Vasilis Kotopoulos

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

NOVEMBER 2026

MANDATE ASSESSMENT · SAMPLE DELIVERABLE

Alternative risk premia sleeve with near-zero stock exposure

> Public pension plan · Crisis-risk-offset portfolio, alternative risk premia allocation

VERDICT

Pass

Meets every mandate rule set in advance. In the held-back test year it stayed ahead of the benchmark, but only by a thin margin.

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

Pass

Meets every mandate rule set in advance. In the held-back test year it stayed ahead of the benchmark, but only by a thin margin.

THE MANDATE

A US public pension plan holds 20% of its assets in a crisis-risk-offset portfolio. Within it, an alternative risk premia allocation (8% target) must have little or no stock-market exposure, hold up in crashes, and beat 90-day T-bills + 2% after costs, in accounts of roughly $200–300M.

7.6%	-4.9%	-0.02	19 of 19
a year after costs, Mar 2007 – Sep 2025	worst fall over the same period	sensitivity to US stocks (beta)	calendar years ahead of cash

Key findings

1. **Near-zero stock exposure.** Beta to the S&P 500 was −0.02 over 2007 – Sep 2025, against a mandate limit we set at 0.2. Sensitivity to long Treasuries (0.07), high-yield credit (0.00) and broad commodities (0.01) was also close to zero.
2. **It held up when stocks fell.** +10.9% in 2008, +0.2% in Q1 2020 and +4.2% in 2022, while the S&P 500 fell 36.8%, 19.4% and 18.2%. In the worst 10% of stock-market months it averaged +0.81%.
3. **Ahead of the benchmark.** 4.0% a year above T-bills + 2% over the design period, and 0.7% above it in the held-back test year (Oct 2025 – Oct 2026).
4. **Survives cost and timing stress.** With double costs and late trades, the design-period Sharpe ratio moves only from 1.51 to 1.42. In the test year the Sharpe above cash moves from 0.79 to 0.68.
5. **Main caveats.** After adjusting for the number of ideas we have tested, the evidence is not statistically conclusive; most of the return comes from one seasonal commodity strategy; and capacity at $200–300M per account is not yet confirmed.

RECOMMENDED NEXT STEP

A **fixed-fee, one-month paid trial**: we run the sleeve on paper every day against your exact policy rules and benchmark, report monthly in your format, and hand over full risk and attribution data. In parallel, we confirm capacity for the commodity component at your account size. No commitment beyond the trial.

SECTION 1

Mandate and constraints

The mandate is drawn from the plan's published investment policy and manager search documents. We translated each requirement into a test we could run before seeing any results.

Requirement	What the plan asks for	How we tested it
Role in the portfolio	Alternative risk premia allocation (8% target) inside a 20% crisis-risk-offset portfolio	Judged as a stand-alone sleeve and on its behaviour in stock-market sell-offs
Stock-market exposure	Low or none	Beta to the S&P 500 below 0.2
Crisis behaviour	Hold up in crashes such as early 2020 and 2022	Positive return in Q1 2020 and in calendar 2022
Benchmark	90-day T-bills + 2%	Annual return above T-bills + 2%, design and test periods
Return quality	Solid return after costs	Sharpe ratio above 0.5 after costs; Sharpe above cash also reported
Consistency	Reliable over time	Calendar years and rolling 12-month periods ahead of cash
Account size	About $200–300M per account, up to four managers	Flagged for confirmation; not modelled in the simulation
Instruments	Liquid, transparent	Liquid exchange-traded bond, gold, currency and commodity markets

> **Pass rule, fixed in advance:** stock-market beta below 0.2, positive in both Q1 2020 and 2022, and a Sharpe ratio above 0.5 in the held-back test period after costs, with the result holding under double costs and late trades.

SECTION 2

Approach

What we built

A multi-source sleeve designed to earn returns that do not depend on the stock market:

- **A seasonal commodity strategy** as the core return engine. It captures recurring calendar patterns in commodity futures that have no structural link to equities.
- **A defensive shelter switch.** When the stock market is in a weak trend, this part holds long-dated government bonds and gold; otherwise it holds cash. This is what protects the sleeve in sell-offs.
- **A small trend-following slice** across alternative markets (currencies, metals, energy, agriculture, credit and emerging markets), sized by risk, for further diversification.

The three parts are combined with fixed weights. There is no machine learning and no discretionary input; every position follows from published prices.

