

Limits of Historical Trading Simulations

This study tests historical US-market trading rules in a backtest, isolating how observation delays, parameter selection, and assumed trading frictions alter simulated performance. For an investor, simulated profit depends on these choices, and the revised Kenneth French data do not represent point-in-time investable products.

SUBJECT

Methodological constraints in evaluating historical market timing signals.

EVIDENCE

Calculations rely on the Kenneth French US market factor dataset^[01]. This revised vintage supplies primary return observations rather than point-in-time investable products.

DOCUMENT TYPE

Private research

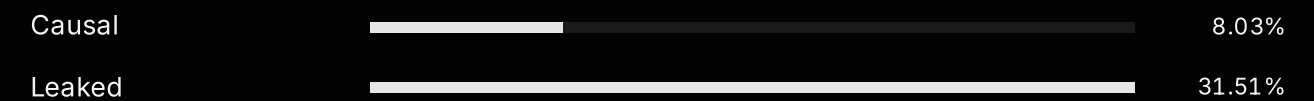
When a Backtest Sees the Month It Earns

This historical US-market trading-rule experiment compares a one-month, zero-threshold signal based on the previous completed month with one using the same month whose return it earns. That timing change adds 23.48494 percentage points to annualized backtest returns ^[01], showing how a signal that sees its earning month can make simulated performance look far stronger than a causal rule.

23.5

Annual Return Difference, Percentage Points

This gap reflects the absolute difference in compound annual returns between the leaked same-month signal and the causal previous-month baseline. It measures exactly 23.48494 percentage points per year.



These bars display compounded annual returns from January 1990 to December 2025. They exclude cumulative wealth. Both metrics derive from an identical fixed US market sample based entirely on our own calculations ^[01].

The fixed sample spans exactly 432 months. Our causal baseline restricts its inputs strictly to prior completed market data, yielding an 8.029 percent compound annual return. The leaked version shifts this observation window. It absorbs the same earning month.

Both rules process identical market factors. The leaked signal generates a 31.514 percent compound annual return across the identical continuous window from January 1990 to December 2025. The resulting gap isolates information availability rather than measured execution delays. Point-in-time datasets were not collected.

This is a hypothetical illustration. It is not a record of live trading. Historical revision means point-in-time availability cannot be certified, and we did not collect point-in-time datasets for this specific comparison.

Historical US-Market Timing Rule Lags the Kenneth French Benchmark

This historical US-market timing experiment selected a rule on training data, but the rule lagged the passive benchmark in both later windows. In the 2020–2025 audit, it earned a lower annualized return than constant market exposure despite a smaller maximum decline, leaving investors with a clear trade-off between return and drawdown ^[01].

-4.2 Annual compounded return difference, percentage points
Annualized return difference between the L9 rule (threshold -1%) and the Kenneth French market benchmark.

Selected rule annual return (%)		10.59%
Passive benchmark annual return (%)		14.82%

Annualized compounded returns for the 2020–2025 audit window. Author calculations based on US market factor data ^[01].

Over the 2020–2025 audit, the timing rule yielded 10.591% annually, trailing the 14.817% return of the Kenneth French passive benchmark. The strategy logged a -20.22% maximum decline against the -24.84% market drop ^[01].

These are not live tests. The retrospective simulations were frozen in 2026, and although the audit window did not dictate the configuration, the historical data remained visible. The selection score maximized annualized excess mean over standard deviation. Multiple testing in a training search over a family of configurations raises the significance threshold ^[03].

Analysis restricted to a US market factor illustration using the revised 202608 CRSP data vintage. This constitutes a retrospective simulation. We assert no point-in-time availability due to historical data revisions. Benchmark is not an investable product.

Testing a US Market-Timing Rule Against Revised History

This historical US market-timing experiment tests when an investor would hold market exposure, using 432 months of revised US market and risk-free returns through December 2025 across unequal training, validation, and audit windows. Revisions reshape older records, and lookbacks cross window boundaries, so the results cannot establish point-in-time data access or independent replication.

