BACKTEST MIRAGE: DECISION-AWARE AUDITING OF FINANCIAL MACHINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Small errors in a model's scores can change the economic conclusion of a financial backtest. A small change near the selection cutoff can replace a stock, while a much larger change elsewhere leaves the portfolio unchanged. We introduce decision-aware economic auditing: rank candidate interventions by the investment decisions they change, then measure economic impact through paired portfolio continuations. Across all 24 future-input conditions, exposure near the cutoff produces smaller score changes but larger economic deviations than exposure farther away. On CSI 300 and CSI 500 with Ridge and LightGBM, our learned Decision priority captures 34% more impact than a learned score-only baseline at the same 10% checking budget in a frozen 2025 test. In a secondary rank-weighted evaluation, exploratory Allocation captures impact from weight changes among the same stocks. Decision information thus directs a fixed checking budget toward errors with larger effects on the economic evidence from backtests. Code is available at Anonymous GitHub.

1 INTRODUCTION

Financial machine learning uses historical backtests to judge whether a model's predictions have economic value. Predictions select stocks, and the backtest simulates the purchases, sales and costs needed to hold them (Gu et al., 2020). An evaluation error can therefore change which model appears useful; its economic effect depends on the investment decisions it changes.

Consider a model that buys the top 10% of ranked stocks. A large score change far from the cutoff can leave every purchase unchanged. A smaller change across the cutoff replaces a selected stock (Figure 1). The account's holdings and cash then determine how that change is traded. Ranking errors by score-change magnitude can therefore direct checks toward cases with smaller economic consequences.

Leakage studies show how invalid information access changes predictive performance (Kaufman et al., 2012; Kapoor & Narayanan, 2023; Roth, 2026). Financial studies examine leakage, point-in-time data and execution assumptions through their effects on investment outcomes (Bhand & Joshi, 2026; Fonseca, 2026; Yin et al., 2026). These studies establish how evaluation errors change reported performance. An economic audit must also decide which supplied interventions to check first. Score diagnostics measure prediction changes; the relevant audit target is the distortion of investment outcomes. At a fixed checking budget, this distinction determines which economic errors receive verification. We address this prioritization problem using the investment decisions that an intervention changes.

We introduce *decision-aware economic auditing*. Each check changes one investment decision and compares portfolio continuations from the same account state. *Replacement* ranks cases by the fraction of selected stocks that change. *Allocation* measures changes in intended weights and equals Replacement under equal weighting. *Decision* learns from historical checks using score and decision descriptors. Decision combines this information with score and boundary descriptors to distin-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

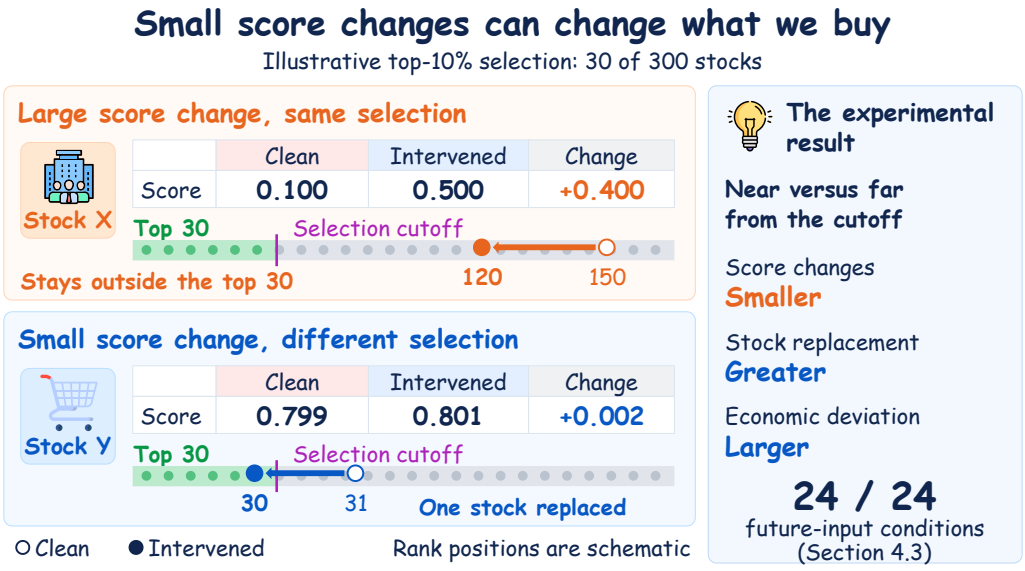


Figure 1: **Score changes can misorder economic consequences.** A large score change leaves Stock X unselected; a small change moves Stock Y across the cutoff. Only one stock's score changes in each illustrative example. The right panel summarizes all 24 future-input conditions; Figure 4 shows the measured ratios.

guish changes with different economic consequences, learning which cases deserve the same limited checks. We measure how much additional impact decision information helps recover from the same cases at the same checking budget.

Our contributions are:

- **An audit objective tied to economic evidence.** We define audit value by the economic impact recovered at a fixed checking budget, and introduce selection-based, allocation-based and learned priorities to choose which paired portfolio checks to run.
- **Why score changes misorder economic consequences.** Five protocol axes, training–scoring comparisons and account-state experiments trace how errors reach portfolio returns. In all 24 future-input conditions, near-cutoff exposure produces smaller score changes but larger economic deviations than exposure farther away.
- **More economic impact recovered with the same checks.** In the frozen 2025 test, Decision captures 34% more impact and finds 40% more high-impact cases than the learned Score-only baseline at a 10% checking budget. A secondary rank-weighted evaluation shows how exploratory Allocation recovers impact from weight changes among retained stocks.

2 Economic Auditing Under a Fixed Checking Budget

2.1 Forecasts, Selections and the Executing Account

At close t, a forecaster ranks eligible stocks for entry at open $t+1$ and scheduled exit at open $t+6$. The target is the five-session open-to-open return

$$y_t = O_{t+6}/O_{t+1} - 1. \tag{1}$$

Predictive accuracy is the daily cross-sectional Spearman correlation between scores and these returns (RankIC). The audit instead follows the decisions those scores induce and the net returns they change. We use an archived BaoStock extract[1] for the Chinese CSI 300 and CSI 500 universes,

[1]BaoStock data platform: http://baostock.com.

with date-effective membership, 19 features (All19), and Ridge and LightGBM forecasters (Hoerl & Kennard, 1970; Ke et al., 2017). We focus on Chinese equities, where trading restrictions make the distinction between intended and executed portfolios explicit. Within this shared market setting, we systematically vary information access, exposure location, account state, and allocation rules to study what makes economic audits effective. Both branches use the same archived panel, which records date-effective membership within a single vendor snapshot. We selected configurations on clean CSI 300 observations from 2016–2021, then fixed them. Each annual fit uses five training years, one validation year and one test year, with label purging and training-only preprocessing (Appendix A).

The primary policy selects $K_t = \max\{1, \lceil 0.10 N_t \rceil\}$ of N_t eligible stocks with deterministic tie-breaking. Each daily sleeve targets one fifth of pre-trade wealth, equally weighted within the selected-stock set; five sleeves overlap. Execution permits internal renewal, retains failed sales and scales purchases to cash without borrowing or substitutes. External fills cost 10 basis points (bp) per side. Appendix H specifies the self-financing account and execution data.

2.2 Audit cases and paired economic checks

Each supplied intervention provides changed inputs and clean/intervened scores. For a given account window, universe, forecaster and date, interventions with identical complete ordered selections form one audit case; unchanged selections remain eligible. A fixed rule chooses one intervention's full diagnostic vector without using outcomes. A check starts two branches from the same account state after open t, changes the next selection in one, then resumes clean selections in both. Market data and trading rules are shared; cash, holdings and fees evolve separately. Case i has economic impact

$$A_i = \frac{10^4}{6} \sum_{h=1}^{6} |r^i_{t+h} - r^0_{t+h}|, \tag{2}$$

the mean absolute paired net-return difference in basis points per day. The six sessions cover entry and scheduled exit, with residual positions retained. Absolute deviation captures both inflated and depressed performance and prevents opposite daily distortions from cancelling.

Membership identifies selected stocks; their order determines rank-dependent weights. The account executes these allocations under cash and trading constraints. Deduplicating by ordered selection therefore preserves within-set changes relevant to rank weighting.

2.3 Impact capture at identical budgets

We group audit cases into retrospective universe–forecaster–month queues. For queue q containing n_q cases, budget b permits each priority g to select $\lfloor b n_q \rfloor$. With $Q_g(b)$ the union of its selections, impact capture (Capture) is

$$C_g(b) = \frac{\sum_{i \in Q_g(b)} A_i}{\sum_i A_i}. \tag{3}$$

Capture reports the percentage of impact recovered. We also count checked high-impact cases, $A_i > 1$ bp/day (Found), and their share of all high-impact cases (Recall). The primary budget is 10%, with identical cases and integer quotas for every priority. Exhaustive offline checks supply the evaluation denominator. Priorities use supplied score pairs and decision-time descriptors, excluding subsequent returns and realized execution outcomes. The budget counts paired checks. Appendix N separately measures diagnostic and continuation costs.

