Research summary

Peilun Cai

Working paper v1; October 30, 2026

Question. How should a reported backtest be assessed when specification searches and overlapping events are not disclosed?

Finding. Under the study's circular-shift procedure, the random-day rule called a direction in 0.314 of cases across 124 selected stocks, against a nominal two-sided 0.10 threshold. The 0.314 estimate tests the final fixed rule; it does not repeat feature or threshold selection and does not estimate the chance of any searched rule passing. A shift-based placebo rejected in 0.102 of cases, but shares the null's shift mechanism, so this is a code check rather than independent calibration evidence.

Event overlap. The median stock retained 0.366 of its qualifying days after requiring 20 trading days between events. This non-overlap ratio is not an effective sample size. Of 43 initially directional real-mask readings, 37 lose their direction after thinning; 11 readings remain directional overall. This does not establish that genuine return effects disappear or identify which readings are correct.

Limits. The selected pool is not a random sample. Stocks share a market calendar; the bootstrap treats them as independent, and interval coverage is unverified. The shifted mask has not been shown exchangeable under all relevant nulls. The work is exploratory and retrospective: its external timestamp postdates the original results and does not establish prospective registration. The first full-run outputs have been reconstructed to recorded local hashes; complete smoke outputs and independent trial completeness remain unresolved. A constant-price fixture exposes a strict-tie rejection artifact. An author-side replay of the original full-sample random-day null arm found no exact ties; split samples, other trials and near ties were not checked. The data are commercial market data (not redistributed). Only derived statistics are reported; no price series or downloadable datasets are included.

The smoke replay used two stocks and twenty replicates each. Its H1b 90% interval [0.10, 0.15] failed the [0.07, 0.13] requirement. The stopping rule has no smoke exemption; the replay nevertheless wrote results, and drafting continued. This violation does not by itself overturn the unchanged full-sample estimates. Original smoke files and runtime remain missing. The 28 documented attempt groups do not establish a complete trial denominator.

Next test. The prospective pooled-panel protocol has been accepted and locked. Calendar timestamp receipts exist; Bitcoin confirmation and independent custody verification remain pending. Retained-data collection cannot begin before confirmation. Full-search and multiple-shift checks are fixed in the protocol. An additional v2 simulation comparison will test false-positive control and power against White's Reality Check and Romano–Wolf procedures under known null and alternative settings. Higher power is a hypothesis, not a result.

Status. Typeset manuscript SHA-256 `922a8f1f6e23e545624e5241080dd98c95817156db78294537f3b37aa5815775`. Frozen scientific manuscript SHA-256 `4867d518213043c4706650ff03c70b44b4df228fe7b5b6174671209cc90626ba`; estimates: `out/results.json`, SHA-256 `6d5b94bfe318ec77097173e5d8ae3537db406566f4c76f4c824eba370e74a32f`. Working paper v1, October 30, 2026. Registration: `https://axiocairn.com/research/prereg/paper1/`. No DOI has been issued. A methodological extension (v2) is in progress.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process. During the preparation of this work, the author used Claude (Anthropic) and Codex (OpenAI) for code development, drafting assistance and language editing. The author designed the study, determined the principal arguments and interpretations, reviewed and revised all content, verified all data and citations, and takes full responsibility for the content of this article.