432

Total dataset observations, 1990-2025, months
The sample spans thirty-six calendar years before any splitting.



Window role	Dates	Month count
Training 199001-200912	1990–2009	240
Validation 201001-201912	2010–2019	120
Audit 202001-202512	2020–2025	72

Our calculations use the full frozen CSV, built from the revised 202608 CRSP vintage and starting in July 1926 ^[01]. It contains all 240 training months, January 1990 through December 2009. The L12 candidate takes January through December 1989 as its warmup; the selected L9 rule takes April through December 1989. Both spans come from the same file.

The selected L9 signal indexes the nine prior completed months, including across validation and audit boundaries. Some observations therefore inform a preceding window and a later window's lookback.

The 202608 vintage is not point-in-time evidence. A January 2025 transition places dividends on ex-dates without measuring revision bias ^[02]. The risk-free rate provider switched from Ibbotson to ICE BofA in June 2024, and the quantitative effect of this change was not measured ^[01].

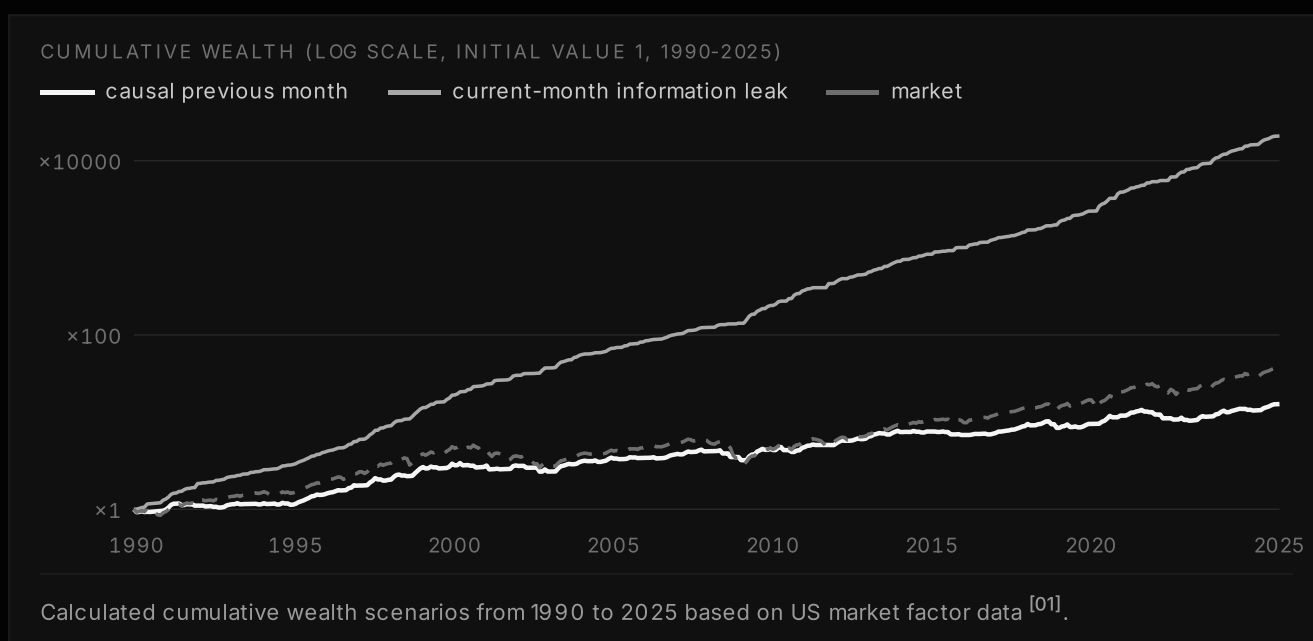
When a Market Rule Uses the Month's Return Too Soon

This historical US-market trading-rule experiment compares a signal using the current month's return with one using only the previous completed month, based on monthly factor returns from 1990 through 2025^[01]. The annual-return gap shows the apparent lift from unavailable information; the separate wealth paths are historical calculations, not proof of executable trading.

