

Significance and Information Timing

QUANTITATIVE BACKTEST CONTROLS

This report examines how statistical inference changes when monthly backtests use standard errors adjusted for serial correlation, and how much a signal depends on when its inputs became observable. It compares an oracle series built with later-known information against a lagged version, then checks daily overlap, subperiod behavior, and the effect of trying multiple strategies. The evidence comes from historical market data and a release - timing example. The findings describe this test design, not general trading skill.

SUBJECT

Statistical significance and information timing in quantitative backtests

EVIDENCE

Historical market returns, release dates, and backtest sensitivity checks

DOCUMENT TYPE

Private research

A large t - statistic cannot establish timely information

The oracle's Newey - West t - statistic is enormous, yet statistical significance cannot certify that a signal was available in time to trade. It uses each month's realized market - excess - return sign inside that same interval; the previous - month control is far smaller despite an identical sample span. ^{[01][02]}

MONTHLY OBSERVATIONS 792 Complete monthly sample	ORACLE HAC T 25.73 Bartlett HAC, L=6	LAGGED HAC T 2.41 Bartlett HAC, L=6
---	---	--

Same-month sign × excess return $o_t = \text{sign}(r_t) \cdot r_t = r_t $		Previous-month sign × excess return $c_t = \text{sign}(r_{t-1}) \cdot r_t$		
CONSTRUCTION	MEAN EXCESS %/MONTH	HAC SE %/MONTH	DESCRIPTIVE MEAN/SD × √12	BARTLETT HAC T, L=6
Market excess return	0.59	0.163	0.46	3.60
Same-month sign × excess return	3.48	0.135	4.25	25.73
Previous-month sign × excess return	0.37	0.154	0.29	2.41

H_0 : $E[x_t] = 0$. All returns are monthly market excess returns. The oracle uses the same - month realized sign; the control uses the prior month's sign, with December 1959 supplying January 1960.

The oracle defines $x_t = \text{sign}(r_t)r_t = |r_t|$, making every observation nonnegative by construction. Its high t - statistic follows mechanically; it does not show that the zero - mean test is miscalibrated or measure the size of timing leakage.

For January 1960, the lagged series uses December 1959's sign. That removes the same - month outcome peek, but a previous - period value in today's revised research file is not proven available at the historical decision. The proxy is not an executable portfolio. ^[01]

INTERPRETATION	The lagged control's HAC t=2.41 is evidence against a zero mean under this specification; it is neither a direct Sharpe - ratio test nor evidence of market - adjusted skill. ^{[02][08]}
IMPLICATION	The ordinary annualized Sharpe ratios, 4.25 and 0.29, use the descriptive monthly mean/SD × √12 convention. They are not HAC - adjusted performance estimates. ^[03]
CAVEAT	The sample uses 792 monthly market excess returns from January 1960 through December 2025 in a revised research dataset. Costs, funding, borrowing, and liquidity are excluded; the lag fixes same - month timing only, not full point - in - time availability. ^[01]

HAC Lowers the t-Statistic for Overlapping Sums

Five-day sums of French daily market excess returns are strongly dependent because adjacent observations recur across windows ^[01]. Windows share four returns. The correction lowers the IID t-statistic while leaving the sample mean unchanged; interpretation still depends on information timing and what an unconditional mean test can establish ^{[02][08]}.

OVERLAPPING WINDOWS:
6,535

6535

Rolling five-day sums.

IID T-STATISTIC: 5.15

5.15

Unadjusted standard error.

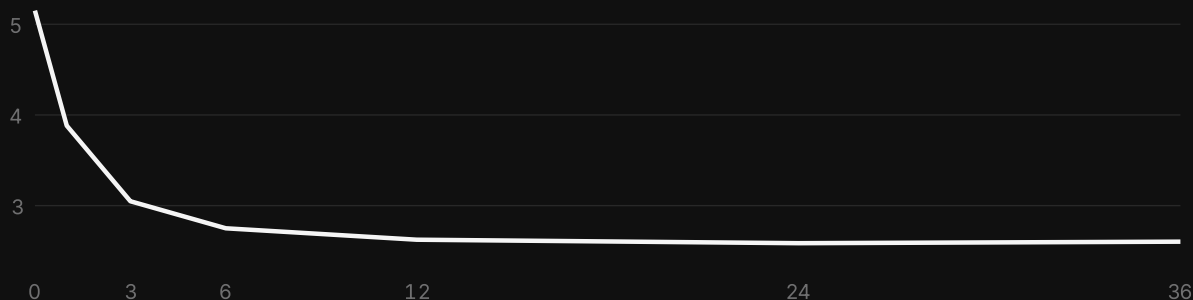
HAC T-STATISTIC, L=20:
2.58

2.58

Bartlett estimate at the
selected lag.