3 Decision-aware audit priorities

In the one-decision audit, identical ordered selection π_t, initial state, subsequent clean decisions and market path produce identical economic paths: $\pi^i_t = \pi^0_t$ implies $A_i = 0$. Figure 2 separates prioritization from verification.

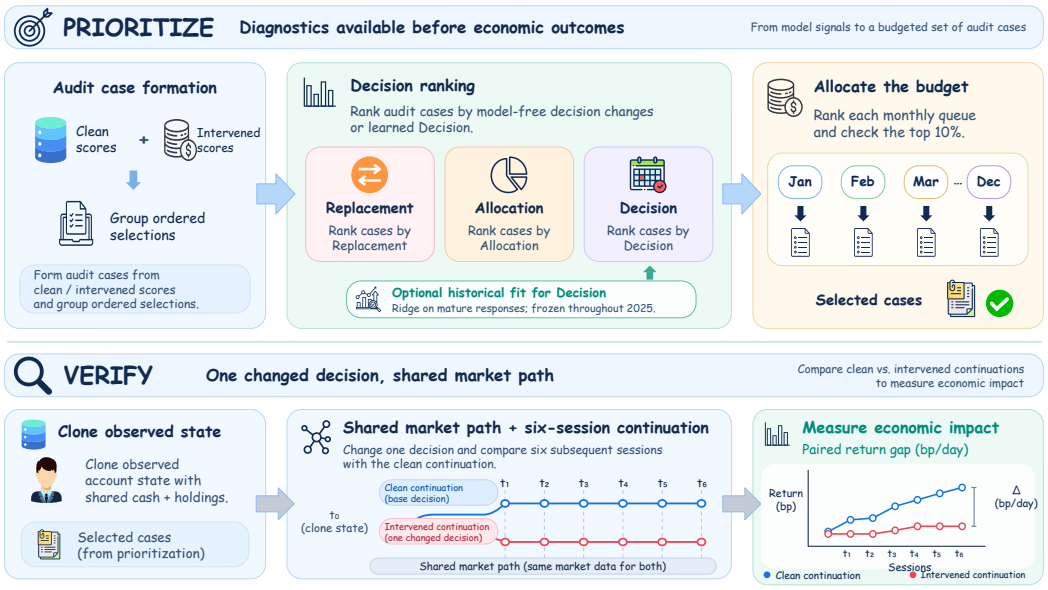


Figure 2: **Prioritize decisions, then verify their economic impact.** Top: rank cases before outcomes are observed; only Decision requires fitting. Bottom: change one selection from a common account state under a shared market path. Impact averages absolute daily return gaps over six sessions. Curves are schematic.

Replacement: a model-free audit priority. For clean and intervened selected-stock sets S_t^0 and S_t^i, selected-stock replacement is

$$R_t^i = \frac{|S_t^0 \setminus S_t^i|}{K_t}. \tag{4}$$

Replacement ranks R_t^i in descending order with the shared outcome-independent tie rule. It uses the selected-stock set alone, requiring neither historical audit responses nor future returns.

For unequal weights, *Allocation* ranks $V_i = \frac{1}{2}\sum_j |w_j^i - w_j^0|$, the difference between clean and intervened intended sleeve allocations. Equal weights give $V_i = R_i$ exactly. Rank weights also expose redistributions among retained stocks (Figure 6). This compares two decisions at one time; actual trading turnover additionally depends on existing holdings and fills.

In a one-period fixed-exposure model, $\Delta\rho = \alpha \Delta w^\top u - \Delta c$: α is the invested wealth fraction, u the asset returns and Δc the cost difference normalized by common initial wealth. Conditional on decision-time information,

$$\mathbb{E}_t|\Delta\rho| \le \alpha \sum_j |\Delta w_j|\, \mathbb{E}_t|u_j| + \mathbb{E}_t|\Delta c|. \tag{5}$$

Equation 5 links weight changes to the return difference they can cause. When return scales are comparable, Allocation measures how much capital changes its exposure. The six-session check then accounts for realized returns, fills and cash constraints. Appendix N.4 gives the derivation and its connection to A_i.

Decision: learning from historical economic checks. The learned score-only priority (Score-only) uses 17 exposure, score and past-volatility descriptors; Decision adds eleven replacement, ordered-selection and cutoff-geometry descriptors. With the same regression procedure, we measure the gain from adding decision descriptors to scores. Appendix E defines every input.

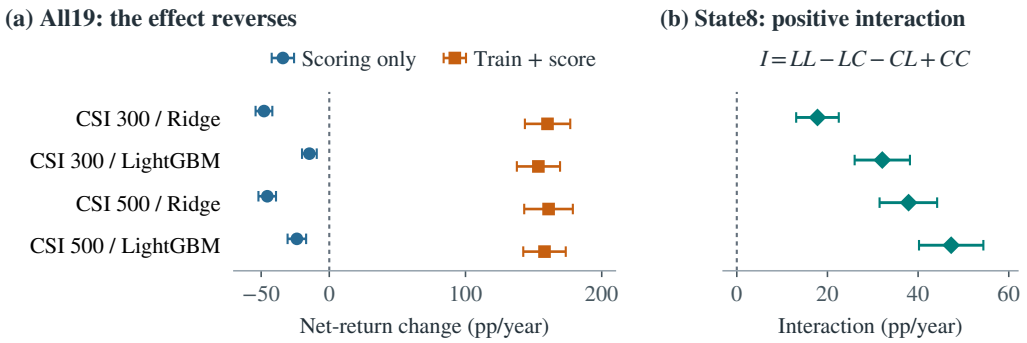


Figure 3: **Information entry can reverse economic effects.** Next-session exposure, 2022–2024. (a) All 19 features (All19), relative to CC. (b) Eight market-state features (State8). C/L: clean/exposed training (first letter) and scoring (second). Bars: pointwise 95% HAC20 intervals (Table 29).

We fit Ridge regressions of A_i by universe and forecaster, using penalty 10, an intercept and training-only standardization, then set negative predictions to zero. Training uses cases with fully observed six-session responses; the 2025 coefficients stay frozen. Decision+State adds 27 account descriptors (Appendix E).

4 Why Decision Information Matters

We examine how evaluation errors become changes in investment returns. The experiments trace four parts of this process: information exposure, its use by the forecaster, its location in the ranking, and execution through the account. Each comparison holds the remaining procedure fixed.

4.1 More Exposure Need Not Cause a Larger Return Change

Comparison. We vary five assumptions over 2022–2024: observation timing (OBS), preprocessing fit (FIT), future universe membership (UNIV), advance knowledge of tradability (TRAD), and execution timing (EXEC). The 84 nonzero universe–forecaster conditions share the portfolio rules. We measure the signed change in annualized arithmetic net return, ΔNet, $100 \times 252\,\mathrm{mean}_t(r_t^a - r_t^0)$, in percentage points (pp).

Result. Full-calendar TRAD exposure replaces fewer than 0.08% of selected stocks and changes returns by less than 0.12 points. In contrast, same-open EXEC changes no scores but lowers returns by 15.77–48.78 points. Exposure can be extensive without changing selections, while execution can change returns without changing predictions. Appendix B gives the full grid and Appendix J gives the public-software cases.

4.2 The Learning Entry Point Can Reverse the Effect

Comparison. We keep the feature representation fixed and vary where future inputs enter: scoring only (CL), training only (LC), or both (LL), relative to clean training and scoring (CC). The first letter denotes training.

Result. With All19, scoring-only exposure lowers returns, whereas joint exposure turns negative clean returns into apparent profitability (Figure 3a). The same kind of future information has a different effect when the model learns to use it. The eight market-state features (State8) also show positive training–scoring interactions (Figure 3b). Sparse exposure can change selections and returns with little RankIC change (Appendix C).

