Vasilis Kotopoulos

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

NOVEMBER 2026

MANDATE ASSESSMENT · SAMPLE DELIVERABLE

Qualifying for a trader-seeding platform's top allocation tier

> Trader-seeding platform · Top-tier ('Gold') allocation

VERDICT

Requires further research

Clears the qualification bar in the held-back test year, but only about 6 in 10 historical one-year windows qualified, and it needs up to about 3× leverage to reach the platform's risk level.

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

Clears the qualification bar in the held-back test year, but only about 6 in 10 historical one-year windows qualified, and it needs up to about 3× leverage to reach the platform's risk level.

THE MANDATE

A trader-seeding platform allocates €50k–€500k to strategies that reach its top tier. To qualify, a strategy needs 8+ months of signal history on the platform and, for example, a one-year return above 20% with return divided by worst fall above 2.5. The platform runs every strategy at a fixed risk level (up to 6.5% monthly value-at-risk) and ranks allocations on monthly return.

+24.6%	4.2	−5.9%	≈60%
return in the held-back test year	return ÷ worst fall, test year (bar: 2.5)	worst fall in the test year	of past one-year windows qualified

Key findings

1. **Qualifies in the held-back test year.** October 2025 – October 2026: +24.6% with a worst fall of −5.9%, a ratio of 4.2. With double costs and late trades: +23.0% and 3.5. Both clear the bar of >20% and >2.5.
2. **Strong long-run record at the platform's risk level.** About 24.8% a year over March 2007 – September 2025 at 15.6% volatility (Sharpe 1.49), ahead of cash in 18 of 19 calendar years, with a worst fall of −14.5%.
3. **Qualification is likely, not assured.** Only about 60% of rolling one-year windows in the design period (126 of 211) met both rules, and no window ending between mid-2020 and early 2023 qualified.
4. **Leverage is the binding constraint.** Reaching the platform's risk level needs leverage of up to about 3×. A version capped at 2×, tested as a pre-declared fix, earned +16.3% in the test year and did not qualify.
5. **Near-zero stock sensitivity** (beta −0.08 in the design period), which helps with the platform's screen for strategies that simply track the market.

RECOMMENDED NEXT STEP

Further research before committing to the platform: confirm futures access, available leverage and financing costs on the platform, re-run the qualification test with those costs, and look for a way to qualify more consistently without more than 2× leverage. If those checks are satisfactory, a **fixed-fee one-month paper trial** at the platform's risk level, reported against the qualification rules.

SECTION 1

Mandate and constraints

The requirements below are taken from the platform's published documentation for its top allocation tier.

Requirement	What the platform asks for	How we tested it
Signal history	8+ months on the platform	Not testable in simulation; flagged as a prerequisite
Return	For example, one-year return above 20%	One-year return in the held-back test year; share of past one-year windows
Return quality	Return ÷ worst fall above 2.5	Same, measured on the same one-year windows
Risk level	Every strategy run at a fixed risk level, up to 6.5% monthly value-at-risk	Sleeve scaled to about 13.7% annual volatility, the equivalent level
Allocation	€50k–€500k, ranked on monthly return	Return shown at the platform's risk level
Diversification	Strategies that do not simply track the market	Stock-market beta and correlation reported
Instruments	What the platform offers	Seasonal commodity component needs futures access (to be confirmed)

> **Pass rule, fixed in advance:** in the held-back test year, a one-year return above 20% and return ÷ worst fall above 2.5 at the platform's risk level, holding under double costs and late trades.

SECTION 2

Approach

What we built

The diversified risk premia sleeve, scaled to the platform's fixed risk level:

- **A seasonal commodity strategy** as the main return engine, with no structural link to equities.
- **A defensive shelter switch** that holds long-dated government bonds and gold when the stock market is weak, and cash otherwise.
- **Risk scaling.** Exposure is adjusted day to day so the sleeve runs at the platform's fixed risk level, using leverage of up to about 3× when markets are calm.

The version was chosen from 23 pre-declared candidates on return relative to worst fall, using design-period data only. There is no machine learning and no discretionary input.

How we tested it

- **Rules locked in advance.** The candidate designs, the selection objective and the pass/fail rules were written down and time-stamped before any test was run.
- **Strict separation of design and test data.** All choices were made on March 2007 – September 2025 data, choosing among a small, pre-declared set of 23 versions using the mandate's own objective. October 2025 – October 2026 was held back and used once, only to check the final rules.
- **Realistic execution.** Signals are taken at the close and traded the next day; trading costs are charged on every position change; uninvested cash earns the T-bill rate.
- **Cost and timing stress.** Every result was re-run with double trading costs and with trades one day later, separately and together.
- **Luck adjustment.** Statistical confidence is adjusted for the roughly 2,000 ideas tested across our research to date, not just the versions tested for this mandate.
- **Statistical stress tests.** A block-bootstrap range for the Sharpe ratio, a random-reshuffle drawdown stress and a random sign-flip test on daily returns.
- **Benchmarks.** Results are compared with US stocks (S&P 500), cash, the mandate's own benchmark and a plain trend-following portfolio, to check the sleeve is not a repackaged version of something simpler.