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 8 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

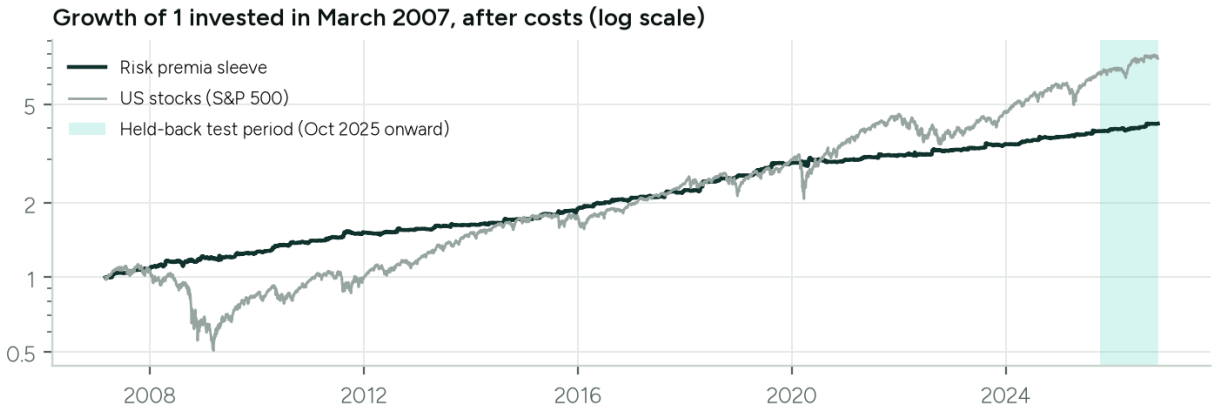
Results

Mandate scorecard

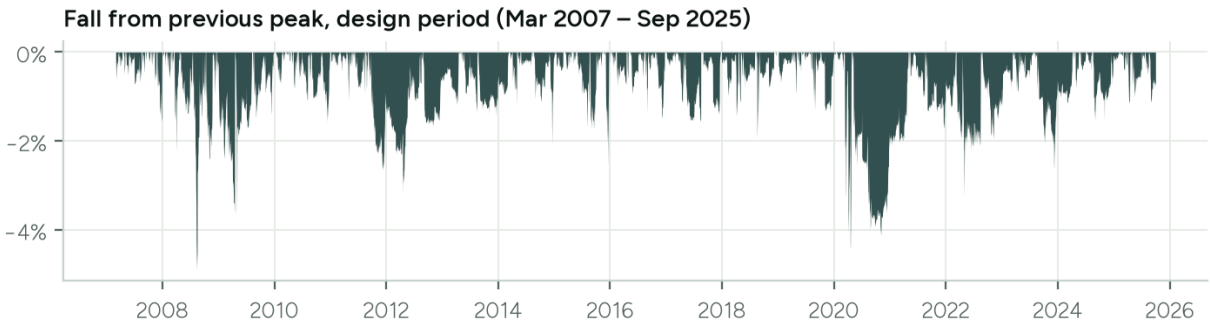
Requirement	Target	Result	Status
Stock-market beta	below 0.2	−0.02 (2007 – Sep 2025)	Met
Q1 2020	positive	+0.2% (S&P 500 −19.4%)	Met
Calendar 2022	positive	+4.2% (S&P 500 −18.2%)	Met
Return vs T-bills + 2%	ahead	+4.0% a year (design); +0.7% (test year)	Met
Sharpe ratio after costs	above 0.5	1.51 (design); Sharpe above cash 1.23 (design), 0.79 (test year)	Met
Holds under double costs and late trades	yes	Sharpe 1.42 (design); above cash 0.68 (test year)	Met
Calendar years ahead of cash	consistent	19 of 19	Met

Design-period profile (March 2007 – September 2025)

Measure	Base case	Double costs + late trades
Annual return	7.6%	7.1%
Volatility	4.9%	4.9%
Sharpe ratio	1.51	1.42
Sharpe above cash	1.23	1.13
Worst fall	−4.9%	−5.0%
Calendar years ahead of cash	19 of 19	18 of 19
Rolling 12-month periods ahead of cash	98.6%	94.3%
Return-to-worst-fall ratio	1.55	1.43



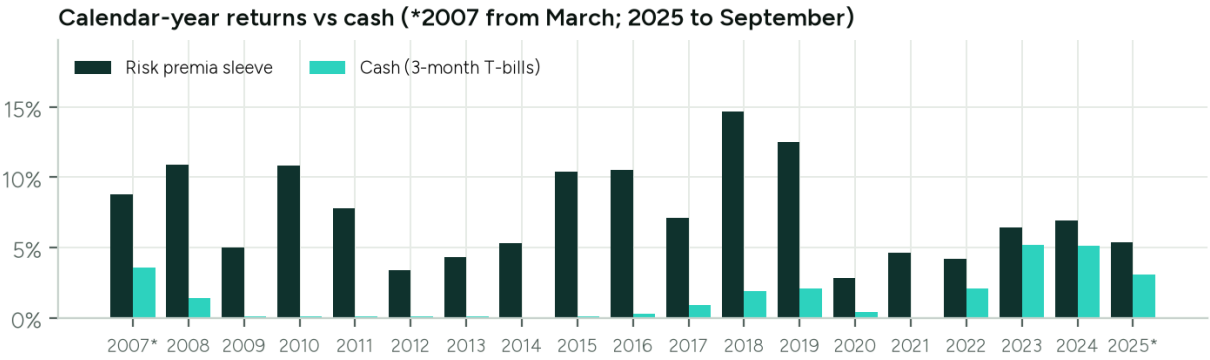
Growth of 1 invested in March 2007, after costs. The shaded area is the held-back test period, which was not used for any decision.