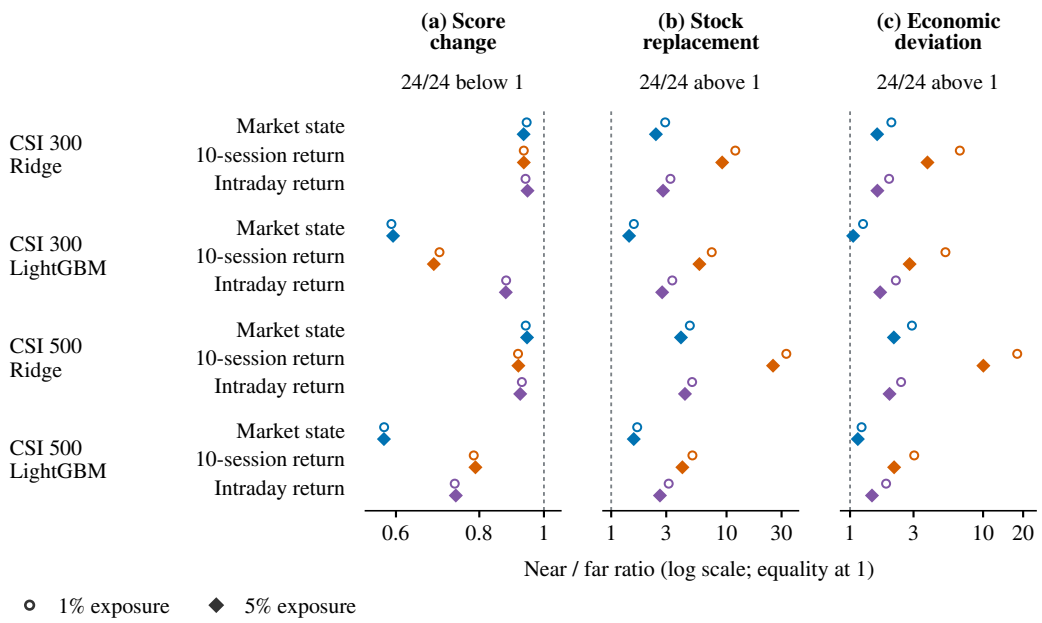


Figure 4: **Smaller score changes accompany larger economic deviations.** All 24 future-input conditions. Each channel has its own row within a universe–forecaster group; 1% and 5% doses occupy separate vertical positions. Points compare pooled near/far means on log scales; one denotes equality. Score and economic means use 726 dates, replacement 685 decision dates. These are descriptive ratios; paired near-minus-far intervals are in Appendix D.

4.3 Smaller score changes can cause larger economic distortions

Comparison. We assign equal exposure near and far from the selection cutoff. Assignment uses clean ranks, with matched counts of selected and unselected stocks over 20 seeds. Three future-feature channels, two doses, two forecasters and two universes give 24 conditions. Unavailable future vectors retain their clean values. We compare score change, selected-stock replacement and economic deviation, the mean absolute daily net-return difference over the full path in bp/day. Symmetric-noise controls test the same location mechanism.

Result. Every condition gives the same reversal: near-cutoff exposure moves scores less but changes selections and returns more (Figure 4). The economic-deviation ratio reaches 18.05. Including noise controls, all forty pooled near-minus-far contrasts have positive pointwise intervals for replacement and economic deviation.

The selection rule explains the reversal. Replacing selected stock i with stock j requires $\delta_j - \delta_i \geq s_i^0 - s_j^0$ under fixed tie-breaking. A score change matters relative to the gap it must cross, rather than through its absolute size alone.

4.4 The account determines how a changed selection is executed

We hold the clean and intervened selections and the market path fixed, then vary starting holdings and cash. A maturing position can be renewed internally, while a new purchase requires available cash. Proposition 1 constructs opposite responses under equal wealth and identical prices. On 684 CSI 300 dates, replacing historical holdings with equal-wealth cash changes mean absolute sensitivity from 0.77 to 0.79 bp/day (Appendix I). The paired check must therefore preserve account state. Section 5 then tests whether observing state also improves audit priorities.

Table 1: **Decision information makes the same checks more productive.** Impact capture and high-impact cases found at the same 10% checking budget.

Priority	2023–2024	2025
A. Capture (%)↑		
Random	9.92	9.92
Raw score change	22.11	21.83
Replacement	26.49	29.04
Score-only	22.64	22.56
Decision	**26.82**	30.24
Decision+State	26.78	**30.49**
Oracle	34.79	41.09
Decision − Score-only (pp)	**+4.19 [3.23, 5.14]**	**+7.68 [4.77, 10.58]**
Decision − Replacement (pp)	+0.34 [-0.17, 0.84]	+1.19 [0.54, 1.85]
B. High-impact cases: Found↑		
Checks available	8,861	3,482
High-impact cases	28,549	7,894
Found: Score-only	6,264	1,792
Found: Replacement	7,552	2,365
Found: Decision	**7,623**	**2,508**

Decision–Score-only is the primary contrast. The direct Decision–Replacement comparison is post hoc.

5 Decision-aware auditing captures more economic impact

The protocol studies locate errors across the backtest; the audit now tests prioritization among supplied future-input interventions. Every priority receives the same cases and checking budget, allowing us to measure the value of observing the decisions those interventions change.

5.1 Candidates and the frozen temporal test

Three next-session substitutions (return, price and volume), three stock fractions (1%, 5%, 10%) and ten uniform masks give 90 interventions per universe–forecaster–date. Cutoff-independent masks let score diagnostics rank cases within channel–dose–date groups under fixed All19 forecasters (Appendix E). Appendix G tests audit ranking at equal score-change magnitude.

Development (2023–2024) has 89,000 audit cases; the frozen 2025 test has 34,984 over 184 decision dates. Candidate design, the 10% budget and the primary Decision-minus-Score-only contrast with its positive-interval criterion were fixed before outcome access (Appendix K). Auditors train on 2022–2024 cases with six-session responses ending before 2025.

Paired 95% intervals use heteroskedasticity- and autocorrelation-consistent inference at lag 20 (HAC20) in development, and four 46-date blocks with a t_3 critical value in 2025. Calibration precedes outcome access (Appendices E and F). Oracle ranks known impacts as an upper reference.

5.2 The same checks recover more economic impact

Decision improves development capture by 4.19 points over Score-only. The frozen test confirms the gain: 30.24% versus 22.56%, a 7.68-point difference [4.77, 10.58] that meets the primary criterion (Table 1). With the same 3,482 checks, Decision finds 716 additional high-impact cases. This is 34% more impact and 40% more cases, using unrounded results. Figure 5 shows the gain across checking budgets.

The result also holds within the audit pools. Decision improves over Score-only in all four universe–forecaster groups in both periods, with eight positive pointwise paired intervals. All nine channel–dose pools improve in both periods. The gain therefore persists when the violation type and dose are the same (Appendices E and F).

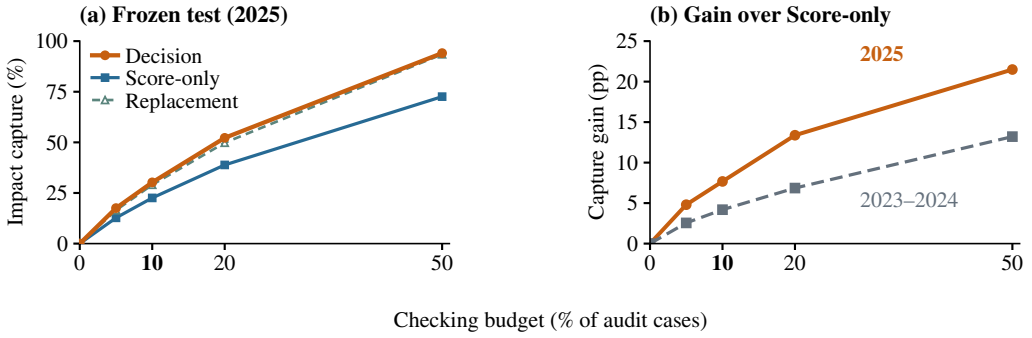


Figure 5: **Decision information improves capture across checking budgets.** (a) Frozen 2025 priorities. (b) Decision minus Score-only in 2023–2024 and 2025. Curves connect measured point estimates. Table 1 reports paired 95% intervals at the primary 10% budget. Cases and quotas are identical within each period.

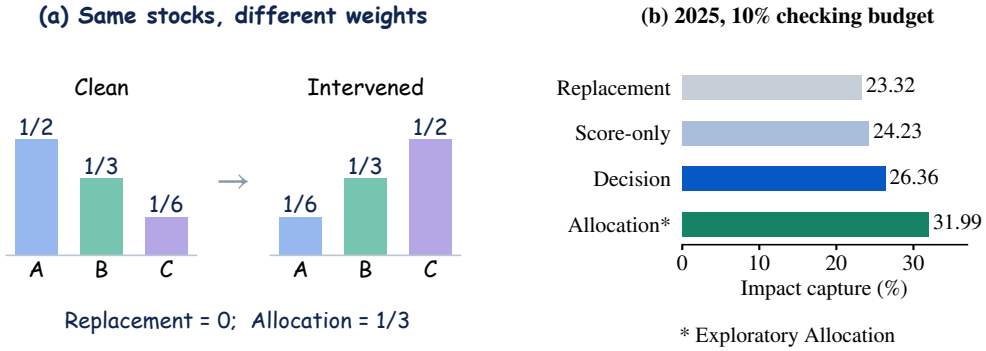


Figure 6: **Allocation captures impact that changed stock names miss.** (a) Illustrative reweighting keeps the same stocks but changes invested capital. (b) Secondary rank-weighted evaluation: 2025 impact capture at 10% checks. Allocation is exploratory. Bars are point estimates; Table 64 gives intervals and full comparisons, and Table 65 gives paired contrasts.