SECTION 3

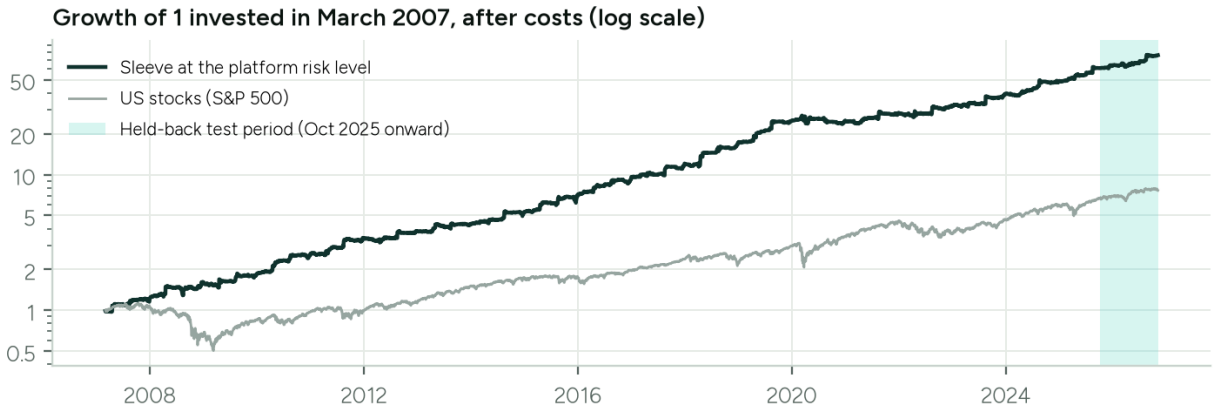
Results

Mandate scorecard

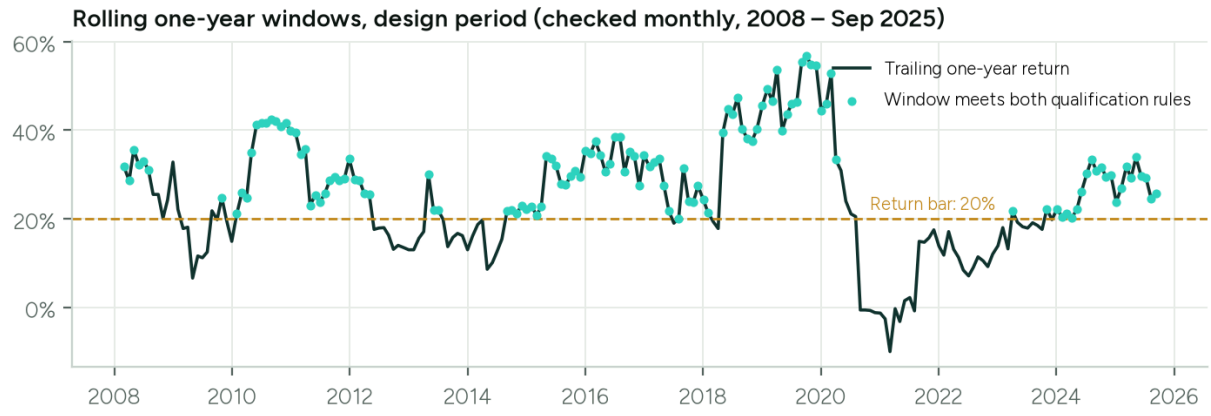
Requirement	Target	Result	Status
One-year return (test year)	above 20%	+24.6%; +23.0% with double costs and late trades	Met
Return ÷ worst fall (test year)	above 2.5	4.2; 3.5 with double costs and late trades	Met
Consistency of qualification	—	About 60% of past one-year windows	Partly
Reaching the risk level	within platform limits	Needs up to about 3× leverage; a 2× cap failed	Partly
8+ months of platform history	required	Not yet started	Not met

Design-period profile (March 2007 – September 2025)

