



QUANT RESEARCH NOTE 06 - FINAL HISTORICAL AUDIT

Post-OPEX Cross-Sectional Effect

A frozen-candidate falsification study of US equity returns in the first two trading sessions after monthly options expiration, with causal timing, cost stress, placebo controls, cycle-level bootstrap, sub-universe replication and immutable forward governance.

ANCHORED-WF SHARPE

1.716

full historical audit, 1x costs

INDEPENDENT OPEX CYCLES

34

monthly event cycles

PSEUDO-OPEX P-VALUE

0.001998

1,000 matched randomized calendars

SHARPE AT 5X COSTS

0.725

same frozen rule

RESEARCH PARTNERS

Roberto Cocco & Amin Hydar Ali

PUBLICATION STATUS

**Frozen high-conviction research candidate - historically robust,
forward validation pending**

i Publication standard

The original Seasonality V2 campaign formally concluded NO ALPHA FOUND. This OPEX rule was preserved only as a frozen research lead. The later falsification audit is strong historical evidence, but it does not rewrite that original gate or substitute for true forward OOS after 2026-10-01.

What the current evidence supports - and what it does not

This memorandum is written for a professional reviewer. It emphasizes research governance, reproducibility, statistical controls and known implementation weaknesses rather than presenting a single attractive equity curve.

FULL WF SHARPE

1.716

87.5% positive folds

1X CAGR

7.82%

historical simulation

MAX DRAWDOWN

-1.41%

WF cost reproduction

2026 SHARPE

0.714

9 events; 44.4% win rate

WHY THE RESULT IS INTERESTING

- **Cost robustness:** Sharpe remains 1.50 at 2x, 1.27 at 3x and 0.73 at 5x modeled screening costs.
- **Event robustness:** all 34 leave-one-cycle-out Sharpe estimates remain positive; worst is 1.48.
- **Placebo specificity:** the frozen Sharpe 1.716 exceeds the 95th-percentile pseudo-OPEX Sharpe 0.971; empirical $p=0.001998$.
- **Cross-sectional breadth:** 20/20 deterministic random half-universes are positive, median WF Sharpe 1.47.
- **Concentration:** largest OPEX cycle contributes 10.8% of absolute cycle P&L; largest stock 13.0% of absolute stock P&L;

WHY THE RESULT IS NOT YET VALIDATED

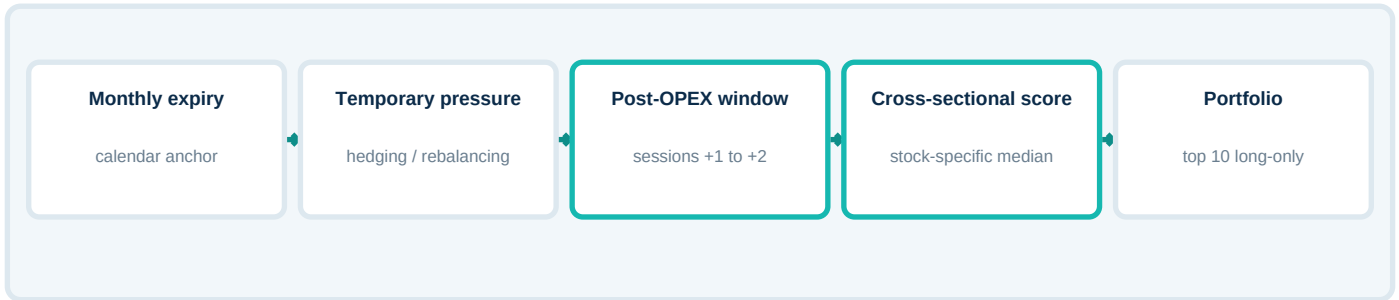
- **Recent weakening:** 2026 Sharpe fell to 0.71 and event win rate to 44.4%.
- **Recent cost fragility:** the 2025-10-01 to 2026-10-01 segment is negative at 5x costs.
- **Small independent sample:** 34 monthly cycles remain far fewer than the hourly bar count suggests.
- **Universe bias:** the current-discovered liquid-stock universe is not point-in-time survivorship-free.
- **Parameter surface:** median immediate-neighbor Sharpe is only 0.53 and the worst neighbor is -0.71.
- **No true forward evidence yet:** historical data stop before the immutable 2026-10-01 cutoff.

i QRD interpretation

Current status: **FROZEN HIGH-CONVICTION RESEARCH CANDIDATE - HISTORICALLY ROBUST, FORWARD VALIDATION PENDING.** The evidence justifies continued forward observation. It does not authorize deployment or a claim of proven alpha.