5.3 What changed selections and learning each contribute

Replacement captures 29.04% of impact in 2025 without fitting an audit model. Decision raises capture to 30.24%; the post hoc paired gain is +1.19 points [0.54, 1.85], with 143 more high-impact cases found. Decision's exclusive checks have lower mean replacement but higher mean impact than Replacement's: learning distinguishes consequences beyond the replacement count. The development increment is +0.34 points [-0.17, 0.84] (Appendix N).

Paired checks preserve account state to measure economic impact. Decision+State changes capture by +0.25 points [-0.40, 0.90] in 2025 and -0.05 [-0.36, 0.27] in development; both paired intervals include zero.

5.4 Rank weights require observing capital allocation

Rank weights redistribute capital among retained stocks, and equal replacement counts can move different amounts of capital. We evaluate the resulting allocation changes by setting $w_r = 2(K - r + 1)/[K(K + 1)]$ for rank r among K selected stocks. Forecasts, interventions, selected-stock sets and the five-session holding rule stay fixed. Score-only and Decision keep their fitted coefficients; past volatility follows the new clean account.

Across the same 34,984 cases and 3,482 checks, Decision improves capture over Replacement by 3.04 points [1.23, 4.85]. Exploratory Allocation reaches 31.99%, gaining 7.77 points [5.62, 9.92] over Score-only and 5.64 points [4.43, 6.84] over Decision (Figure 6b).

Cases that preserve the selected-stock set account for 26.08% of total impact. Allocation finds 130 high-impact cases in this group; Replacement and frozen Decision check none at 10%. They contribute 1.97 points to Allocation's gain over Decision, with 3.67 points from cases that change stock names. This explains why the diagnostic should follow intended weights when the portfolio rule makes rank determine capital.

These evaluations test decision-aware auditing in two Chinese equity universes under equal and rank weighting. They measure impact capture for controlled interventions at fixed paired-check budgets; Appendix M details the scope and operational use.

6 Related work

Backtest reliability. Data-snooping tests and backtest-overfitting analysis ask whether selecting a strategy from historical results creates misleading evidence (White, 2000). Research protocols organize safeguards for financial learning (Arnott et al., 2019). We measure the economic impact of supplied audit cases and test which priorities recover more impact with the same checks.

Leakage and implementation. Preprocessing and temporal splits can change predictive estimates and model comparisons (Moscovich & Rosset, 2022; Ji et al., 2023). Roth relates leakage mechanisms and dose to predictive effects (Roth, 2026). Financial studies examine leakage, point-in-time construction and cross-engine disagreement through investment outcomes (Bhand & Joshi, 2026; Fonseca, 2026; Yin et al., 2026). We connect these questions to audit decisions: observed selection and allocation changes guide which economic effects to measure first.

Choosing what to inspect. ActiveClean selects records to clean while updating a predictive model (Krishnan et al., 2016). Active testing selects labels to estimate model risk (Kossen et al., 2021); guided exploration finds confident prediction errors (Lakkaraju et al., 2017). We use decision descriptors to select supplied audit cases for paired economic checks, aiming to capture more impact. Surrogate-based evaluation further reduces labeling costs by learning to estimate predictive risk (Kossen et al., 2022). Our objective rewards the economic impact recovered by the selected checks, making portfolio selection and allocation central to the acquisition rule.

Prediction and decisions. Decision-focused learning trains predictors for downstream utility (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022), with theory addressing decision regret (El Balghiti et al., 2019). We instead use decision information to evaluate a fixed forecaster. Appendix L compares these research objectives.

7 Conclusion

Backtest errors affect economic evidence through selection, allocation and execution. Decision-aware auditing uses this connection to choose better checks: the frozen 2025 test captures 34% more impact than Score-only at the same 10% budget. Post hoc analysis measures a further gain over Replacement. The secondary rank-weighted evaluation shows why the diagnostic must also follow intended weights, with exploratory Allocation capturing 32% of impact. The resulting audit connects prediction changes to the economic evidence a backtest provides.

AI USE STATEMENT

A generative AI assistant assisted with polishing the paper.

ETHICS STATEMENT

This study evaluates historical financial-research procedures and does not deploy a trading strategy or involve human participants. Its results concern the interpretation of backtests, not investment recommendations or evidence of live profitability. We distinguish controlled interventions from feasible information access in trading, identify data and software sources, and do not redistribute material without appropriate rights.

REPRODUCIBILITY STATEMENT

The main text defines the economic audit, decision-aware priorities, controlled mechanisms and temporal evaluation. The appendices specify the data and forecasting models, self-financing account, complete intervention grids, audit descriptors, mature-response fitting, paired uncertainty and public-software cases. The 2025 confirmation appendix records the fixed primary contrast, pre-outcome calibration, independent account windows and verification scope. Complete numerical tables accompany the compact main-text results. Source data and reconstruction procedures are documented; archived market observations remain subject to the provider's redistribution terms.

REFERENCES

Rob Arnott, Campbell R. Harvey, and Harry Markowitz. A backtesting protocol in the era of machine learning. *The Journal of Financial Data Science*, 1(1):64–74, 2019. doi: 10.3905/jfds.2019.1.064.

Kavya Bhand and Aadi Joshi. The illusion of alpha: Quantifying hidden data leakage in financial machine learning. Research Square preprint, 2026. URL https://doi.org/10.21203/rs.3.rs-9180656/v1.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, NIPS '17, volume 30, Long Beach, CA, USA, 2017. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper/2017/hash/3fc2c60b5782f641f76bcefc39fb2392-Abstract.html.

Othman El Balghiti, Adam N. Elmachtoub, Paul Grigas, and Ambuj Tewari. Generalization bounds in the predict-then-optimize framework. In *Advances in Neural Information Processing Systems*, NeurIPS '19, volume 32, Vancouver, BC, Canada, 2019. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper/2019/hash/a70145bf8b173e4496b554ce57969e24-Abstract.html.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022. doi: 10.1287/mnsc.2020.3922. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922.

Xavier Fonseca. Point-in-time backtesting of momentum-trend equity strategies: A formal bias taxonomy, ATR trailing stop analysis, and investor-experience metrics. *Mathematics*, 14(12):2182, 2026. doi: 10.3390/math14122182. URL https://www.mdpi.com/2227-7390/14/12/2182.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.

Rustam Ibragimov and Ulrich K. Müller. t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, 28(4):453–468, 2010. doi: 10.1198/jbes.2009.08046. URL https://www.princeton.edu/~umueller/tstat.pdf.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993. doi: 10.1111/j.1540-6261.1993.tb04702.x.

Yitong Ji, Aixin Sun, Jie Zhang, and Chenliang Li. A critical study on data leakage in recommender system offline evaluation. *ACM Transactions on Information Systems*, 41(3):75:1–75:27, 2023. doi: 10.1145/3569930.

Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023. doi: 10.1016/j.patter.2023.100804.

Shachar Kaufman, Saharon Rosset, Claudia Perlich, and Ori Stitelman. Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6(4):15:1–15:21, 2012. doi: 10.1145/2382577.2382579.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, NIPS '17, volume 30, Long Beach, CA, USA, 2017. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, ICML '17, volume 70, pp. 1885–1894, Sydney, Australia, 2017. PMLR. URL https://proceedings.mlr.press/v70/koh17a.html.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Tom Rainforth. Active testing: Sample-efficient model evaluation. In *Proceedings of the 38th International Conference on Machine Learning*, ICML '21, volume 139, pp. 5753–5763, Virtual, 2021. PMLR. URL https://proceedings.mlr.press/v139/kossen21a.html.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Thomas Rainforth. Active surrogate estimators: An active learning approach to label-efficient model evaluation. In *Advances in Neural Information Processing Systems*, NeurIPS '22, volume 35, pp. 24557–24570, New Orleans, LA, USA, 2022. Curran Associates, Inc. doi: 10.52202/068431-1783. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/9b9cfd5428153ccfbd4ba34b7e007305-Paper-Conference.pdf.

Sanjay Krishnan, Jiannan Wang, Eugene Wu, Michael J. Franklin, and Ken Goldberg. ActiveClean: Interactive data cleaning for statistical modeling. *Proceedings of the VLDB Endowment*, 9(12):948–959, 2016. doi: 10.14778/2994509.2994514. URL https://www.vldb.org/pvldb/vol9/p948-krishnan.pdf.

Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Eric Horvitz. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '17, volume 31, pp. 2124–2132, San Francisco, CA, USA, 2017. AAAI Press. doi: 10.1609/aaai.v31i1.10821.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Víctor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 81:1623–1701, 2024. doi: 10.1613/jair.1.15320. URL https://www.jair.org/index.php/jair/article/download/15320/27076/40841.

Amit Moscovich and Saharon Rosset. On the cross-validation bias due to unsupervised preprocessing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1474–1502, 2022. doi: 10.1111/rssb.12537.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

Simon Roth. Which leakage types matter? A quantitative landscape across 2,047 benchmark datasets. *ArXiv preprint*, arXiv:2604.04199, 2026. doi: 10.48550/arXiv.2604.04199. URL https://arxiv.org/abs/2604.04199.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *ArXiv preprint*, arXiv:1707.06347, 2017. doi: 10.48550/arXiv.1707.06347. URL https://arxiv.org/abs/1707.06347.

Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000. doi: 10.1111/1468-0262.00152.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '19, volume 33, pp. 1658–1665, Honolulu, HI, USA, 2019. AAAI Press. doi: 10.1609/aaai.v33i01.33011658.

Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, pp. 1–8, Virtual, 2020. Association for Computing Machinery. doi: 10.1145/3383455.3422540. URL https://doi.org/10.1145/3383455.3422540.

Dong Yin, Takeshi Miki, Vladislav Lesnichenko, and Vasyl Gural. Implementation risk in portfolio backtesting: A previously unquantified source of error. *ArXiv preprint*, arXiv:2603.20319, 2026. doi: 10.48550/arXiv.2603.20319. URL https://arxiv.org/abs/2603.20319.

GUIDE TO SUPPORTING EVIDENCE

Use this guide to trace the main claims to their definitions, complete comparisons and uncertainty calculations. For the central empirical results, start with Appendices F and N; Appendix E defines the audit inputs and Appendix H specifies execution.

Claim-to-evidence index.

Question	Evidence to read	Role
Does decision information improve audit value?	Table 1; Appendix F.3 (p. 65) for paired inference, and Tables 42–44 for complete results.	Frozen primary test; secondary breakdowns
What do Score-only and Decision observe, and how are they fitted?	Appendix E.2 (p. 60) for cases and forward fits; Appendix E.3 (p. 61) for every descriptor.	Main audit definitions
What does learning add beyond Replacement?	Appendix N.1 (p. 85), Table 63: paired gains and the cases selected by only one priority.	Post hoc comparison
Why does Allocation capture additional impact?	Figure 6; Appendices N.2–N.4 (pp. 86–88) for policy transfer, impact decomposition and the structural relation.	Secondary evaluation; exploratory Allocation
Can score change misorder economic deviation?	Figure 4; Appendix D (p. 54) for matched exposure, all conditions and noise controls.	Controlled mechanism evidence
How do protocol and learning entry affect the backtest?	Appendix B (p. 16) for the five protocol axes; Appendix C (p. 27) for feature, training–scoring and exposure grids.	Complete mechanism comparisons
How do execution and account state enter the target?	Appendix H (p. 72) for self-financing execution; Appendix I (p. 74) for state controls and sensitivity.	Accounting and state controls
How does the audit relate to prior research?	Appendix L (p. 83), Tables 61–62: experimental objects, objectives and supported decisions.	Research positioning

Reading the evidence. Table headers use the short metric names introduced with each question. Brackets accompanying estimates give 95% CIs; the text specifies the paired comparison and temporal procedure. Arrows mark the preferred metric direction. Bold, darker cells identify the best displayed comparable non-oracle result, including ties; the main table also emphasizes the primary contrast. Neither bold nor color denotes statistical significance. Units remain in headers or captions so that a result can be read without reconstructing its scale.

Representations across audit designs. The main heterogeneous audit uses Score-only (S17), Decision (D28) and Decision+State (55 descriptors), defined in Appendix E.3. Appendix G reports the separate equal-amplitude control with local 22-descriptor Decision and 49-descriptor State variants.

Data and study context. Appendix A (p. 13) specifies data and forecasting models; Appendix J (p. 79) gives public-implementation cases; Appendix K (p. 81) records the study sequence and pre-outcome specifications; Appendix M (p. 85) defines the evidence scope and operational use.

A DATA, FEATURES, AND FORECASTING MODELS

Historical sample and clocks. The BaoStock extract was downloaded on July 25, 2026 and covers 2010–2025.[2] Historical constituent membership determines the CSI 300 and CSI 500 universes on each date within this archived snapshot. Both are constituent sets in one Chinese equity market; the

[2]BaoStock data platform: http://baostock.com.

snapshot does not contain historical vendor vintages. Raw daily bars and status fields determine whether trades can execute. Backward/post-adjusted (HFQ) prices provide return features, labels and valuations in adjusted units, whose accounting is specified in Appendix H.

Historical membership identifies the constituent set effective on each date; a vendor vintage also records the values available at that historical time. Here, clean denotes the reference pipeline without the specified intervention. Shared data and execution rules let paired comparisons isolate that intervention within the archived environment. Historical revisions can interact with features, fitted scores and portfolio choices, so evaluating historical as-of performance would additionally require vendor-vintage evidence.

At close t, the model predicts the five-session return $y_t = O^{\mathrm{HFQ}}_{t+6}/O^{\mathrm{HFQ}}_{t+1} - 1$. The clean ranking selects $K_t = \max\{1, \lceil 0.10N_t \rceil\}$ of N_t eligible stocks, using a fixed outcome-independent hash to break ties. The resulting sleeve enters at open $t+1$ and is scheduled to exit at open $t+6$. In 2022–2024, predictive diagnostics use 726 dates, new-sleeve decisions use 685 dates through November 4, 2024, and portfolio returns use all 726 execution/valuation dates. The final 41 sessions retain exits and failed-sale retries. Late-2024 labels and future-feature endpoints use support through January 31, 2025. The separate 2025 confirmation accounts are specified in Appendix F. The original development paths use 1,461 economic dates and 1,420 sleeve-opening dates in 2016–2021, with the analogous November 4, 2021 decision cutoff.

The clean samples on sleeve-opening dates contain 415,526 CSI 300 asset–dates in development and 204,397 CSI 300 / 339,347 CSI 500 asset–dates in 2022–2024. Their distinct-security counts are 533, 398 and 772, respectively; constituent changes explain why these exceed the nominal index sizes. Mean cross-sectional sizes are 292.62, 298.39 and 495.40. Both forecasting models use identical clean keys.

Features and preprocessing. Table 2 defines all 19 inputs. Its superscripts distinguish current-vintage backward/post-adjusted (HFQ) from raw fields. Feature windows end at t and are computed for each security before restricting the sample to index members. Features are not cross-sectionally normalized. Missing market dates are neither filled nor compressed; the longest feature requires 61 consecutive market dates. The eight-feature *State* representation contains the three volatility, two range and three volume features. It excludes signed returns, close-to-moving-average ratios, the intraday return, log dollar volume and the up-day fraction. These market-state features describe prices and trading activity; the portfolio-state descriptors used later describe holdings and cash.

Table 2: **The features separate price direction from market state.** This distinction allows the later experiments to vary what future information a model receives.

Feature	Definition
ret_k, $k \in \{1, 5, 10, 20, 60\}$	$\log(c^{\mathrm{HFQ}}_t / c^{\mathrm{HFQ}}_{t-k})$.
vol_k, $k \in \{5, 20, 60\}$	Sample standard deviation (degrees of freedom 1) of ret_1 over $t-k+1{:}t$.
$\mathrm{hl_range}_1$	$\log(h^{\mathrm{raw}}_t / l^{\mathrm{raw}}_t)$.
hl_range_5_mean	Five-session mean of $\mathrm{hl_range}_1$.
$\mathrm{oc_gap}_1$	$\log(c^{\mathrm{raw}}_t / o^{\mathrm{raw}}_t)$.
close_ma$k$_ratio, $k \in \{5, 20, 60\}$	c^{HFQ}_t divided by its k-session moving average.
log_volume	$\log(1 + v^{\mathrm{raw}}_t)$.
$\mathrm{vol_ratio}_k$, $k \in \{5, 20\}$	v^{raw}_t divided by its k-session moving average.
log_dollar_volume	$\log(1 + a^{\mathrm{raw}}_t)$, where a is traded amount.
up_days_ratio_20	Mean of $\mathbf{1}\{\mathrm{ret}_1 > 0\}$ over 20 sessions; a zero return maps to zero.

Each annual fold has five training years, one validation year and one test year, with training labels crossing the fold boundary purged. Each fold fits its own model. Clean preprocessing winsorizes each feature at the training-sample 0.5th and 99.5th percentiles, using linearly interpolated quantiles, then standardizes by its winsorized training mean and population standard deviation. A zero standard deviation is replaced by one; missing values are not imputed.

Compact setting labels abbreviate LightGBM as LGBM. Configurations are selected once on clean CSI 300 development data and held fixed across interventions and CSI 500. We maximize the equal-weight mean daily cross-sectional Spearman correlation with the label over 1,426 validation dates and 412,640 asset–dates. Scores within 10^{-4} of the best are tied. Ridge ties favor larger α; LightGBM ties favor fewer leaves, shallower depth, larger minimum child count, larger ℓ_2 penalty, then fewer trees. All five Ridge candidates tie by this rule (selected validation RankIC -0.00959); LightGBM has a unique selected candidate (RankIC 0.04023).

The validation RankIC values above describe development configuration selection. The all-feature clean annualized net returns of -8.57%, -5.01%, -0.49% and -1.55% concern the four model–universe pairs in 2022–2024; they are distinct from the 2025 audit evaluation. Earlier CSI 300/LightGBM clean performance is +14.96% in 2016–2021 (Table 9). Audit impact measures how an error changes the evidence about a fixed model, including when an intervention makes an unprofitable reference appear profitable. The comparisons vary interventions within each selected forecasting setup; forecasting quality is not a separate experimental factor.

