Vasilis Kotopoulos

AI-native quant research · From strategy abundance to allocation conviction

NOVEMBER 2026

MANDATE ASSESSMENT · SAMPLE DELIVERABLE

Crypto prediction tournament: a daily coin-ranking signal

> Crowd-sourced prediction tournament · Crypto ranking signal

VERDICT

Requires further research

The ranking is accurate and holds up on the full coin universe, including delisted coins. Originality, which the tournament mainly pays for, is not yet shown.

PREPARED BY

Vasilis Kotopoulos

Case study against a publicly described mandate, not a past or current client. All results are simulated. Past and simulated performance is not indicative of future results. Not investment advice.

EXECUTIVE SUMMARY

Executive summary

VERDICT

Requires further research

The ranking is accurate and holds up on the full coin universe, including delisted coins. Originality, which the tournament mainly pays for, is not yet shown.

THE MANDATE

A crowd-sourced crypto tournament asks participants for a daily ranking of about 300 coins, scored against the following 20 days of returns. Its payout weights originality, meaning what a signal adds beyond the crowd's combined signal, about ten times more heavily than raw accuracy.

0.098
rank correlation with next-20-day returns (good signals: 0.01–0.03)

4.7
t-statistic over 275 weeks

178
eligible coins a week on average; 520 in total incl. delisted

0.016
added by the crowding ingredient on its own (not reliable)

Key findings

1. **Accurate ranking.** Over June 2020 – September 2025 the ranking's correlation with next-20-day returns averaged 0.098, against 0.01–0.03 for good tournament signals, with a t-statistic of 4.7 across 275 weekly snapshots.
2. **Not flattered by survivors.** Re-run on the exchange's full list of perpetual futures, including 25 coins delisted along the way, with each coin admitted only when it qualified on that day: 0.098, in line with 0.095 on the original 85-coin list. On the 300 most liquid coins each week: 0.099.
3. **Stable and free of look-ahead.** 0.094 in 2020–2023 and 0.107 in 2024 – September 2025. Using the signal one or two days late makes it slightly weaker (0.095 and 0.092), as expected when no future information has leaked in.
4. **Originality is the weak point.** The well-known 'calm coins do better' pattern scores 0.120 on its own. With it removed, the futures-funding ingredient adds only 0.016 (t-statistic 1.2), which is not reliable.
5. **Cost-sensitive as a traded book.** As a weekly long-short book, the Sharpe ratio is 1.05 on the 40 most-traded coins (0.83 at double costs) and 0.91 on all liquid coins (0.61 at double costs).

RECOMMENDED NEXT STEP

Submit the signal live to the tournament for **8–12 rounds with no money at stake**, and judge it on the originality score. In parallel, research a second, independent ingredient to replace the weak funding component. Stake only once originality is shown.

SECTION 1

Mandate and constraints

The requirements below are taken from the tournament's published rules and scoring documentation.

Requirement	What the tournament asks for	How we tested it
Universe	A daily ranking of about 300 coins	All perpetual futures on a major exchange, including delisted coins, with point-in-time eligibility
Target	Returns over the next ~20 days, starting two days after submission	Rank correlation of each weekly snapshot with returns from day 2 to day 22
Scoring: accuracy	Correlation with the target	Average rank correlation and its t-statistic, by period
Scoring: originality	Contribution beyond the crowd's combined signal, weighted about ten times more than accuracy	Not testable offline (needs the crowd's signal); proxied by what the signal adds beyond a well-known pattern
Robustness	Signals that keep working live	Signal-delay tests; split by period; liquidity-restricted universe

> **Pass rule for accuracy, fixed in advance:** average rank correlation above 0.01 with a t-statistic above 2. Originality cannot be tested without the tournament's live scoring, so it is the deciding open question.

SECTION 2

Approach

What we built

A transparent, rule-based score with no machine learning. It combines two ingredients:

- **Calmness.** Coins whose prices have been less volatile recently rank higher. This is a well-documented pattern across many markets.
- **Crowding in futures.** Coins where traders have paid little to hold leveraged long positions rank higher; heavily crowded coins rank lower.

The two ingredients are blended with fixed, equal weights. Rankings are produced at the daily close.