Can post-expiration equity pressure leave a repeatable cross-sectional signature?

The strategy is not simply a market-wide "OPEX day" seasonal. It ranks individual stocks by their own historical post-expiration behavior and asks whether that cross-sectional pattern survives causal execution and falsification.



Potential economic mechanisms

Mechanism	Why it could create post-OPEX behavior	What this research establishes
Dealer hedging unwind	Delta/gamma hedges established into expiration may be reduced after contracts expire.	Not identified directly. The equity effect is consistent with, but does not prove, this mechanism.
Index / benchmark rebalancing	Expiry-linked flows can interact with institutional inventory, benchmark and liquidity needs.	No direct fund-flow or dealer-positioning data are used.
Temporary price pressure	Execution demand around expiry can push prices away from short-horizon equilibrium.	The test asks whether subsequent stock-specific returns are predictable from prior post-OPEX outcomes.
Behavioral / liquidity persistence	Recurring liquidity provision or order-flow patterns may be stable by name.	The strategy is cross-sectional and stock-specific; persistence is tested statistically, not economically identified.



Mechanism discipline

The causal mechanism remains an open research question. A statistically unusual calendar effect should not be described as "dealer gamma alpha" without independent options-positioning or flow evidence.

IBKR-only historical data, with explicit limits on what the universe can prove

The audit uses Interactive Brokers historical data and an automatically replaced liquid-stock universe. This improves operational continuity but does not create a point-in-time security master.

DAILY PANEL

150 stocks

2016-10-05 to 2026-09-30

INTRADAY PANEL

60 stocks

1h target built from complete RTH blocks

1H HISTORY

4,485 bars

2023-10-03 to 2026-09-30

FORWARD CUTOFF

2026-10-01

immutable; exclusive in historical mode

Layer	Implementation	Professional review comment
Vendor	IBKR API only. Daily uses adjusted history; intraday uses RTH trades.	Operationally reproducible through the broker, but intraday history is not a specialist fully adjusted institutional database.
Universe acquisition	150 daily and 60 intraday valid common-stock histories; failed or suspicious symbols are replaced.	Replacement prevents incomplete panels but can embed current-universe / survivorship selection.
Bar construction	30-minute RTH atomic data; complete 1h/2h/4h blocks only. Partial terminal blocks are discarded.	Reduces irregular aggregation and partial-bar contamination.
Calendar	Monthly options-expiration proxy anchored to the third Friday / nearest observed prior session when necessary.	A deterministic proxy, not a complete official instrument-specific expiration calendar.
Metadata	Current IBKR sector metadata unless a frozen metadata file is supplied.	Historical sector membership is therefore not established by the default run.

i

Universe limitation

Survivorship-bias disclosure: automatic replacement is a data-availability safeguard, not a survivorship-bias cure. Publication-grade replication should use a frozen point-in-time security master or historical constituent file and point-in-time sector / size metadata.

The exact candidate is immutable: SEA_1h_opex_post_f30968bec9fe

The purpose of the final audit is to try to break this rule without changing it. Parameter neighbors are diagnostics only and never become replacement strategies.

Parameter	Frozen value	Role
Timeframe	1h	Complete regular-hours hourly blocks.
Family	opex_post	Active only after the monthly expiration anchor.
Turn window	2	First two post-OPEX sessions.
Estimator	median	Robust historical location estimator.
Lookback occurrences	40	Maximum same-bucket historical outcomes retained.
Minimum occurrences	12	No score until at least 12 eligible prior outcomes exist.
Holding horizon	12 hourly bars (~2 sessions)	Historical target and time exit horizon.
Exit	trail_3	3% trailing exit or time exit at the horizon.
Book	long_only	No short leg in the frozen candidate.
Selection	Top 10	Highest eligible cross-sectional scores.
Weighting	Inverse volatility	Reduces exposure to more volatile names.
Money management	vol_target_dd	12% target-vol scaling with drawdown throttles.
Target volatility	12%	Scale is clipped and never intentionally exceeds 125% gross in share-level audit.
Sign filter	True	Requires supportive sign / direction.
Filter mode	none	No extra trend or regime filter.