Table 3: **Forecasting choices are fixed before the interventions.** Both models retain the configurations selected on clean development data.

Model	Search space	Fixed configuration
Ridge	$\alpha \in \{0.01, 0.1, 1, 10, 100\}$.	$\alpha = 100$; intercept; LSQR solver; tolerance 10^{-8}; maximum 10,000 iterations.
LightGBM	Cartesian product of (depth, leaves) in $\{(4, 15), (6, 31), (8, 63)\}$, minimum child count in $\{20, 100\}$, $\lambda_2 \in \{0, 1\}$, and (learning rate, trees) in $\{(0.05, 200), (0.03, 400)\}$.	GBDT squared-error objective; depth 8; 63 leaves; minimum child count 100; $\lambda_2 = 0$; learning rate 0.03; 400 trees; full row and feature sampling; deterministic seed 20260810; no early stopping.

Model fitting uses scikit-learn 1.9.0 and LightGBM 4.6.0. Additional LightGBM settings are row/feature sampling fractions one, sampling frequency zero, ℓ_1 penalty zero, maximum bin count 255, minimum split gain zero, deterministic column-wise fitting, feature pre-filtering off, and one worker; all model/data/sampling seeds equal 20260810. The forecasting Ridge penalty in Table 3 is distinct from the fixed penalty 10 used for the audit regressions.

A.1 Study roles and outcome definitions

The study map and account identities collect the common definitions used by the protocol and mechanism comparisons.

Table 4: **The evidence moves from mechanisms to audit value.** Each study answers a different question along the path from information access to economic consequences.

Study	Scope	Question
Protocol severity	Five protocol axes; 2022–2024	Do larger protocol departures or greater exposure produce larger return effects?
Information and entry	Six feature groups; four leads; training/scoring 2×2	How do information content and exposure during training or prediction affect returns?
Sparse exposure	Date, stock–date and feature doses; 2022–2024	Can limited exposure change selected stocks and portfolio returns?
Exposure location	Forty conditions; three locations; twenty seeds	Can smaller score changes produce larger economic effects?
Initial holdings	Fixed selections; historical versus all-cash state; 2019–2021	What changes economically when only the starting account changes?
Heterogeneous audit	Three channels, three doses; 2023–2024	Do selection changes help prioritize economic checks under the same budget?
Temporal audit test	Frozen priorities and new 2025 forecasts	Does the specified audit improvement hold in a new period?

The main text distinguishes predictive accuracy (RankIC), changed selections (Rep.), and economic consequences. The appendix follows the same sequence. Score change compares predictions without labels; signed return change (ΔNet) shows inflation or depression, while full-path deviation (Dev., written $A(a)$ below) measures the distortion that remains when opposing daily effects cannot cancel:

$$\Delta\mathrm{Net}(a) = 100 \times 252 \ \mathrm{mean}_t(r_t^a - r_t^0) \tag{6}$$

$$A(a) = \frac{10^4}{T} \sum_{t=1}^{T} |r_t^a - r_t^0|. \tag{7}$$

For reference levels, annualized arithmetic net return (Net) is $100 \times 252 \ \mathrm{mean}_t \, r_t$, in percent. Its paired change ΔNet is in pp. Predictive change (ΔRankIC) subtracts the clean mean daily RankIC; the accompanying annualized daily Sharpe change (ΔSharpe) records the change in risk-adjusted performance. These signed diagnostics describe how the evidence moves, not whether an intervention is acceptable. Full-path deviation $A(a)$ is in bp/day over T dates; the one-decision audit uses its six-session counterpart. With cash c_{t-1}, adjusted lot units $u_{\ell,t-1}$, mark $\widetilde{P}_{\ell,t}$, external sales S_t^{ext}, purchases B_t^{ext} and per-side fee $\kappa = 0.001$, the account satisfies

$$\begin{aligned} V_t^- &= c_{t-1} + \sum_{\ell} u_{\ell,t-1} \widetilde{P}_{\ell,t}, & c_t^s &= c_{t-1} + (1-\kappa) S_t^{\mathrm{ext}}, \\ c_t &= c_t^s - (1+\kappa) B_t^{\mathrm{ext}} \geq 0, & V_t^+ &= V_t^- - \kappa (S_t^{\mathrm{ext}} + B_t^{\mathrm{ext}}). \end{aligned} \tag{8}$$

For remaining purchase intentions $D_t > 0$, the fill scale is $\min\{1, c_t^s / [(1+\kappa) D_t]\}$. Daily net return is $r_t = V_t^+ / V_{t-1}^+ - 1$. Internal renewal changes the exit schedule without an external trade.

B Complete protocol-severity evidence

This appendix reports exposure, uncertainty and annual results for the 84 comparisons in Section 4.1. Each of the four model–universe settings uses its own clean reference and the common adjusted-unit, self-financing rule in Appendix H. The five axes use different information units; we compare their paired economic effects without treating dose as a common scale. Each comparison changes one assumption with fixed configurations and economic rules. The first level of each grid is clean; the separate OBS learning-entry design is given in Section 4.2.

B.1 EXPOSURE DEFINITIONS AND COMPLETE-PATH COMPARISONS

Table 5: **Five interventions probe different stages of the backtest.** Each changes one information or execution assumption while retaining the remaining protocol.

Axis	Levels	Change from clean
Observation (OBS)	$L \in \{0, 1, 2, 5, 10\}$ market sessions	At $L = 0$, use the 19-feature vector formed through close t. At $L > 0$, shift every trailing-feature endpoint for the same asset to $t + L$ while retaining its window length; if the future vector is unavailable or non-finite, use the entire date-t vector.
Preprocessing fit (FIT)	$q \in \{0, 1, 5, 10, 25, 100\}\%$	At $q = 0$, estimate winsorization and standardization from training rows only. Otherwise also use feature rows from the earliest stated percentage of the ordered validation-and-test decision calendar; no validation or test label enters model fitting.
Universe (UNIV)	$\Delta \in \{0, 1, 5, 20, 60, 240\}$ market sessions	Replace only date-d membership M_d by $M_{d+\Delta}$. Training and economic candidate sets may change, while availability, feature sufficiency, status, and tradability remain at date d.
Tradability (TRAD)	$q \in \{0, 1, 5, 10, 25, 100\}\%$	On the q fraction of decision dates selected by each of 20 fixed nested masks, reveal next-open fillability, skip an unfillable selected name, and take the next-ranked fillable name. Each mask is simulated to its endpoint before the 20 endpoints are averaged equally.
Execution (EXEC)	next open; same close; same open	Keep features, scores, and labels fixed. The clean category enters at open $(t+1)$ and exits at open $(t+6)$; the alternatives enter at close t and exit at close $(t + 5)$, or enter at open t and exit at open $(t + 5)$. Raw-bar eligibility and the HFQ wealth mark use the corresponding open or close throughout each path.

Table 5 defines the coordinates and interventions. OBS advances all 19 feature endpoints in training and scoring. FIT changes only the rows used to estimate preprocessing. Its ordered denominator is the entire validation-plus-test decision calendar, including dates with no candidate rows. Only feature values enter the estimate; validation and test labels remain excluded. The nominal fraction of dates can therefore differ from the realized fraction of admitted feature rows.

UNIV changes training membership and the economic candidate set to their future values, keeping features, availability, status and tradability contemporaneous. Predictive comparisons retain clean stock–date keys: excluded clean members receive shadow scores but cannot be traded. We report membership changes separately from model-specific changes in clean selected names: the altered training sample can change rankings even among names shared by the two universes.

TRAD uses twenty outcome-independent permutations of the common sleeve-opening calendar. Within each permutation, the mask at fraction q takes its first $\lfloor qD \rfloor$ dates; the full calendar is used at $q = 100\%$. Masks are nested within a family, and every mask is simulated as a complete portfolio path. In 2022–2024, the five positive doses select 6, 34, 68, 171 and 685 of the 685 sleeve-opening dates. The 2016–2021 development anchor at 5% selects 71 of 1,420 dates. TRAD endpoints average the twenty complete-path endpoints equally; its temporal uncertainty uses the corresponding within-date average of the paired daily return series.

Realized exposure is reported in each intervention's native denominator: complete future vectors used for OBS, feature rows admitted to estimation for FIT, membership and clean-selection changes for UNIV, clean selected names replaced for TRAD, and dates with changed net returns for EXEC. The UNIV and TRAD selection denominators count intended names, not executed fills. In 2022–2024, predictive and observation diagnostics use 726 dates; new sleeves begin on 685 dates, and economic valuation continues through the remaining 41 sessions.