How we tested it

- **Full, survivorship-safe universe.** All 576 perpetual futures listed on the exchange over the period, including coins that were later delisted. 520 coins qualified at some point; on average 178 a week, up to 432.
- **Point-in-time eligibility.** A coin enters the ranking only if, on that day, it had enough price and funding history and met a minimum trading-volume threshold, using past data only.
- **The tournament's timing.** Each ranking is scored against returns from two days after the snapshot to 22 days after, as the tournament requires.
- **Sealed data.** Nothing dated 1 October 2025 or later was loaded. The test window ends on 30 September 2025, so this re-run has no held-back period.
- **Decomposition.** Each ingredient is scored on its own, and the crowding ingredient is also scored after removing everything it shares with calmness, to measure what it adds.
- **Delay and cost tests.** Scores with the signal used one and two days late; and a weekly long-short book on the ranking at base (0.20% per unit traded, plus actual funding), double and triple costs.

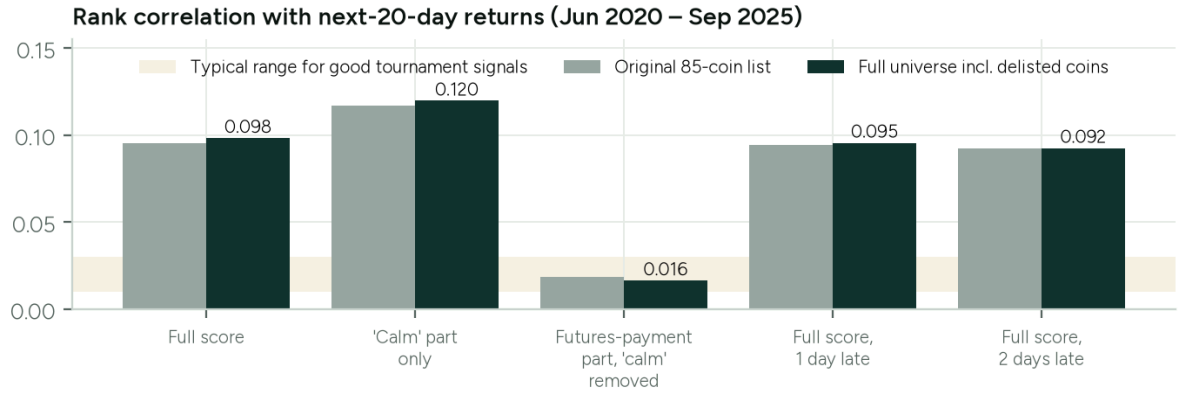
SECTION 3

Results

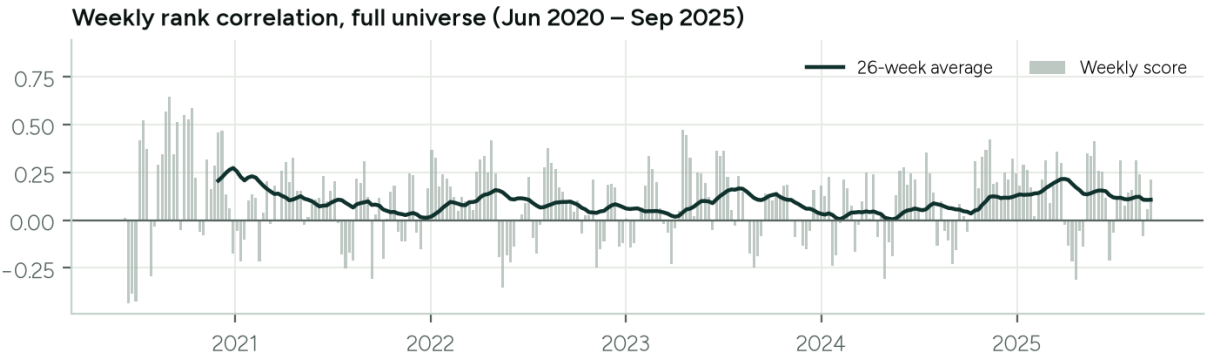
Accuracy scorecard (June 2020 – September 2025)

Measure	Full universe incl. delisted	Original 85-coin list
Full score	0.098	0.095
t-statistic	4.7	4.2
2020 – 2023	0.094	0.093
2024 – Sep 2025	0.107	0.100
300 most liquid coins each week	0.099	—
'Calm' ingredient alone	0.120	0.117
Crowding ingredient, 'calm' removed	0.016	0.019
Full score, 1 day late	0.095	0.094
Full score, 2 days late	0.092	0.092
Coins ranked per week (average)	161	72

Accuracy pass rule (above 0.01 with t-statistic above 2): **Met**. The 'coins ranked' row counts coins with a full 20-day result; eligibility counts in the text include all coins that qualified.