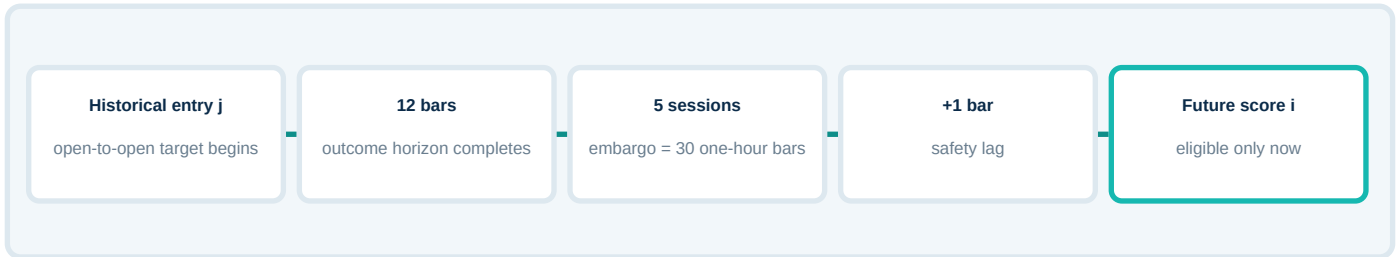


No retuning

The best neighboring configuration in the final audit reached Sharpe 1.743, slightly above the frozen 1.716. It is deliberately **not selected**. Switching to it would contaminate the freeze and convert a robustness diagnostic into a new optimization.

The design delays learning until the full historical outcome plus embargo has elapsed

For the 1h candidate, the 12-bar horizon is combined with a 5-session embargo and one safety bar. At six 1h bars per session this creates a 43-bar eligibility lag before a historical outcome may influence a future score.



Execution component	Screening layer	Share-level audit
Entry timing	Score for bar T uses only eligible completed prior outcomes. Calendar state is known before T open.	Orders are priced from the opening bar with modeled adverse slippage.
Baseline friction	5 bps one-way; multiplied 1x/2x/3x/5x.	Per-share commission + pass-through + regulatory + clearing + routing fees.
Spread/slippage	Embedded in conservative screening friction.	Spread floor + liquidity term + volatility component + square-root ADV participation impact.
Exit	12-bar horizon with trail_3 outcome logic.	3% trailing stop or time exit; OHLC bars determine hit logic.
Sizing	Portfolio weights + historical money-management scale.	Actual-share / fractional-share ledger; target notional from equity, weights and scale.

i Execution realism
Important code limitation: historical OHLC bars cannot reconstruct tick order inside a bar. If both a new intrabar high and a trailing-stop threshold are touched, the exact live path and queue position are unknown. The stop model is therefore a deterministic approximation, not tick-perfect execution.

The candidate was frozen because it was interesting - not because it passed

The most important integrity fact in this research is that the original Seasonality V2 campaign did not declare a deployable alpha. The candidate was retained in the manifest only so later evidence could falsify it without retrospective selection.



Campaign diagnostic	Observed result	Interpretation
Targeted search	4,000 experiments requested across OPEX, intraday clock, institutional calendar and controls.	Large search space creates material data-mining risk. Raw Sharpe is insufficient.
Negative-control 95th percentile	WF Sharpe 1.086.	The candidate had to distinguish itself from pseudo-calendar and pay-cycle control structures.
Frozen OPEX lead	WF Sharpe 1.71; 2x 1.50; 3x 1.26; 5x 0.72.	Strong cost degradation curve, but the absolute campaign gate still failed.
Campaign CPCV/PBO	PBO 0.3571 across 60 models.	Campaign-level overfitting diagnostic; not a candidate-specific guarantee.
Reality Check	p = 0.0036 across 5,000 paths.	Evidence of non-random structure in the model set, but still campaign-level.
Formal V2 conclusion	NO ALPHA FOUND.	Frozen leads were retained for falsification only, not deployment.

i

Why continue a gate-fail lead?

This final audit must not be interpreted as retroactively changing the V2 decision. Its scientific value comes precisely from keeping the failed-gate rule frozen and subjecting it to additional tests without moving the goalposts.

Ten stages designed to attack temporal, implementation and data-mining failure modes

Historical mode hard-clips all observations at the immutable 2026-10-01 cutoff. The manifest and source hashes are checked before testing. No post-cutoff observations are permitted into the historical audit.

01 Load frozen V2 engine

Imports the original research engine rather than re-implementing the rule.

06 Share-level attribution

Stock and cycle P&L; concentration under more detailed friction.

02 Verify manifest + acquire IBKR data

Hash checks and automatic replacement of unusable histories.

07 Independent sub-universes

Liquidity/size halves, sectors, and 20 deterministic random half-universes.