T-STATISTIC (DIMENSIONLESS)

— French daily market excess return, five-day sum



Daily sample: January 3, 2000 to December 31, 2025; source: Kenneth French Data Library ^[01].

HAC TRUNCATION LAG (DAILY OBSERVATIONS)

From January 3, 2000, through December 31, 2025, 6,539 daily observations generate 6,535 rolling five-day windows. Each neighboring sum shares four daily returns with the next, mechanically inducing serial dependence in this constructed series ^{[01][08]}.

The mean is 0.163 percentage points per five-day sum under both IID and HAC estimation. These values are arithmetic sums of daily excess returns, not compounded portfolio returns; changing the standard error does not alter that construction ^{[01][02]}.

INTERPRETATION

HAC leaves means unchanged. The t-statistic falls from 5.15 under IID assumptions to 2.88 at lag 4 and 2.58 at lag 20, while the estimated mean stays fixed ^{[02][08]}.

CAVEAT

Timing remains uncorrected. HAC adjusts sampling uncertainty around a positive unconditional mean; even a significant mean test does not establish market-adjusted skill or verify that information was available when returns were formed ^{[02][08]}.

Bandwidth shifts the two t - statistics in opposite directions

The direction is clear. Across the tested Bartlett bandwidths, the future -informed series keeps a large t-statistic but declines as the truncation window widens. The lagged series moves upward instead, a sensitivity result that does not make a longer window the better choice. ^{[02][08]}

BANDWIDTH SETTINGS

7

Bartlett truncation lags included in the sweep

ORACLE MINIMUM

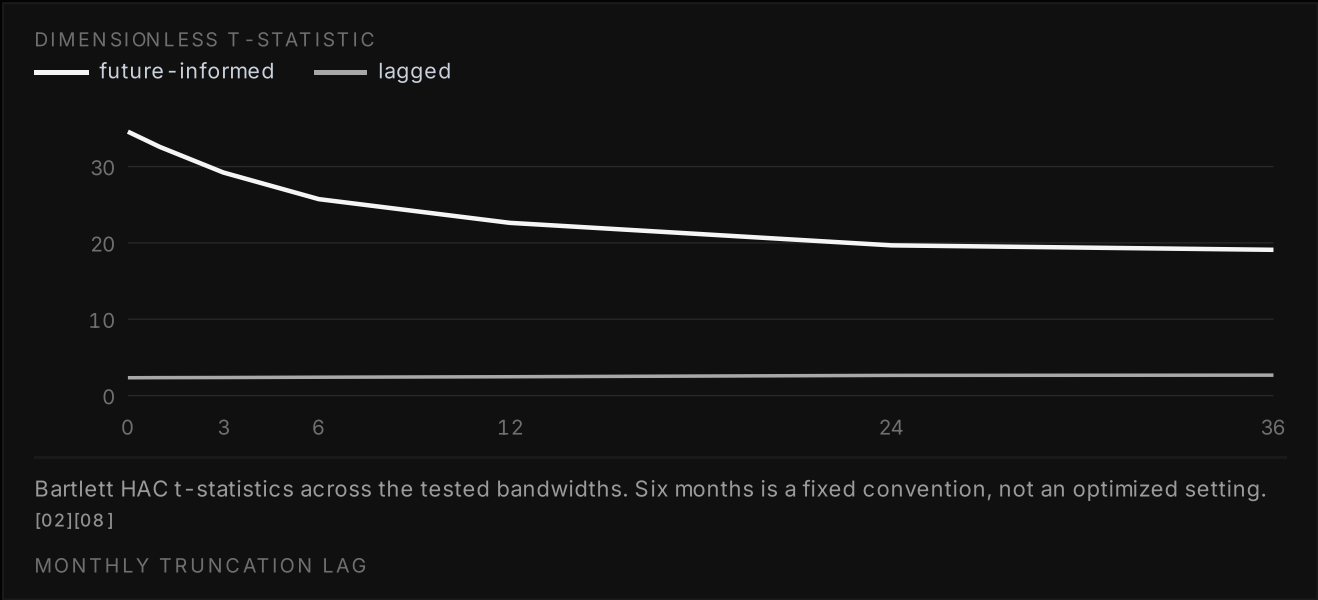
19.09

Lowest future -informed t - statistic in the sweep

LAGGED MAXIMUM

2.68

Highest lagged t -statistic in the sweep



Two paths emerge. Across monthly truncation lags of 0, 1, 3, 6, 12, 24 and 36, the future -informed t - statistic falls from 34.57 to 19.09, while the lagged estimate climbs from 2.34 to 2.68. ^{[01][02]}