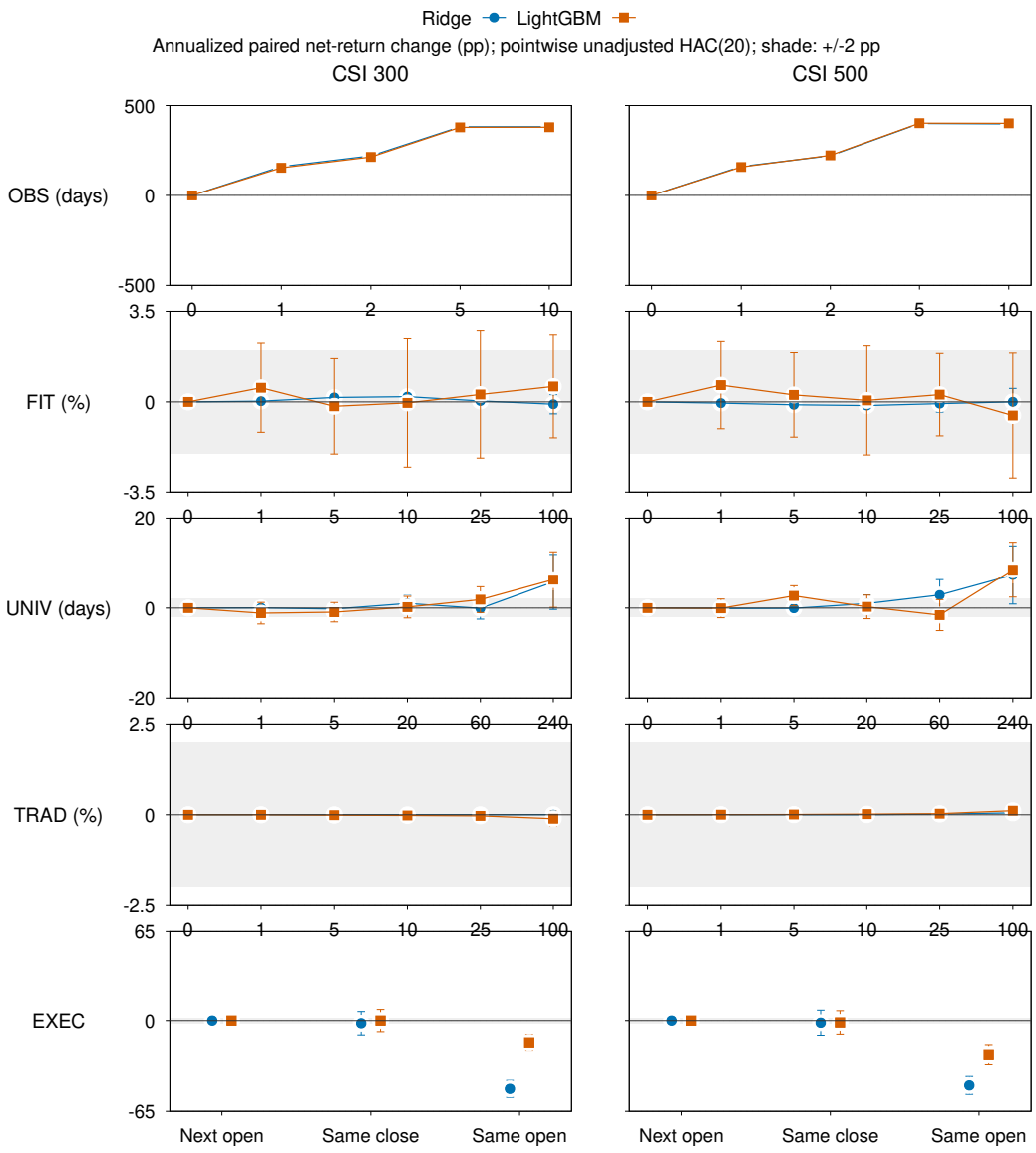


Figure 7: **The complete five-axis severity comparison.** All 84 intervention-minus-clean net-return effects in 2022–2024, in annualized percentage points. OBS exposes all 19 features in both training and scoring. Lines connect evaluated numerical coordinates; execution clocks are categorical. Whiskers are descriptive, pointwise lag-20 Newey–West 95% intervals, without simultaneous coverage. Shading marks the ± 2-point practical reference region. Row-specific vertical scales keep small and negative responses visible.

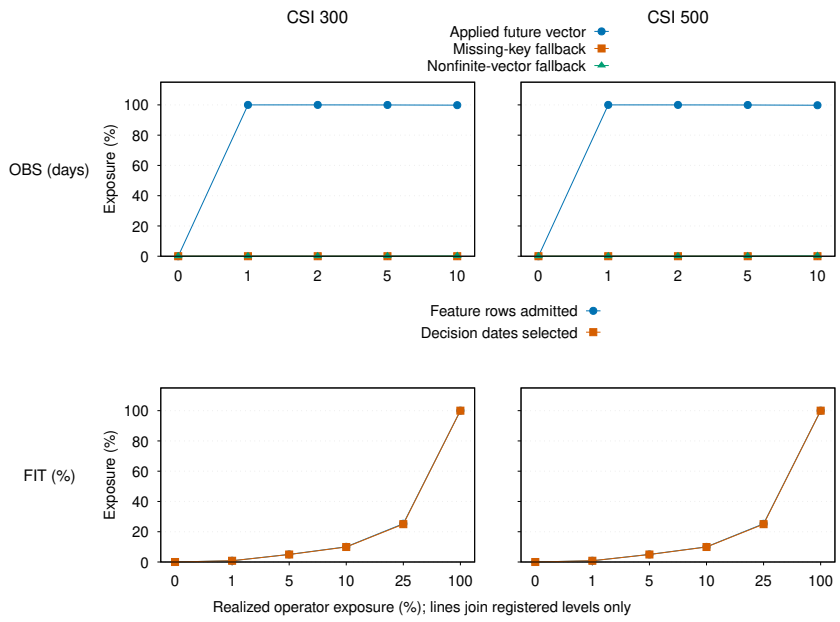


Figure 8: Realized exposure for observation timing and preprocessing scope in 2022–2024. OBS records complete future-vector use and whole-vector fallback; FIT compares selected decision dates with admitted feature rows. Lines connect evaluated coordinates only. Universes share a vertical scale within a row, while interventions retain their own exposure definitions.

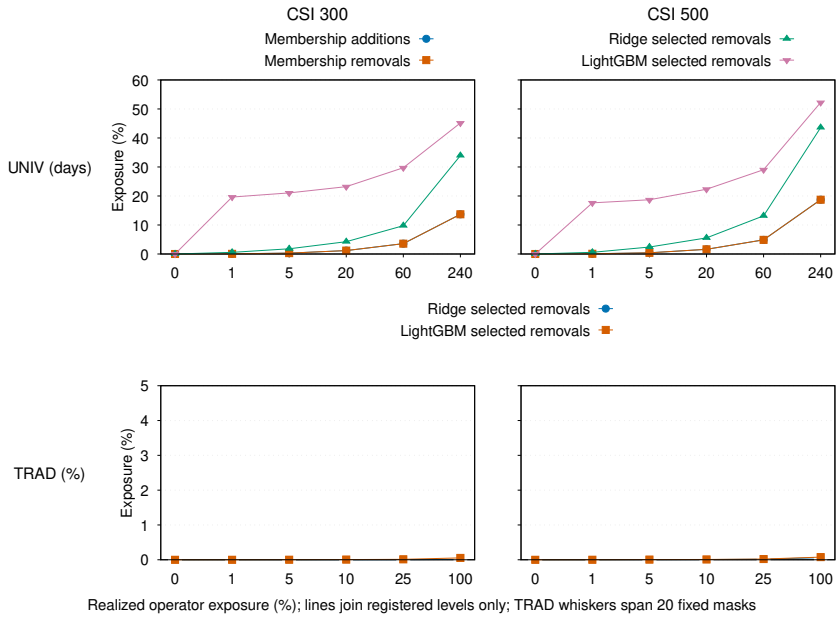


Figure 9: Realized exposure for future membership and future fillability in 2022–2024. Membership changes and model-specific clean-selection replacements are distinguished. For TRAD, curves show equal-weight means across twenty complete paths; ranges show mask-to-mask variation, not confidence intervals. Lines connect evaluated coordinates only; universes share a vertical scale within each row.

B.2 EVERY POOLED EFFECT AND THE PRACTICAL REFERENCE REGION

Table 6 reports all 84 paired effects, the four clean references, realized exposure, Sharpe change and RankIC change. Negative, negligible and nonmonotone effects are retained alongside the large OBS effects. The 95% CIs use lag-20 Bartlett–Newey–West standard errors and a normal 1.96 multiplier. They describe temporal variability in this explored period, without simultaneous or selection-adjusted coverage. Rows are paired to their own universe–model clean path. They are distinct from the four-block rule for the frozen 2025 audit confirmation.

Table 6: **Information access is not a common scale of economic severity.** OBS produces large positive effects, same-open EXEC produces large negative effects, and FIT and TRAD remain much smaller.