03 Exact cost reproduction + pre/post split

1x/2x/3x/5x costs; compares earlier and recent historical segments.

08 Event studies

Descriptive OPEX-relative paths; pre-entry segment explicitly non-causal.

04 Temporal/date robustness

Calendar years, expanding starts, truncated ends, rolling windows, LOYO.

09 Matched pseudo-OPEX placebo

1,000 randomized calendars with the same frozen portfolio logic.

05 Cycle robustness + bootstrap

Leave-one-OPEX-cycle-out and 10,000 cycle-level block-bootstrap replications.

10 Parameter neighborhood

39 immediate neighbors mapped diagnostically; none promoted.

Operational note

The 1,000-placebo test consumed 1h36m25s of a 1h42m39s total audit. This is intentionally heavy historical work. Routine monthly forward monitoring should not rerun the full placebo battery.

The complete historical sample remains strong under large cost multipliers

The frozen candidate produces 34 historical OPEX events. The anchored walk-forward result is strong at baseline and degrades gradually rather than collapsing immediately under cost stress.

WF SHARPE 1X

1.716

87.5% positive folds

WF SHARPE 2X

1.502

87.5% positive folds

WF SHARPE 3X

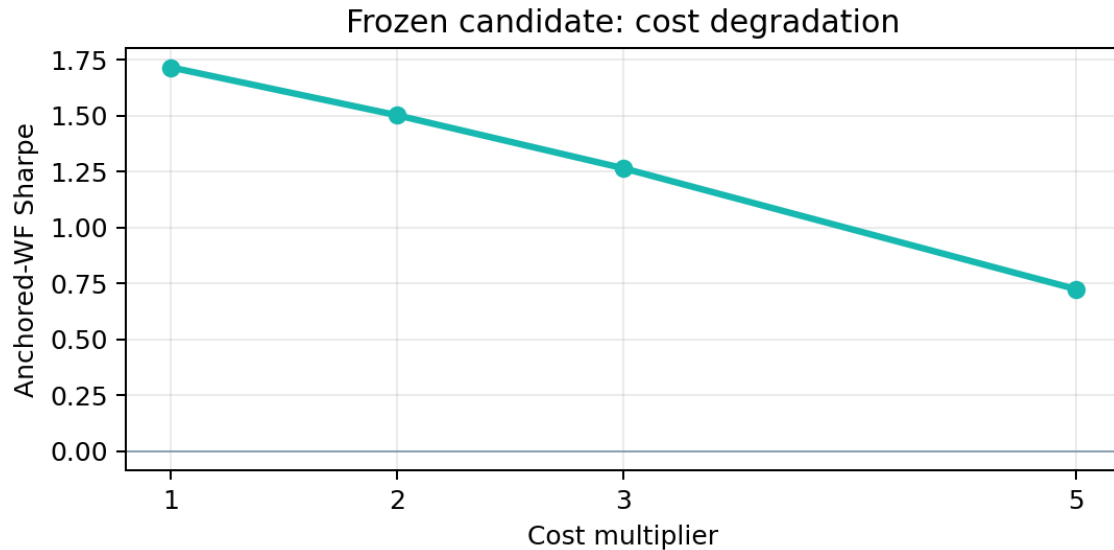
1.265

75.0% positive folds

WF SHARPE 5X

0.725

62.5% positive folds



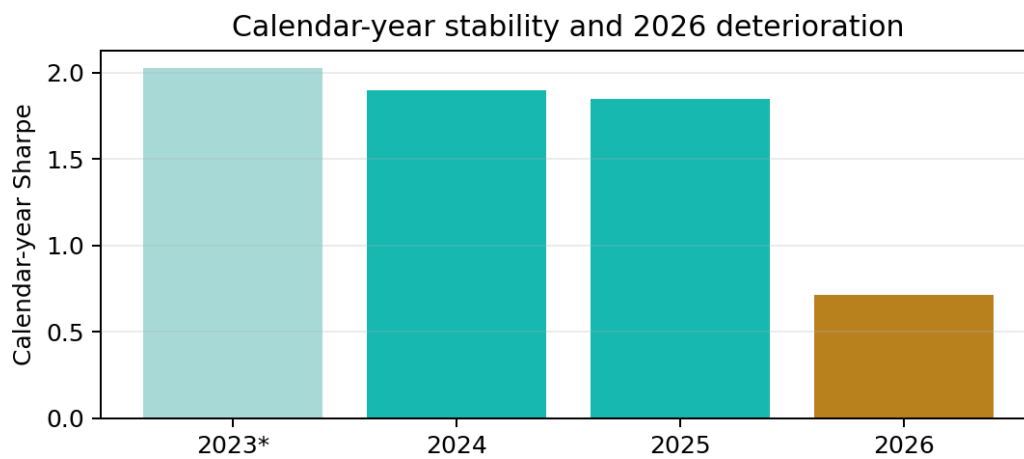
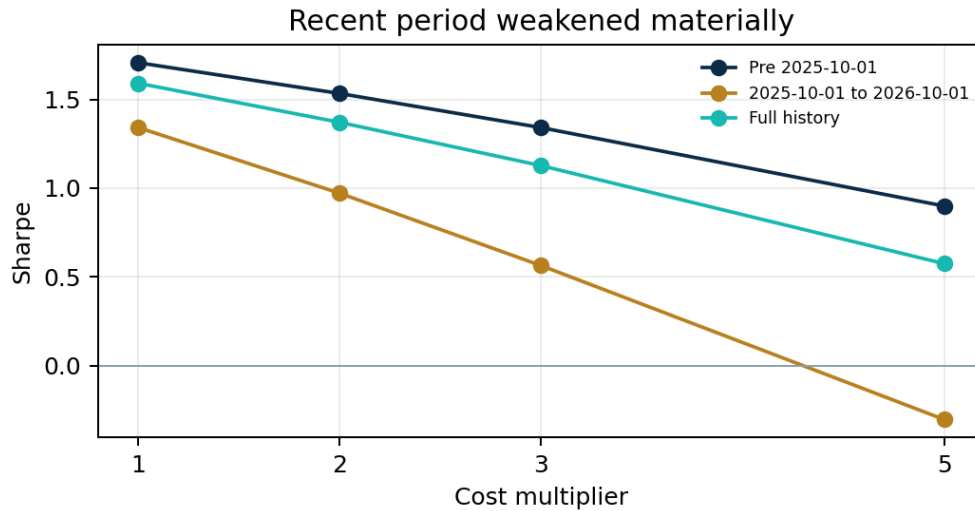
Cost	WF Sharpe	CAGR	Max drawdown	Positive folds