The worst fall in the design period (−4.9%) was in August 2008, before the most acute phase of the financial crisis; the sleeve finished 2008 up 10.9%.

Behaviour in stock-market stress

Episode	Sleeve	S&P 500
2008 (calendar year)	+10.9%	−36.8%
Q1 2020 (COVID crash)	+0.2%	−19.4%
2022 (rates and equity sell-off)	+4.2%	−18.2%
Worst 10% of stock-market months (average)	+0.81%	—



Calendar-year returns of the sleeve against cash. The sleeve was ahead of cash in every year, including 2008, 2020 and 2022.

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

Measure	Base case	Double costs + late trades
Return above T-bills + 2%	+0.7%	—
Sharpe above cash	0.79	0.68

The test year is 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 record is unlikely to be noise on its own terms.** The block-bootstrap range for the design-period Sharpe ratio is 1.24 to 1.79, and a random sign-flip test was never matched in 1,000 trials.
- **But we have tested many ideas.** Adjusting for the roughly 2,000 ideas screened across our research, the luck-adjusted confidence is about 19%, well short of the 95% we would need to call the result statistically proven. The pass therefore rests on how closely the sleeve fits the mandate, its economic logic and its consistency, not on statistical proof. A live paper record is the right way to build the missing evidence.
- **Small version-selection risk.** Only eight pre-declared versions were compared for this mandate, so the scope for picking a lucky version was limited.

Key risks

- **Concentration.** Most of the return comes from the seasonal commodity strategy (monthly correlation with it: 0.92). Diversification inside the sleeve is real but limited, and that strategy trades infrequently, so there are fewer independent bets than daily data suggest.
- **A thin margin in the test year.** The sleeve beat T-bills + 2% by only 0.7% in the test year, and its Sharpe above cash (0.79) was well below the design-period level (1.23). Higher cash rates raise the bar for this kind of low-volatility sleeve.
- **Deeper falls are possible.** The random-reshuffle stress suggests a fall of about −8.3% in an unlucky sequence, versus −4.9% in history. Risk budgets should allow for that.
- **Capacity.** We have not confirmed that the commodity component can absorb $200–300M per account without moving prices. This needs checking before any allocation.
- **Simulation limits.** All results are simulated. Market impact, financing and operational frictions are modelled simply.

> **Bottom line.** The sleeve behaves the way a crisis-risk-offset allocation should: close to zero stock exposure, positive in every major sell-off in the sample, and consistently ahead of cash. The open questions are statistical strength, concentration and capacity, not mandate fit.

SECTION 5

Verdict and rationale

Pass Meets every mandate rule set in advance. In the held-back test year it stayed ahead of the benchmark, but only by a thin margin.

Why it passes

- Every rule set in advance was met: beta −0.02 (limit 0.2), positive in Q1 2020 and 2022, and a test-year Sharpe comfortably above 0.5 after costs.
- The result holds under double costs and late trades, in both the design period and the test year.
- It did what the mandate exists for: it made money in 2008, early 2020 and 2022, when stocks fell sharply.
- It stayed ahead of the plan's T-bills + 2% benchmark in the held-back test year.

What keeps it from being a stronger pass

- Statistical confidence after adjusting for the number of ideas tested is low (about 19%).
- The test-year margin over the benchmark was thin (+0.7%).
- Capacity at the plan's account size is unconfirmed.

SECTION 6

Recommended next steps

1. **One-month paid trial (fixed fee).** We run the sleeve on paper every day against your exact policy rules and benchmark, send a monthly report in your format, and hand over full position, risk and attribution data. No commitment beyond the trial.

2. **Capacity review.** Confirm trading volumes and expected market impact for the commodity component at $200–300M per account, and agree an implementation plan.

3. **Decision point.** Review the paper record against the same pass rules. Extend the trial if you want a longer live record before allocating.

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.
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.
Correlation	A measure from −1 to +1 of how closely two return streams move together. Near zero means they move independently.
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.
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.
Sign-flip test	We randomly flip the sign of daily returns 1,000 times. If the real average return is rarely matched, the result is unlikely to be noise.
Rolling 12-month periods ahead of cash	The share of all 12-month windows, checked month by month, in which the sleeve returned more than cash.

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.