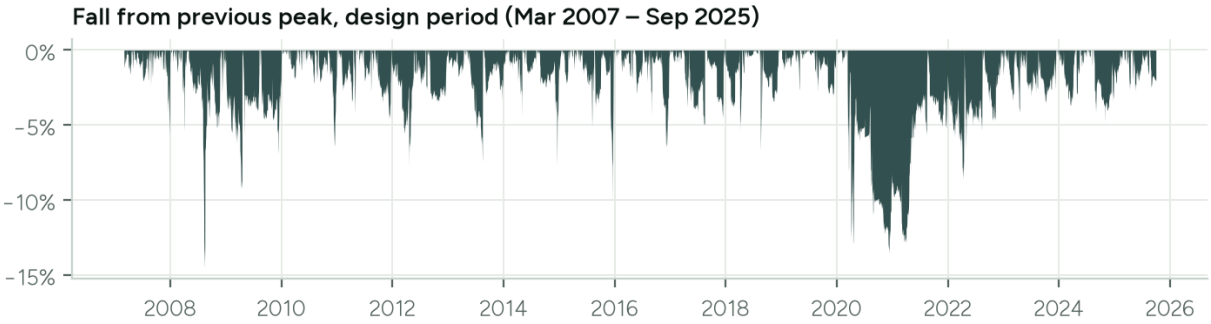
Measure	Base case	Double costs + late trades
Annual return	24.8%	24.0%
Volatility	15.6%	15.7%
Sharpe ratio	1.49	1.45
Sharpe above cash	1.40	1.36
Worst fall	-14.5%	-14.4%
Calendar years ahead of cash	18 of 19	18 of 19
Return-to-worst-fall ratio	1.71	1.66
Stock-market beta	-0.08	-0.08



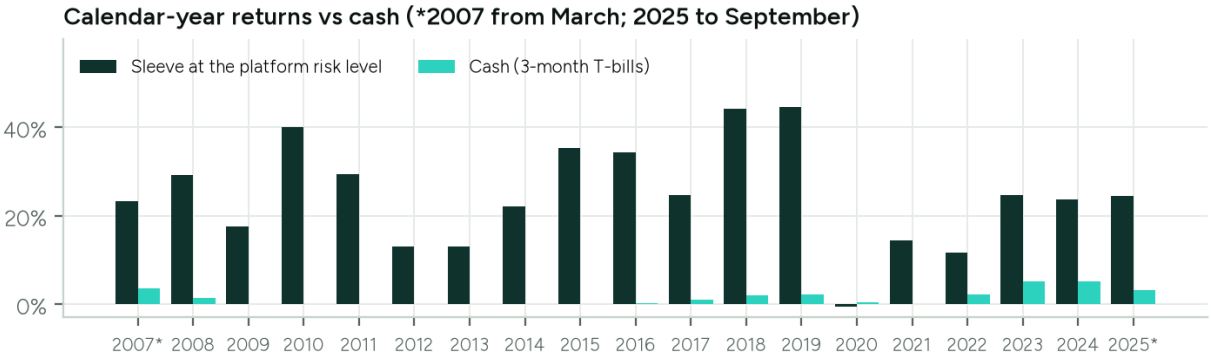
Growth of 1 invested in March 2007 at the platform's risk level, after costs. The shaded area is the held-back test period.



Trailing one-year return checked monthly. Dots mark windows that met both rules (return above 20% and return ÷ worst fall above 2.5): 126 of 211. No window ending between mid-2020 and early 2023 qualified.



Fall from previous peak at the platform's risk level. The deepest fall (−14.5%) was in 2008.



Calendar-year returns against cash. 2020 was the only losing year (−0.6%).

Held-back test year (October 2025 – October 2026)

Measure	Base case	Double costs + late trades	Version capped at 2× leverage
Return	+24.6%	+23.0%	+16.3%
Worst fall	−5.9%	—	—
Return ÷ worst fall	4.2	3.5	—
Qualifies?	Yes	Yes	No

Reported exactly as it came out; nothing was changed after seeing it. Only the figures above are reported for this period.

SECTION 4

Robustness and risk assessment

How much confidence the evidence supports

- **The selection itself looks sound.** The overfitting check across the 23 candidate versions came out at 14%: the version that looked best in one half of history usually stayed above average in the other half.
- **The record is unlikely to be noise on its own terms.** Block-bootstrap Sharpe range 1.23 to 1.75; a random sign-flip test was never matched in 1,000 trials.
- **Luck-adjusted confidence is about 50%** after allowing for the roughly 2,000 ideas we have tested. That is better than most results at this stage but well short of the 95% we would need to call it proven.
- **Cost and timing.** Double costs and late trades cost about 0.8% a year in the design period and 1.6% in the test year; the qualification result survives both.