1x	1.716	7.82%	-1.41%	87.5%
2x	1.502	6.56%	-1.79%	87.5%
3x	1.265	5.32%	-2.49%	75.0%
5x	0.725	2.85%	-3.88%	62.5%

i Cost-model distinction

Cost robustness is a strong point, but the headline WF table uses the screening friction model. The share-level audit has a more detailed commission/spread/volatility/ADV-impact model. A professional reviewer should distinguish these two layers rather than treating all cost numbers as identical.

The most important warning: the most recent historical year is materially weaker

The post-2025-10-01 segment remains positive at baseline costs but deteriorates much faster under cost stress. This segment was still part of the original research history, so it is diagnostic - not independent forward OOS.



* 2023 contains only one event and is not statistically informative.

i Recent degradation
At 1x costs, the recent 12-event segment has Sharpe 1.34 versus 1.71 earlier. At 5x costs, the recent segment falls to -0.30. Calendar-year 2026 Sharpe is 0.71 with a 44.4% event win rate. This is not evidence that the effect is dead, but it is meaningful evidence that the edge may be decaying.

Thirty-four monthly cycles are a better unit of evidence than thousands of hourly bars

The economically relevant independent observations are monthly OPEX cycles. The audit therefore evaluates delete-one-cycle stability and resamples the cycle sequence directly rather than pretending every hourly bar is independent.