The estimator matters. Bartlett HAC uses gamma estimates with denominator n, no prewhitening and no finite - sample adjustment (HC0). Its asymptotic basis requires dependence and stationarity conditions, as well as suitable bandwidth conditions. A sweep cannot prove them. ^{[02][08]}

IMPLICATION	Choose the six-month setting in advance and retain it as a reporting convention. This sensitivity sweep diagnoses dependence on bandwidth; it does not optimize inference. ^{[02][08]}
CAVEAT	These t -statistics test unconditional mean returns under an asymptotic HAC approximation. They do not establish market -adjusted skill or account for model selection; dependence, stationarity and bandwidth conditions remain assumptions, and this diagnostic sweep is no automatic significance rule. ^{[02][08]}

Serial covariance cuts the oracle ratio by 33.4%

Conventional annualization of monthly excess returns assumes zero serial covariance. On summed annual excess returns, the oracle's naive ratio is 4.25; the covariance -adjusted estimate is 2.83. Market and lagged constructions move little. The gap reflects serial dependence, not a significance test. ^{[01][03]}

ORACLE NAIVE	COVARIANCE - ADJUSTED	REDUCTION
4.25	2.83	33.4%
Conventional zero-covariance scaling	Stationary annual aggregation	Adjusted relative to naive ratio

$$SR_{12} = 12\mu / \sqrt{(12\gamma_0 + 2 \sum_{j=1}^{11} (12-j)\gamma_j)}$$

CONSTRUCTION	NAIVE ANNUAL RATIO	COVARIANCE - ADJUSTED RATIO
Market excess returns	0.46	0.45
Oracle excess returns	4.25	2.83
Lagged excess returns	0.29	0.30

Covariance-adjusted denominator for summed annual excess returns: $\text{sqrt}(12*\gamma_0 + 2*\text{sum}((12-j)*\gamma_{\text{maj}}, j=1..11))$; sample covariances use denominator n. ^[03]

The monthly series spans January 1960 through December 2025, with 792 observations across market, oracle, and lagged excess-return constructions. Annual ratios compare conventional scaling with a stationary aggregation denominator that incorporates monthly autocovariances through lag 11. ^{[01][03]}

The oracle's adjusted ratio is 33.4% below its naive value; market moves from 0.46 to 0.45, while lagged rises from 0.29 to 0.30. Using n rather than n-1 in covariance estimates changes little; serial covariance drives the comparison. ^[03]

IMPLICATION	Use the covariance-adjusted ratio for stationary annual aggregation, and show the naive figure beside it to preserve its zero-covariance assumption. Neither ratio establishes market-adjusted skill.
CAVEAT	These estimates sum monthly excess returns into a year under a stationary covariance structure; they are not compound annual Sharpe ratios, HAC t-statistics, or p-values. The oracle uses future information by construction, so its ratio is an unattainable reference. ^[03]

Q2 GDP arrived on two different dates

A quarter label identifies the period measured. Availability on a decision date depends on when the estimate was published and received. For Q2 2024, BEA's first annualized real GDP estimate arrived after quarter - end; a later release revised the figure for that same quarter. ^{[05][06]}

FIRST - RELEASE LAG	REVISION SIZE	RELEASE INTERVAL
25	+0.2	35
Quarter-end to advance estimate	Advance-to-second estimate change	Days separating publication dates

2024-06-30

Quarter-end

June 30, 2024

2024-07-25

Advance estimate

July 25, 08:30 EDT · 2.8% annualized ^[05]

2024-08-29

Second estimate

August 29, 08:30 EDT · 3.0% annualized ^[06]

RELEASE	DATE	GROWTH, ANNUALIZED (%)
Advance estimate	2024-07-25	2.8%
Second estimate	2024-08-29	3.0%

The quarter ended June 30, 2024. BEA published its advance estimate on July 25 at 08:30 EDT, 25 calendar days later, reporting 2.8% annualized real GDP growth. On August 29 at the same time, its second estimate reported 3.0% for that quarter. ^[05]
^[06]

Observed, the second release raised the estimate by 0.2 percentage points, 35 days after the advance release. Inference: a quarter - only key can expose a backtest to a value that was unavailable on a decision date; the key cannot show when a provider delivered it or when the system ingested it.

