What we're testing next

- Next issue: the OOS fate of all six configs above -- does the surviving family keep its edge?
- Council vs. solo OOS: does the ensemble advantage persist after tuning?
- Monthly: formal multiple-testing correction once several OOS windows accumulate.

Related research

- Stability Index Run 2 -- marketmania.ai/research/reports/si-run-2.pdf
- Weekly Calibration #1 -- marketmania.ai/research/reports/weekly-calibration-2026-08-03.pdf
- Weekly Model Watch #1 -- marketmania.ai/research/reports/model-watch-2026-08-03.pdf
- Consensus Watch #1 -- marketmania.ai/research/reports/consensus-watch-2026-08-03.pdf

CITE THIS REPORT

```
@techreport{mm_config_watch_2026w33,
  title      = {Config Watch #1},
  author     = {{MarketMania Research}},
  institution = {MarketMania},
  year       = {2026},
  month      = aug,
  day        = {13},
  type       = {Weekly Research Report},
  series     = {Config Watch},
  number     = {1},
  note       = {Methodology v1.1 + Config Watch amendments, hash e66c7e8c864a2233},
  url        = {https://marketmania.ai/research/reports/config-watch-2026-08-05.pdf}
}
```