OPEX CYCLES

34

independent monthly cycles

PROFITABLE CYCLES

70.6%

1x cost audit

MEDIAN CYCLE RETURN

0.409%

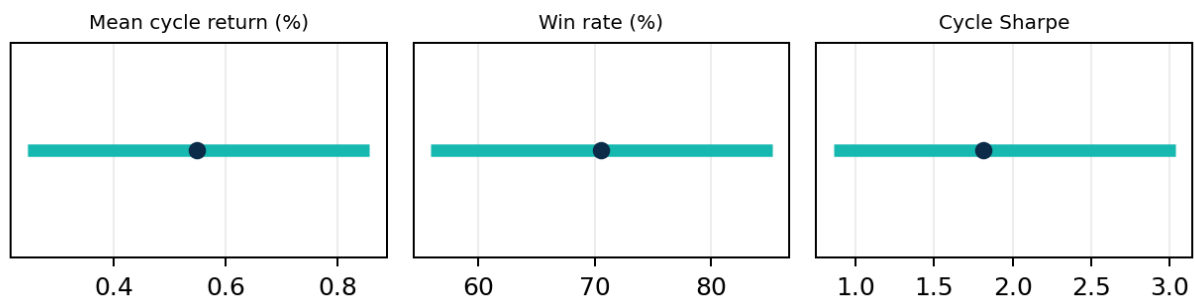
net event return

WORST DELETE-1 SHARPE

1.481

all 34 deletions stay positive

Cycle-level bootstrap: observed point and 95% interval



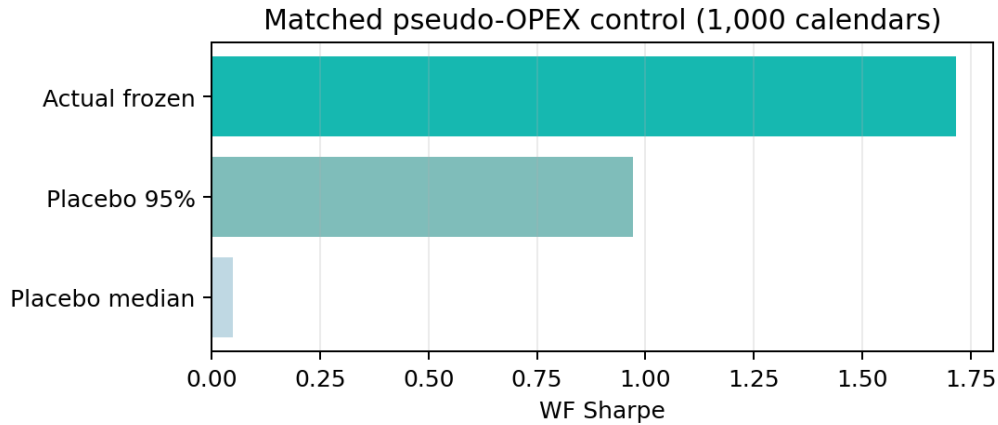
Cycle diagnostic	Observed / interval	Interpretation
Full cycle-level Sharpe	1.815 annualized	Consistent with a strong event-level effect.
Mean cycle return	0.548%; 95% bootstrap ~0.246% to 0.857%	Positive throughout the bootstrap interval under the observed cycle process.
Win rate	70.6%; 95% bootstrap ~55.9% to 85.3%	Broad uncertainty remains because 34 cycles is still small.
Leave-one-cycle-out	Median 1.559; worst 1.481; 100% positive	No single OPEX month is carrying the result.
Largest Sharpe drop after deletion	0.110	Low sensitivity to a single event relative to the full Sharpe 1.591.

i Bootstrap interpretation

Bootstrap replications are not new market histories. They quantify sampling uncertainty conditional on the 34 observed cycles and the assumed dependence structure. They do not create 10,000 independent OPEX experiments.

The actual OPEX calendar is far more informative than matched randomized calendars

The placebo test preserves the same frozen horizon, estimator, portfolio construction and money-management rules while randomizing the calendar anchor. It asks whether "some monthly date" can generate the same result.



ACTUAL WF SHARPE

1.716

frozen candidate

PLACEBO MEDIAN

0.048

1,000 randomized calendars

PLACEBO 95TH PCTL

0.971

actual remains far above

EMPIRICAL P-VALUE

0.001998

approximately 2 in 1,001 ranking formula

What this test does well

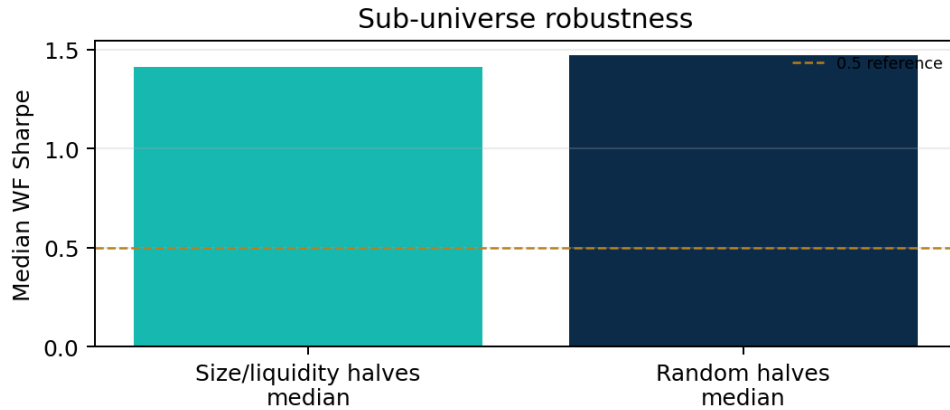
- Controls for generic calendar mining by asking whether the exact real OPEX dates are special relative to matched pseudo-dates.
- Uses the same frozen portfolio rules, so the test is not advantaged by changing the strategy after seeing the random calendars.
- Expands the earlier 200-repetition placebo to 1,000 repetitions and materially tightens the tail estimate.