23.5

Leaked minus causal annual return (percentage points, 1990-2025)

Quantifies the annualized yield advantage of using current-month data. This rate differential differs mathematically from the plotted cumulative wealth.



Market excess return means the market return above the risk-free rate. The same-month rule uses that month's excess-return sign to set exposure for that same month, although the final sign is unavailable before the month closes. Using the previous completed month's sign removes this specific timing leak. In the 1990–2025 US factor-data calculation, the same-month rule's annualized return is exactly 23.48494 percentage points higher^[01].

The chart follows cumulative wealth from an initial value of 1 on a logarithmic scale, which makes proportional changes visible across a wide range of outcomes^[01]. The 23.48494-point figure compares annualized returns; it is not the difference between the chart's ending wealth values.

This research relies on a revised 202608 data vintage. Point-in-time information availability cannot be guaranteed. Results illustrate hypothetical signal timing mechanics, not executable investment strategies.

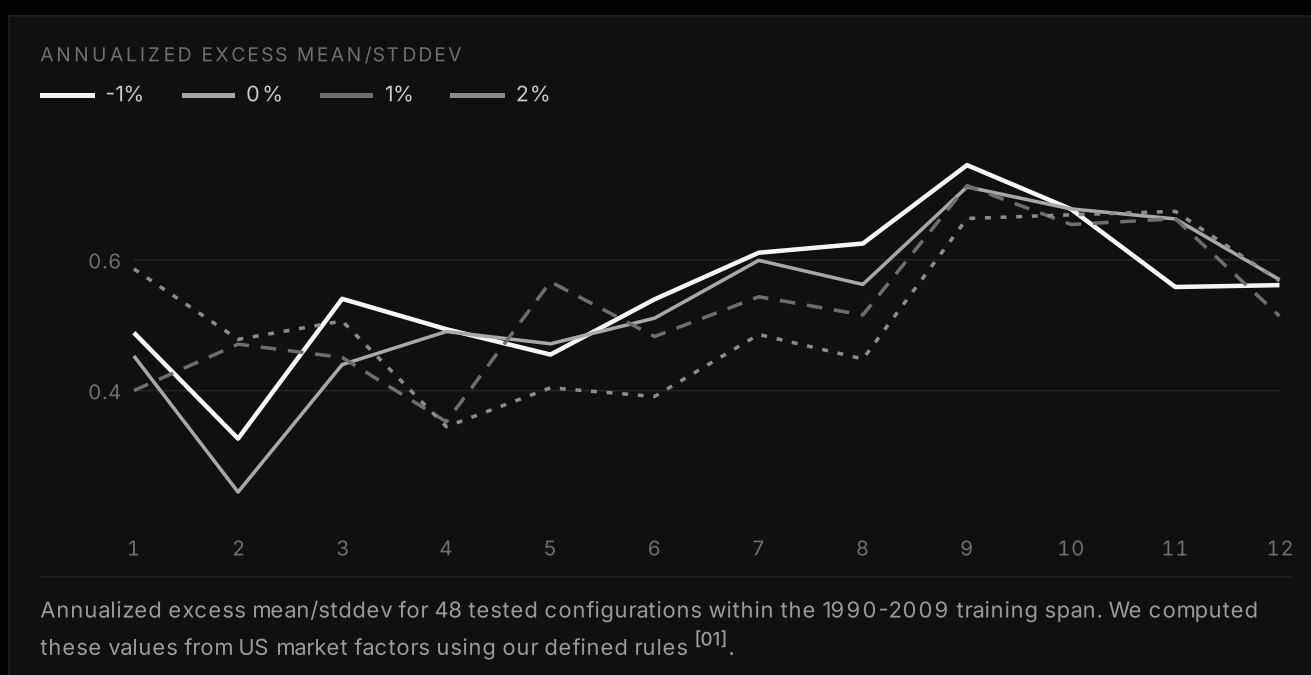
A Historical Search for US Market-Timing Rules

This historical US market-timing experiment compares 48 combinations of lookback periods and signal thresholds in the 1990-2009 training window, ranking each by annualized mean excess return relative to its standard deviation. That winner is a retrospective selection; the maximum score alone proves neither investment skill nor how the strategy would perform beyond the training period.