IMPLICATION	ALFRED preserves real-time periods and data vintages, which help identify what was published when. It does not prove when your own system received or ingested a value; store those timestamps separately. ^[07]
CAVEAT	These releases establish publication timing and a revision for one quarter. They do not establish that GDP revisions explain strategy profit, and they do not represent a claim about the latest GDP series today. ^{[05][06]}

Repeated tests raise the chance of a false positive

Repeated testing raises the family-level chance of at least one false positive. Here, the curve assumes independent true-null tests, each with exact 5% size. It is a defined mathematical scenario. It says nothing about how often investment strategies are tested.

TEST SIZE

5%

Assumed exact size for each true-null test.

AT 20 TRIALS

64.2%

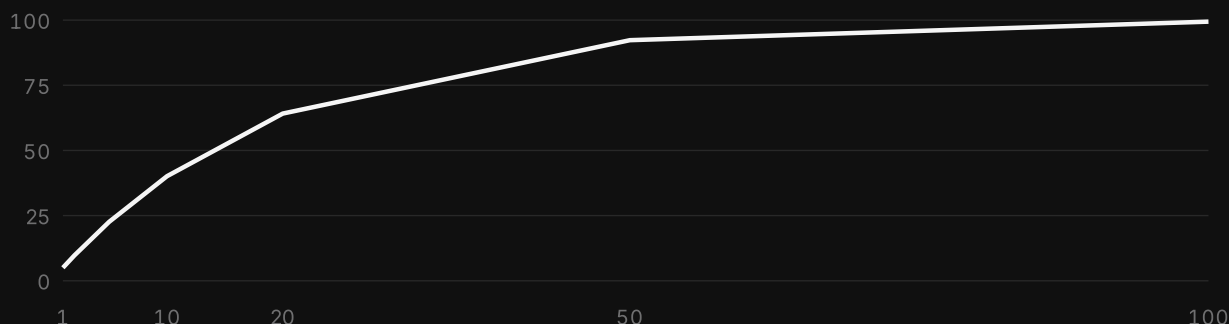
Scenario family probability under independence.

AT 100 TRIALS

99.4%

Scenario family probability under independence.

PROBABILITY (%)
— Probability (%)



Analytical scenario: independent true-null tests at exact 5% size. No calendar period.
INDEPENDENT NULL TRIALS

$$P(\geq 1) = 1 - (1 - 0.05)^m$$

Assumptions matter. The calculation uses $1 - (1 - 0.05)^m$. Each test is an independent true null with exact 5% size. The resulting family probability is 5.0% at one trial and 9.75% at two. The chart extends the same scenario through 100 trials.

Keep the distinction clear. The Deflated Sharpe Ratio addresses selection, non-normal returns, and backtest overfitting, and requires information about trial counts and dependence^[04]. This chart reports no DSR calculation. Its values describe a defined scenario, without measuring strategy prevalence or effective independent trial counts.

CAVEAT

Actual trials may be dependent, and their dependence is unknown here. The plotted probability cannot estimate the effective trial count. Dependence can shift family-level risk, while tests whose true size differs from 5% also fall outside this scenario.

Oracle t-statistics stay high in every partition

The oracle's Newey-West t-statistic stays between 13.74 and 15.89 across three retrospective partitions, while the lagged series weakens in the latest window. The partitions differ in length and were not pre-registered holdouts; all are slices of the same historical record.
[01][02]

SUBPERIODS

3

Retrospective partitions with unequal lengths

LOWEST ORACLE T

13.74

Minimum across all three windows

RECENT LAGGED T

0.77

2000-2025 partition



The oracle's HAC t-statistics are 13.74 for 1960-1979, 15.89 for 1980-1999, and 15.51 for 2000-2025. The windows contain 240, 240, and 312 monthly observations, respectively; all estimates use a six-month Bartlett Newey-West HAC estimator.
[01][02]

Lagged mean is 0.3708% per month. $\ln E[\text{position}] \times E[\text{market}] + \text{Cov}(\text{position}, \text{market})$, average exposure and sample covariance contribute 0.1174 and 0.2534 percentage points per month. This positive mean mixes both; neither establishes deployable predictive skill, and covariance is not alpha or validated prediction.
[01]