What this test does *not* solve

- The p-value is conditional on the already-selected candidate. It does not erase the 4,000-experiment campaign or earlier research search space.
- It does not prove an economic mechanism; it shows calendar specificity under the test design.
- It does not remove survivorship, corporate-action or execution-model limitations.

The effect is not obviously dependent on a few stocks or one extraordinary expiry month

Breadth tests are important because a cross-sectional seasonal can look robust while actually being driven by one sector, one mega-cap cohort or a small number of event dates.



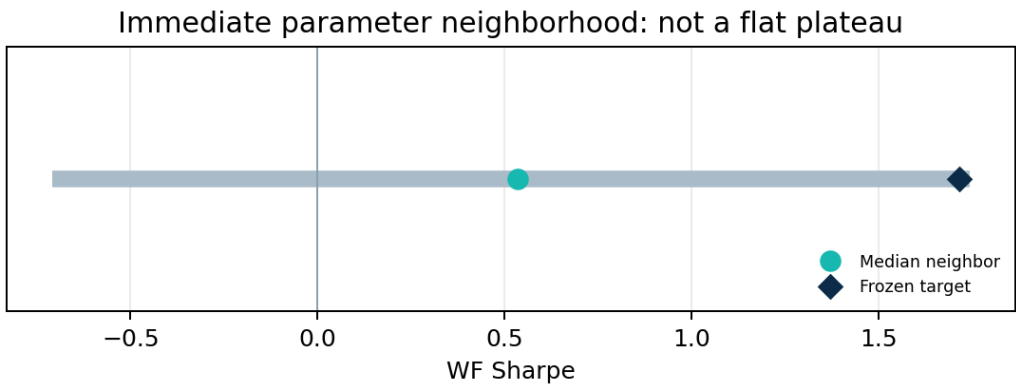
Diagnostic	Result	Interpretation
Liquidity/size halves	2/2 positive; median WF Sharpe 1.413	The effect survives a coarse split of the 60-stock intraday universe.
Random half-universes	20/20 positive; median WF Sharpe 1.470; all ≥ 0.5	Strong evidence that the result is not dependent on one particular stock subset.
Stocks with positive attribution	69.4% of 49 traded stocks	Breadth is favorable, though not universal.
Largest stock absolute-P&L; share	13.0%	No single stock dominates the result.
Top five stock absolute-P&L; share	43.7%	Moderate concentration; still material enough to monitor forward.
Largest OPEX cycle absolute-P&L; share	10.8%	No single expiry cycle dominates.
Top five cycle absolute-P&L; share	36.1%	Event concentration improved versus the earlier rewind audit.

i Sector test

The sector test is **inconclusive**: five current-classification sector groups were processed but the summary Sharpe is NaN. Current IBKR sector labels are not point-in-time historical metadata. This result should neither be promoted as a pass nor treated as a meaningful failure.

The local surface is positive more often than not - but it is not a broad flat plateau

A fragile overfit often appears as one sharp optimum surrounded by failures. Here most immediate neighbors remain positive, but the dispersion is wide enough that parameter sensitivity remains a real weakness.



NEIGHBORS

39

diagnostic only

MEDIAN NEIGHBOR SHARPE

0.534

well below frozen 1.716

POSITIVE NEIGHBORS

76.9%

23.1% are non-positive

SHARPE >= 0.5

56.4%

moderate, not universal

Neighborhood fact	Value	Professional interpretation
Worst neighbor	-0.708	Some nearby choices destroy the effect; the surface is not uniformly robust.
Best neighbor	1.743	Only slightly above the frozen target; it is not promoted.
Frozen target	1.716	Near the upper end of the local surface, which warrants caution about selection pressure.
Median neighbor	0.534	Positive but substantially below the target; robustness is moderate rather than exceptional.

i

No parameter rescue

The correct governance response is not to move to the 1.743 neighbor. The correct response is to keep the 1.716 rule frozen and let forward evidence decide whether the selected point was genuine or lucky.