48

48 attempted rules in 1990-2009

Total count of parameter combinations evaluated exclusively on the training window.



We construct the trading signal as the product of one plus the monthly market excess return measured over the prior L fully completed months, minus one. An allocation weight of one applies if this signal sits strictly above the target threshold. Otherwise, the weight zeroes out. We calculate the monthly portfolio return as RF plus the weight multiplied by $Mkt-RF$. RF defines the risk-free monthly return. $Mkt-RF$ denotes the market excess ^[01].

We test 48 variants using L parameters from 1 to 12 and thresholds of -1%, 0%, 1%, and 2%. Training happens only on 1990-2009 data. We compute the score as the square root of 12 multiplied by the mean monthly excess return, divided by the standard deviation of the monthly excess return. The $L9$ parameter with a -1% threshold won. Finding a maximum proves neither skill nor overfitting ^[04]. We selected no dates using later outcomes in the code. We claim no human-blind research process.

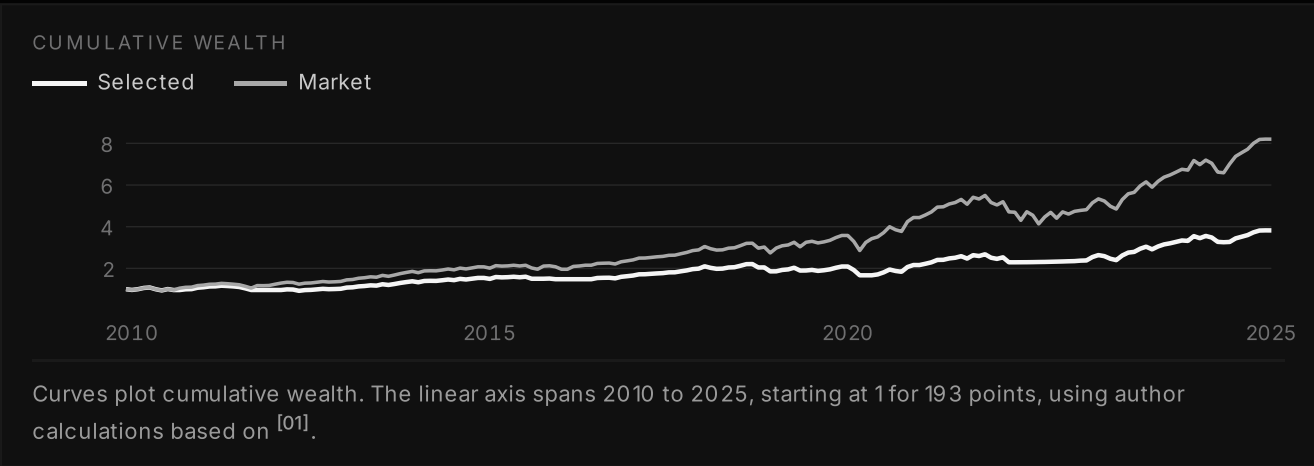
We evaluate a retrospective selection family. Our team knew late market history when writing the code. We did not estimate the probability of backtest overfitting. We did not collect independent out-of-sample data to verify this result. The specified training block controls these metrics entirely.

A Frozen US Timing Rule Trails the Market in Validation

A retrospective 2026 US-market timing experiment freezes the L9 rule at a -1% threshold and compares monthly compounded returns along one historical path with a fully invested market benchmark. Across adjacent later windows, the validation shortfall tests whether the training-era selection held up, while unequal exposure leaves its cause unresolved in this non-risk-matched comparison.