Key risks

- **Qualification is uncertain in any given year.** About 40% of past one-year windows would not have qualified, including a run of more than two years from mid-2020. A qualifying test year is encouraging, not a guarantee.
- **Leverage and financing.** Reaching the platform's risk level needs up to about 3× leverage. The simulation scales returns to that level without charging a financing cost for the extra exposure; on the platform, financing or holding charges would reduce returns. The pre-declared 2× cap did not qualify, so leverage cannot simply be reduced.
- **Deeper falls are possible.** The worst historical fall at this risk level was −14.5%, and the random-reshuffle stress points to about −21.8% in an unlucky sequence. A fall of that size in a qualification year would push return ÷ worst fall well below 2.5.
- **Instrument access.** The seasonal commodity component needs futures access on the platform, which is not yet confirmed.
- **Time to allocation.** The platform requires 8+ months of history on its own systems before any allocation, so the earliest allocation is at least that far away.

> **Bottom line.** The numbers clear the bar in the test year, but the path to an allocation depends on platform-specific facts we have not yet confirmed (futures access, leverage, financing) and on a qualification rate of only about 6 in 10 years.

SECTION 5

Verdict and rationale

Requires further research

Clears the qualification bar in the held-back test year, but only about 6 in 10 historical one-year windows qualified, and it needs up to about 3× leverage to reach the platform's risk level.

Why further research is required

- The test-year result qualifies with room to spare (+24.6%, ratio 4.2) and survives double costs and late trades.
- But qualification rests on leverage of up to about 3×, the one pre-declared fix to reduce it (a 2× cap) failed, and financing costs at that leverage are not yet modelled.
- Historically only about 6 in 10 one-year windows qualified, so the chance of qualifying in a given year is meaningfully below certain.
- Futures access and the 8-month platform history are prerequisites that cannot be confirmed in simulation.

What would move it to a pass

- Confirmation that the platform offers the needed instruments and leverage, and a re-run showing the result holds after the platform's financing costs.
- A version that qualifies in a clearly higher share of years, or with lower leverage, tested under the same rules.

SECTION 6

Recommended next steps

1. **Platform fact-check.** Confirm futures (or equivalent) access for the commodity component, maximum leverage, and financing or holding costs. Re-run the qualification test with those costs included.

2. **Targeted research.** Look for an additional, independent return source that lifts the qualification rate or cuts the leverage needed, tested under the same locked-rules process.

3. **One-month paper trial (fixed fee).** If the checks pass, run the sleeve on paper at the platform's risk level, reported monthly against the qualification rules, as the first step toward the 8-month history requirement.

APPENDIX

Definitions in plain English

Term	What it means
Annual return	The average yearly growth rate over the period, compounded, after modelled trading costs.
Volatility	How much returns swing around, expressed per year. Higher means a bumpier ride.
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.
Sharpe above cash	The same ratio after first subtracting the return available on cash (3-month T-bills). It shows the reward for taking risk rather than holding cash, and is the stricter of the two.
Worst fall (maximum drawdown)	The largest drop from a previous high point to a subsequent low, before a new high is reached.
Return ÷ worst fall	The one-year return divided by the size of the worst fall within that year. The platform's bar is 2.5.
Value-at-risk (VaR)	A loss level that should be exceeded only rarely; here, the monthly loss expected to be exceeded about 1 month in 20. The platform caps it at 6.5%.
Leverage	Running more exposure than the capital in the account. Here, up to about 3× in calm markets.
Stock-market sensitivity (beta)	How much the strategy tends to move when the stock market moves 1%. Zero means no systematic link; 1 means it moves with the market.
Design period	The historical window in which all choices were made. Nothing after 30 September 2025 was used to design or select anything.
Held-back test period	Data from 1 October 2025 onward, kept aside and used only once, to check the final, already-fixed rules. It is a simulation on unseen data, not live trading.
Double costs and late trades	A stress test that doubles every modelled trading cost and delays each trade by one extra day. A strategy that only works with perfect timing or very cheap trading usually fails this test.
Luck-adjusted confidence	Testing many ideas makes it likely that some look good by chance. This statistic (a "deflated Sharpe ratio") estimates the probability that the record reflects real skill, after allowing for the total number of ideas we have tested. We require 95% to call a result statistically proven.
Overfitting check	Estimates how often the version that looked best in one half of history ends up below average in the other half. Lower is better; below 50% means selection added value.
Bootstrap range	We resample blocks of the historical returns 1,000 times and recompute the Sharpe ratio. The range shown covers the middle 90% of outcomes.
Random-reshuffle stress	We shuffle the order of historical daily returns 1,000 times and record the worst fall each time. The figure shown is a bad-case (1-in-20) outcome, a guide to how deep a fall could be in an unlucky sequence.

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.