A professional review should know exactly where the backtest can still be wrong

The research code contains several safeguards, but it is not an exchange simulator, a point-in-time institutional database, or an options-flow model. The table below distinguishes limitations that are structural from those that are simply awaiting forward evidence.

Weakness	Why it matters	Mitigation / next evidence
Current-discovered universe	Survivorship and selection bias are not fully removed by replacing unavailable symbols.	Replicate with frozen point-in-time constituents / security master.
IBKR intraday TRADES history	Corporate-action discontinuities can remain; anomaly filters are not equivalent to professionally adjusted intraday data.	Cross-check with a specialist adjusted database for publication-grade replication.
OHLC stop ordering	Bars do not reveal tick path, opening-auction queue position or exact trail-stop sequence.	Paper/live fills; tick or quote-level replay if available.
Screen vs share-level cost models	Headline WF stress uses 5 bps one-way screening friction; share-level audit uses a more detailed dynamic model.	Report both layers separately and prioritize realized forward slippage.
Modeled market impact	Square-root ADV impact and spread/volatility terms are assumptions, not realized fills.	Record actual IBKR executions and compare predicted vs realized friction.
OPEX calendar proxy	Third-Friday logic may differ from special expiries, holiday conventions or instrument-specific expiration dates.	Treat this proxy as frozen v1; official-calendar research must be a new version.
Only 34 independent cycles	Event-level uncertainty remains large despite thousands of bars.	Accumulate 6-12+ frozen forward cycles; do not infer confidence from bar count.
Data-mining history	Candidate emerged from a 4,000-experiment campaign and prior seasonality research.	Preserve the failed original gate; rely on frozen forward evidence, not retrospective p-values.
Recent deterioration	2026 Sharpe 0.71; recent 5x-cost segment negative.	Forward monitoring should test whether decay persists.
Parameter sensitivity	23% of immediate neighbors are non-positive; worst neighbor -0.71.	No retuning; forward validation of the frozen target only.
Sector test inconclusive	Current metadata and sparse groups prevent a clean sector inference.	Use point-in-time sector / market-cap metadata.
No true forward OOS yet	All attractive metrics remain historical.	Post-2026-10-01 registry is the decisive next dataset.

i

Audit posture

The brochure intentionally does not hide these weaknesses. The strategy is scientifically interesting because it has survived multiple falsification attempts, not because the backtest is flawless.

From 2026-10-01 onward, the research question changes from backtesting to observation

The historical dataset has been exhausted for this exact frozen configuration. Additional historical optimization would reduce the credibility of the project. The next evidence must be timestamped forward cycles generated by the immutable rule.



Forward registry field	What must be stored before / after each cycle
Cycle identity	Official month, frozen calendar anchor, POST1 / POST2 timestamps.
Pre-trade state	Eligible universe, scores, selected top 10 names, weights, target notional, predicted costs.
Execution	Actual IBKR orders/fills, timestamps, commissions, realized slippage, rejects and partial fills.
Outcome	Gross and net P&L, exit reason, bars held, realized drawdown and concentration.
Audit trail	Append-only cycle record; no deletion, backfill or parameter changes. Missing-data cycles remain documented.
Evaluation	Cumulative net return, cycle win rate, cycle Sharpe, realized-vs-modeled costs, concentration and stability.

Forward milestones

6 completed OPEX cycles: early checkpoint only. Enough to identify gross implementation failure or obvious sign reversal, not enough to validate the strategy.

12 completed cycles: materially more informative first forward evaluation. Still not an automatic live-trading approval.

Beyond 12 cycles: continue the same frozen registry if evidence remains economically coherent; any parameter change creates a new version and a new forward clock.

i Research conclusion

Final professional conclusion: the historical evidence is unusually strong for a sparse calendar effect - especially the 1,000-placebo control, 34-cycle delete-one robustness, bootstrap intervals, cost degradation and random-half replication. The two most important unresolved risks are recent performance decay and the absence of genuinely independent forward OOS. The strategy should remain frozen and observed, not optimized further.

Methodology sources: IBKR_SEASONALITY_TARGETED_VALIDATION_AUDIT_V2_COMPLETE.py

IBKR_OPEX_FINAL_AUDIT_AND_FORWARD_MONITOR_V2_PROGRESS_COMPLETE.py | QRD Research Standard v1.0

Frozen manifest SH-A256: 8773546c53b19944a480c5dddb5f1c19ffa86c871d731334649158e7d8782c7c

Disclosure: This document reports quantitative research methodology and simulated historical results. It is not investment advice, an offer, a recommendation, or a representation of future performance.