-6.0

Selected minus market annual compounded return in 2010-2019, percentage points
Measures annual underperformance. It compares the fixed training winner against continuous market exposure during the ten-year validation window.



Role & Dates	Months	Selected Ann. %	Market Ann. %	Gap pp/year	Selected Annualized Excess Mean/SD	Market Annualized Excess Mean/SD
Training 1990-2009	240	11.6	8.4	3.2	0.74	0.36
Validation 2010-2019	120	7.6	13.6	-6.0	0.68	1.01
Audit 2020-2025	72	10.6	14.8	-4.2	0.59	0.72

The weights were binary. Across the 1990-2009 training, 2010-2019 validation, and 2020-2025 audit windows, mean market weight was 72.08%, 86.67%, and 79.17%, respectively, while the strategy held cash in 67, 16, and 15 months, respectively ^[01].

Those percentages are time averages, not continuously fractional allocations. Exposure differed across the three windows, but that alone does not identify why the strategy lagged the market in validation.

We tested one adjacent path retrospectively in October 2026. We did not build risk-matched portfolios. We did not collect data to verify universal degradation. Calculations rely on primary observations from 1990 to 2025 ^[01]. These capture pure market factors, not investable products.

Testing Trading Costs in a Historical US Market-Timing Rule

This historical US-market timing-rule test applies five assumed one-way trading-cost levels to the strategy’s 2020 to 2025 evaluation and records eight absolute changes in market weight across the window. For investors, the comparison shows how execution assumptions narrow the rule’s reported annual return, without estimating any broker’s actual spread.

0.72

Gross minus 50-basis-point annual return 2020-2025, percentage points
Calculated difference between the frictionless scenario and the highest assumed friction.



Selected rule compound annual return in 2020-2025 under assumed one-way cost scenarios of 0, 5, 10, 25 and 50 basis points ^[01].

Changing market exposure demands liquidity. We applied fixed penalties to every position change during the 2020 to 2025 evaluation, deducting these zero to fifty basis-point scenarios directly from primary market data returns ^[01]. They remain assumptions. Measured bid-ask spreads vary across assets.

Compound returns drop. The frictionless baseline yields 10.59% in the final window, while applying the maximum fifty basis-point penalty reduces that result to 9.87%. We excluded taxation, market impact and financing constraints. Actual execution introduces latency.

These 0, 5, 10, 25 and 50 basis-point one-way cost levels are assumed scenarios, not measured bid-ask spreads. We deduct cost once per absolute market weight change, carrying previous period weight across the audit boundary without artificial liquidation fees. We model no taxes. Calculations rely on downloaded primary rows ^[01].

Evidence on a Historical US-Market Timing Rule

We tested a historical US-market timing rule that switches between market exposure and the risk-free rate, selected by its training-period excess-return-to-volatility score, against passive market exposure in a separate audit window. The estimated arithmetic monthly gap is negative, and its serial-dependence-adjusted confidence interval crosses zero, so the audit does not establish superior average returns.

1.00 Training selected Bonferroni-adjusted p-value

The unadjusted p-value of 0.207 is multiplied by the 48 training attempts and capped at a maximum of 1.00.

HAC lag months	mean paired gap pp/month	lower 95% pp/month	upper 95% pp/month	one - sided positive - mean p
0	-0.360	-1.074	0.354	0.839
3	-0.360	-0.936	0.216	0.890
6	-0.360	-0.915	0.195	0.898
12	-0.360	-0.880	0.159	0.913

Table displays Bartlett HAC lag sensitivity calculations for the audit paired mean difference.

Across 48 training configurations, the selected rule's annualized excess mean-to-standard-deviation ratio was its training selection score, distinct from its raw superiority p-value of 0.207427, while our Bonferroni procedure multiplies that p-value by 48 and caps it at 1.00^[01]. No p-value cleared 0.05. Multiple testing raises the significance hurdle^[03].