INTERPRETATION	The oracle payoff is built algebraically; repeated significance across retrospective partitions therefore offers no causal guarantee for a strategy formed with information available at the time. [01]
CAVEAT	The unequal windows are retrospective partitions of one record, not trained holdouts or independent replications; their sample sizes are 240, 240, and 312 months. Six-month Bartlett Newey-West HAC inference adjusts serial dependence, but cannot turn an algebraically constructed oracle into a causal test. [01][02]

Same-month outcome access changes the measured signal

Across the frozen sample, the oracle construction records a Newey-West HAC t-statistic of 25.73. The prior-month signal yields 2.41.^{[01][02]} This comparison isolates sensitivity to same-month outcome access, while timestamp provenance and execution evidence remain outside the test.

HAC T REDUCTION 23.32 Subtraction of two statistics, not a difference test.	NAIVE SHARPE REDUCTION 3.97 Descriptive comparison of oracle and lagged constructions.	ORACLE ROWS DEPENDENT 792 Period-index ledger count, not a timestamp audit.
--	---	--

Timing evidence

Capture both timestamps. Store observed_at and ingested_at for every input, then retain decision and order times for each signal. The zero same-month dependency is structural, derived from period-index formulas. Provider delivery, revisions, and data provenance remain unaudited here.^[07]

Inference

Keep the tests separate. HAC inference adjusts uncertainty for serial dependence. The positive unconditional mean alone establishes no market-adjusted skill. The 23.32 t-statistic subtraction describes sensitivity between constructions, without testing their mean difference or measuring total leakage damage.^{[02][08]}

Experiment record

Register every trial. Preserve a dated registry of variants, selection rules, and holdout boundaries before evaluating outcomes. Record transaction costs, funding, borrow, and realized orders alongside the signal. A genuine holdout and point-in-time inputs remain requirements for any broader performance claim.^{[04][07]}

OBSERVED	The row ledger flags same-period outcome dependency in all 792 oracle rows and none of the 792 lagged rows. ^[01] This count follows formula-period alignment. It is not an audit of vendor, receipt, or decision timestamps.
CAVEAT	The 23.32-point change subtracts two t-statistics, without testing the mean difference, quantifying effect size, or measuring total leakage damage. The frozen January 1960 to December 2025 sample sets the bounds, with no extension to new years or refreshed vintages. ^{[02][08][01][07]}

Sources and reproducibility

Eight cited records document the data inputs, inference methods, selection adjustment, GDP release vintages, and historical- data conventions relevant to the calculations.

-
- [01] **Kenneth French Data Library**
Monthly and daily files. French factor data provide the historical market series used here.
-
- [02] **Newey and West, HAC covariance estimator**
Presents a covariance estimator accounting for serial dependence and heteroskedasticity.
-
- [03] **Andrew Lo, The Statistics of Sharpe Ratios**
Sampling uncertainty matters. Derives statistical properties of estimated Sharpe ratios.
-
- [04] **Marcos M. Lopez de Prado, Deflated Sharpe Ratio**
Selection changes inference. Adjusts Sharpe-ratio evaluation for multiple testing.
-
- [05] **BEA, Q2 2024 GDP advance estimate**
The first Q2 2024 GDP estimate, released 25 July 2024.
-
- [06] **BEA, Q2 2024 GDP second estimate**
The second Q2 2024 GDP estimate, released 29 August 2024.
-
- [07] **Federal Reserve Bank of St. Louis, ALFRED**
ALFRED records data vintages. Supports retrieval by release vintage.
-
- [08] **Lazarus, Lewis, Stock and Watson, HAR Inference: Recommendations for Practice**
Variance and dependence matter. Recommends inference procedures allowing heteroskedasticity and autocorrelation.
-

Source^[01] provides French monthly/daily files, downloaded 7 October 2026 from the August 2026 CRSP vintage. Frozen retrospective windows: monthly 1960 - 2025, daily 2000 - 2025. December 1959 initializes signals. Package retains hashes, validation, and exact URLs.

[02][03] document HAC inference and Sharpe-ratio sampling considerations. These are research-only diagnostics. Sharpe ratios are annualized, while portfolio returns use arithmetic aggregation. Figures describe historical calculations, not executable investment performance.

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

blckridge.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.