Rank correlation with next-20-day returns. Nearly all the strength comes from the calmness ingredient; what crowding adds on its own is small.



Weekly scores are noisy, as expected for a 20-day target: 69% of weeks were positive, and the 26-week average stayed above zero throughout (its low was 0.004, in May 2024).

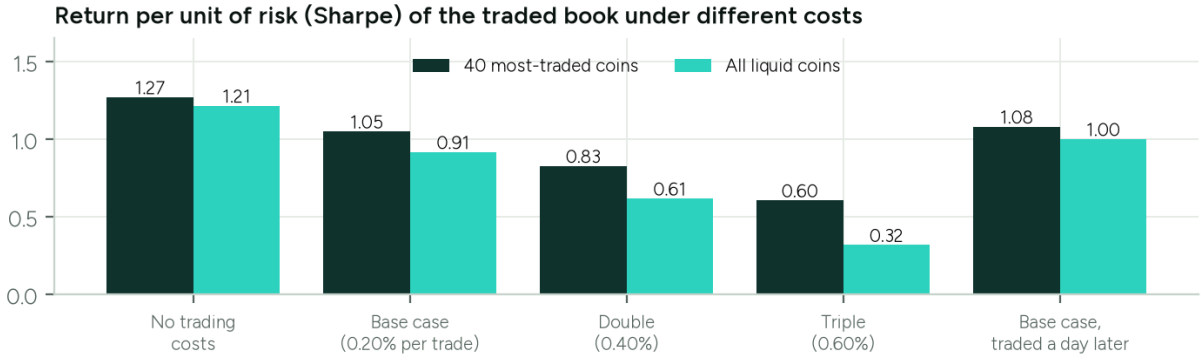
The ranking as a traded long-short book

Tournament payouts do not depend on trading the signal, but a book shows whether it is usable as a stand-alone strategy. Each week the book buys the top-ranked fifth of coins and sells the bottom fifth, risk-weighted and balanced against bitcoin.

Measure	40 most-traded coins	All liquid coins
Sharpe ratio, base costs	1.05	0.91
Annual return, base costs	25.0%	15.3%
Worst fall, base costs	-23.8%	-27.9%
Sharpe, double costs	0.83	0.61
Sharpe, triple costs	0.60	0.32
Sharpe, 2020 – 2023 / 2024 – Sep 2025	0.87 / 1.66	0.77 / 1.42
Correlation with bitcoin	0.03	0.06
Turnover (times a year)	27	26



Growth of 1 for the weekly long-short book, April 2020 – September 2025, after costs.



Sharpe ratio of the traded book as costs rise. The broader universe is far more sensitive to costs because smaller coins are more expensive to trade.

SECTION 4

Robustness and risk assessment

How much confidence the evidence supports

- **Accuracy is well supported.** A t-statistic of 4.7 over 275 weeks, similar results in both sub-periods, on both coin lists and on the 300 most liquid coins.
- **No sign of look-ahead.** Using the signal late makes it slightly weaker rather than stronger, which is what a clean signal should do.
- **Survivorship handled.** Including delisted coins and admitting coins only when they qualified on the day left the result essentially unchanged.
- **Originality is not supported.** The crowding ingredient's own contribution (0.016, t-statistic 1.2) is too weak to rely on. The rest of the score is a well-known pattern that other participants probably already use, so the tournament's originality score may be low.