Both annual returns were positive. Across 2020-2025, the arithmetic paired mean (selected minus market) was -0.360139 percentage points per month, a relative gap, not an absolute monetary loss or negative strategy return, and it does not imply lagging every month. The one-sided test of $H_0: \text{mean} \leq 0$ against $H_1: \text{mean} > 0$ returned $p = 0.889906$. At HAC lag 3, the approximate asymptotic-normal two-sided 95% interval for the arithmetic paired mean was $[-0.935867, +0.215589]$ pp/month^[05].

Small-sample uncertainty and possible nonstationarity limit this asymptotic inference. Bartlett weights estimate long-run covariance for the standard error, not smooth returns. The table compares HAC lags 0, 3, 6, and 12. The approximate interval concerns the arithmetic monthly mean gap, not CAGR.

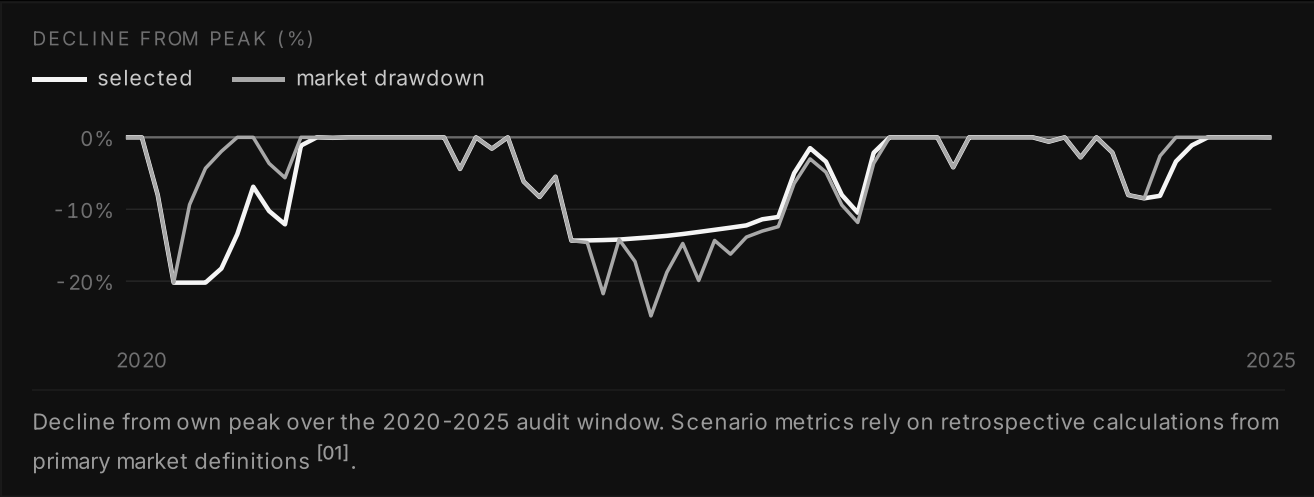
US Market Timing, Drawdowns, and the Limits of Historical Evidence

This historical US-market timing test compares a fixed monthly rule with the market from January 2020 through December 2025, with an October 2026 parameter freeze. Our frozen 202608 CSV ^[01] puts the selected rule's deepest decline at -20.216%, versus -24.845% for the market over 72 months. CRSP records the January 2025 CIZ switch and ex-date dividend reinvestment ^[02], while future capital protection remains unproven.

-20.2

Selected rule deepest decline from peak, 2020-2025 audit (percent)

The chosen configuration suffered a smaller maximum decline than the broader market. Absolute returns fell short of passive exposure. This historical outcome provides no guarantee of future safety.



signal clock	annual %	months
prior-completed-month cutoff	10.6	72
one extra month cutoff delay	11.4	72

The rule stayed fixed. With the unchanged L9 rule, -1% threshold and prior-completed-month signal, one extra full month of cutoff delay lifted annual return from 10.591% to 11.444% in this 72-month scenario covering January 2020 through December 2025 ^[01]. That signal multiplies (1 + Mkt-RF) across the prior nine completed months, so the stated rule allowed 2019 inputs in early-2020 calculations ^[01].

Configuration search can inflate apparent historical performance and raise overfitting risk, a general methodological concern supported by ^[04]. This experiment provides no estimate of overfitting. That risk alone does not invalidate every forecast. Live execution remains unmeasured. The audit uses simulated monthly returns, without actual fills or order-latency observations, so it cannot establish realized trading performance.

Register the rules in advance and hold them unchanged throughout future validation. Preserve point-in-time inputs and timestamps, and capture actual fills after registration.

Sources and Evidence Limits for the Historical Market-Timing Test

This sheet documents a historical US-market timing-rule experiment assessed against monthly market returns in a single, revised 202608 CRSP history, with no point-in-time evidence or independent replications. For investors, standard errors approximate sampling uncertainty; the cited papers provide no independent validation of returns or evidence of live execution.

5

addressed sources in the 1990-2025 analysis

Five distinct references supply the primary data and statistical methods.

[01] Kenneth French Data Library F-F Research Data Factors

This source supplies primary US market and risk-free monthly return definitions, representing raw observations rather than investable products.

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

[02] Kenneth French Data Library Documentation

CRSP's FIZ format ended after the December 2024 release; releases from January 2025 use CIZ, with dividends reinvested on ex-dates.

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[03] Harvey, Liu, and Zhu (2016)

This paper demonstrates that multiple testing requires a higher significance hurdle. We do not apply its specific factor model.

<https://www.nber.org/papers/w20592>

[04] Bailey, Borwein, Lopez de Prado, and Zhu

The authors show how configuration selection can overfit. Our report does not estimate PBO or use combinatorially symmetric cross-validation.

<https://www.davidhbailey.com/dhbpapers/backtest-prob.pdf>

[05] Newey and West (1987)

This work defines Bartlett-weighted autocovariances to adjust standard errors under serial dependence. Short samples and nonstationarity remain limitations.

<https://www.nber.org/papers/t0055>

This summary rests on a 432-month sample of US market returns from a single historical path. One underlying dataset constrains our evidentiary boundary. We claim no universal prevalence. Retrospective analysis does not guarantee future live trading execution.

Independent research, introductions clearly defined.

BlackRidge is an independent research bureau introducing private investors to quantitative traders through multiple strategy providers. We publish quantitative research. Investors access strategies through PAMM accounts at the broker. We are not a fund or broker and never hold client money.

01

Your funds stay at the broker. The account is opened in your own name. BlackRidge never receives, holds, or has withdrawal rights over client capital.

02

No fee on your profits. A one-time access payment of 6.7% of the agreed trading level, paid to the selected strategy provider. No management fee, no performance fee, no profit share, in any year.

03

You fund the risk, not the exposure. Allocations are notionally funded: you agree a trading level and fund the margin and drawdown allowance behind it. Trading losses can exceed the deposit without applicable negative balance protection; the separate 6.7% access payment is nonrefundable.

CONTINUE READING

blackridge.com/research: published research. [Notional funding](#) and [PAMM accounts](#) explained in full, and [answers](#) to the questions this raises.

We use AI models to gather and aggregate source material and to help prepare each report. Read it with its cited sources, sample, methods and limitations, and send corrections through [research methodology and corrections](#).

The client pays the selected strategy provider a one-time access payment of 6.7% of the agreed trading level. BlackRidge receives an introduction fee from that provider and does not receive broker compensation. [About BlackRidge](#).

Minimum funded capital \$25,000. The strategy, its live records, the broker, the selected strategy provider and the full commercial terms are presented on an introductory call and confirmed in writing before any payment or deposit.

This report is published for information only. It is not investment advice and does not take account of your circumstances. Trading leveraged instruments including CFDs carries substantial risk and is not suitable for all investors; you may lose the capital you fund and, depending on your broker's terms, may owe more than your deposit. Findings in this report may combine third-party evidence, historical calculations and illustrative scenarios; they are not verified live trading results of any strategy provider. Past performance does not indicate future results.