Key risks

- **Payout risk.** Because the tournament pays mostly for originality, an accurate but unoriginal signal may earn little, or lose stake if staked.
- **No held-back period in this re-run.** To keep the data sealed, this re-run stops at September 2025. Live rounds are the out-of-sample test.
- **Costs, if traded.** On all liquid coins the book's Sharpe ratio falls from 0.91 to 0.61 at double costs and 0.32 at triple. The 40 most-traded coins are sturdier (1.05 to 0.83).
- **Uneven book returns.** The traded book had a weaker 2020 – 2023 (Sharpe 0.87 on 40 coins) than 2024 – 2025 (1.66), and worst falls of -23.8% to -27.9%.
- **Universe gap.** The tournament ranks about 300 coins; our eligible universe averaged about 178 a week, so some tournament coins would need a default rank.

> **Bottom line.** The signal ranks coins well, and that result is robust. What it adds that the crowd does not already have is unproven, and that is what the tournament pays for.

SECTION 5

Verdict and rationale

Requires further research

The ranking is accurate and holds up on the full coin universe, including delisted coins. Originality, which the tournament mainly pays for, is not yet shown.

Why further research is required

- The accuracy pass rule is met comfortably and the result survives every robustness test we ran.
- But the tournament's payout depends mainly on originality, which cannot be measured offline, and our best proxy for it, what the crowding ingredient adds, is weak.
- This re-run has no held-back period, so live rounds are needed to confirm accuracy out of sample.

What would move it to a pass

- A positive originality score over 8–12 live rounds.
- Or a second, independent ingredient that adds measurable accuracy beyond calmness, tested under the same rules.

SECTION 6

Recommended next steps

1. **Unstaked live rounds.** Submit the ranking live for 8–12 rounds with no money at stake and judge it on the tournament's originality score as well as accuracy.

2. **Replace the weak ingredient.** Research a second, independent ingredient to replace the crowding component, under the same locked-rules process.

3. **Stake decision.** Stake only if originality is positive over the live rounds; otherwise stop, or consider the 40-coin traded book as a separate, cost-tested strategy.

APPENDIX

Definitions in plain English

Term	What it means
Rank correlation (IC)	How well the ranking lines up with what happened next, from −1 to +1. 0 means no link. For 20-day crypto returns, 0.01–0.03 is already considered a good signal.
t-statistic	How many standard errors the average score is above zero. Above 2 is usually taken as unlikely to be chance.
Originality	What a signal adds beyond the crowd's combined signal. The tournament measures this live; it cannot be computed without the crowd's submissions.
Calmness (low volatility)	How little a coin's price has swung recently. Calmer coins have tended to do better, a pattern seen in many markets.
Funding fee (crowding)	A recurring payment between long and short holders of perpetual futures. A high fee paid by longs signals crowded bullish positioning.
Delisted coins	Coins later removed from the exchange. Including them avoids flattering results by testing only on survivors.
Point-in-time eligibility	A coin is included only if it qualified using information available on that day.
Signal delay test	Re-scoring with the signal used one or two days late. A clean signal should get slightly weaker, not stronger.
Sharpe ratio (return per unit of risk)	Average return divided by volatility, per year. Above 1 is generally considered good for a single strategy. Unless stated otherwise, we show it on total return.
Worst fall (maximum drawdown)	The largest drop from a previous high point to a subsequent low, before a new high is reached.
Turnover	How much of the book is traded per year.
Double and triple costs	The traded book re-run with trading costs at two and three times the base case of 0.20% per unit traded.

Important information

This document is a sample deliverable. It was prepared against a mandate described in public sources; the organisation that published that mandate is not a past or current client and did not commission, review or endorse this work.

All performance figures are simulated (backtested) and hypothetical. They were produced by applying fixed rules to historical market data, after modelled trading costs, and do not reflect actual trading or real money. Simulated results have inherent limitations: they are prepared with the benefit of hindsight, cannot fully account for liquidity, market impact, financing costs, operational constraints or the effect of real trading decisions, and may differ materially from results actually achieved. The held-back test period is also simulated, not live trading.

Past and simulated performance is not indicative of future results. Any investment can lose value, and leverage magnifies losses as well as gains.

This document is provided for information and discussion only. It is not investment advice, a recommendation, or an offer or solicitation to buy or sell any security, fund or strategy, and it does not take into account the objectives, financial situation or needs of any investor. Recipients should make their own independent assessment and seek professional advice before making any investment decision.