Universe / model	Intervention	ΔNet (pp)	CI	Exposure (%)	ΔSharpe	ΔRankIC
CSI 300 Ridge	Clean	+0.0000	–	0.00000	+0.00000	+0.000000
	OBS $L = 1$	+160.2273	[+143.5873, +176.8673]	99.99706	+6.13582	+0.396785
	OBS $L = 2$	+220.0406	[+198.6450, +241.4363]	99.99412	+8.06901	+0.519983
	OBS $L = 5$	+381.8849	[+348.7682, +415.0016]	99.95255	+12.44938	+0.976492
	OBS $L = 10$	+382.1565	[+348.6928, +415.6202]	99.82692	+11.83217	+0.972162
	FIT $q = 1\%$	+0.0284	[-0.0461, +0.1030]	0.83581	+0.00102	+0.000003
	FIT $q = 5\%$	+0.1729	[+0.0055, +0.3403]	5.01320	+0.00619	-0.000004
	FIT $q = 10\%$	+0.2039	[+0.0090, +0.3989]	10.02874	+0.00741	-0.000011
	FIT $q = 25\%$	+0.0375	[-0.2105, +0.2855]	25.29714	+0.00145	+0.000014
	FIT $q = 100\%$	-0.0896	[-0.4636, +0.2843]	100.00000	-0.00355	+0.000138
	UNIV $\delta = 1$	+0.0442	[-0.2473, +0.3357]	0.56448	+0.00159	+0.000039
	UNIV $\delta = 5$	-0.2132	[-0.8803, +0.4539]	1.79562	-0.00750	+0.000296
	UNIV $\delta = 20$	+1.0288	[-0.7547, +2.8124]	4.23844	+0.03815	-0.000245
	UNIV $\delta = 60$	-0.0700	[-2.4623, +2.3222]	9.76156	-0.00382	-0.000078
	UNIV $\delta = 240$	+5.8204	[-0.2960, +11.9368]	33.92214	+0.20694	-0.014548
	TRAD $q = 1\%$	+0.0000	[+0.0000, +0.0000]	0.00000	+0.00000	+0.000000
	TRAD $q = 5\%$	+0.0003	[-0.0003, +0.0009]	0.00024	+0.00001	+0.000000
	TRAD $q = 10\%$	+0.0006	[-0.0012, +0.0024]	0.00122	+0.00002	+0.000000
	TRAD $q = 25\%$	+0.0016	[-0.0033, +0.0065]	0.00341	+0.00006	+0.000000
	TRAD $q = 100\%$	+0.0034	[-0.0096, +0.0164]	0.00973	+0.00013	+0.000000
CSI 300 Ridge	EXEC same close	-1.9239	[-10.3723, +6.5244]	95.17906	-0.10154	+0.000000
	EXEC same open	-48.7791	[-54.9000, -42.6582]	95.17906	-1.70486	+0.000000
CSI 300 LightGBM	Clean	+0.0000	–	0.00000	+0.00000	+0.000000
	OBS $L = 1$	+153.5336	[+137.6869, +169.3803]	99.99706	+6.04226	+0.362227
	OBS $L = 2$	+213.8193	[+192.9652, +234.6733]	99.99412	+8.23292	+0.493565
	OBS $L = 5$	+378.8954	[+344.3697, +413.4211]	99.95255	+12.64925	+0.957615
	OBS $L = 10$	+379.4673	[+343.8098, +415.1249]	99.82692	+11.71779	+0.952938
	FIT $q = 1\%$	+0.5505	[-1.1784, +2.2794]	0.83581	+0.02186	+0.001917
	FIT $q = 5\%$	-0.1694	[-2.0227, +1.6840]	5.01320	-0.00745	+0.000151
	FIT $q = 10\%$	-0.0390	[-2.5325, +2.4545]	10.02874	-0.00036	+0.002007
	FIT $q = 25\%$	+0.2894	[-2.1804, +2.7592]	25.29714	+0.01133	+0.000790
	FIT $q = 100\%$	+0.6011	[-1.3951, +2.5972]	100.00000	+0.02231	-0.000270
	UNIV $\delta = 1$	-1.1396	[-3.5406, +1.2615]	19.63504	-0.04466	-0.000570
	UNIV $\delta = 5$	-0.9210	[-3.0858, +1.2439]	21.06083	-0.03432	-0.000712
	UNIV $\delta = 20$	+0.2102	[-2.1981, +2.6185]	23.21168	+0.00884	+0.000076
	UNIV $\delta = 60$	+1.8781	[-0.9925, +4.7487]	29.73236	+0.07318	+0.000989
	UNIV $\delta = 240$	+6.3382	[+0.1556, +12.5209]	45.16788	+0.24097	-0.005798
	TRAD $q = 1\%$	-0.0003	[-0.0008, +0.0003]	0.00024	-0.00001	+0.000000
	TRAD $q = 5\%$	-0.0110	[-0.0284, +0.0064]	0.00219	-0.00041	+0.000000
	TRAD $q = 10\%$	-0.0191	[-0.0528, +0.0145]	0.00560	-0.00070	+0.000000
	TRAD $q = 25\%$	-0.0300	[-0.0917, +0.0316]	0.01241	-0.00109	+0.000000
	TRAD $q = 100\%$	-0.1093	[-0.3173, +0.0988]	0.05353	-0.00392	+0.000000

Continued on next page

Universe / model	Intervention	ΔNet (pp)	CI	Exposure (%)	ΔSharpe	ΔRankIC
CSI 300 LightGBM	EXEC same close	+0.0104	[-7.9943, +8.0151]	95.17906	-0.01623	+0.000000
	EXEC same open	-15.7742	[-21.4143, -10.1340]	95.17906	-0.59390	+0.000000
CSI 500 Ridge	Clean	+0.0000	–	0.00000	+0.00000	+0.000000
	OBS $L = 1$	+160.9565	[+143.0934, +178.8195]	99.99500	+5.88741	+0.386847
	OBS $L = 2$	+220.8489	[+197.7849, +243.9129]	99.98989	+8.04685	+0.498602
	OBS $L = 5$	+400.2330	[+365.0631, +435.4028]	99.93549	+12.83596	+0.955153
	OBS $L = 10$	+397.6511	[+362.9399, +432.3623]	99.76912	+12.21050	+0.950867
	FIT $q = 1\%$	-0.0471	[-0.1501, +0.0560]	0.83518	-0.00156	+0.000004
	FIT $q = 5\%$	-0.1118	[-0.2701, +0.0466]	5.00181	-0.00369	+0.000007
	FIT $q = 10\%$	-0.1412	[-0.3939, +0.1115]	9.99602	-0.00466	-0.000001
	FIT $q = 25\%$	-0.0672	[-0.3953, +0.2608]	25.25734	-0.00223	+0.000103
	FIT $q = 100\%$	+0.0039	[-0.5179, +0.5257]	100.00000	+0.00013	+0.000208
	UNIV $\delta = 1$	-0.0902	[-0.3772, +0.1968]	0.57327	-0.00295	+0.000105
	UNIV $\delta = 5$	-0.0591	[-0.7543, +0.6360]	2.39836	-0.00191	-0.000584
	UNIV $\delta = 20$	+0.9899	[-0.9257, +2.9055]	5.54841	+0.03244	-0.001415
	UNIV $\delta = 60$	+2.9061	[-0.5347, +6.3469]	13.15297	+0.09472	-0.003826
	UNIV $\delta = 240$	+7.3681	[+0.9049, +13.8314]	43.57999	+0.24569	+0.004389
	TRAD $q = 1\%$	+0.0006	[-0.0009, +0.0022]	0.00117	+0.00002	+0.000000
	TRAD $q = 5\%$	+0.0016	[-0.0063, +0.0094]	0.00366	+0.00005	+0.000000
	TRAD $q = 10\%$	+0.0019	[-0.0119, +0.0156]	0.00614	+0.00006	+0.000000
	TRAD $q = 25\%$	+0.0195	[-0.0245, +0.0636]	0.01550	+0.00065	+0.000000
	TRAD $q = 100\%$	+0.0532	[-0.1101, +0.2164]	0.06142	+0.00176	+0.000000
	EXEC same close	-1.5328	[-10.5893, +7.5238]	95.17906	-0.05859	+0.000000
	EXEC same open	-46.3033	[-52.7588, -39.8478]	95.17906	-1.48991	+0.000000
CSI 500 LightGBM	Clean	+0.0000	–	0.00000	+0.00000	+0.000000
	OBS $L = 1$	+158.0114	[+142.3869, +173.6359]	99.99500	+5.92316	+0.373675
	OBS $L = 2$	+222.9949	[+201.6588, +244.3310]	99.98989	+8.16461	+0.494715
	OBS $L = 5$	+401.9624	[+366.2366, +437.6882]	99.93549	+13.21502	+0.956565
	OBS $L = 10$	+401.2135	[+364.8977, +437.5293]	99.76912	+12.39994	+0.951703
	FIT $q = 1\%$	+0.6514	[-1.0418, +2.3446]	0.83518	+0.02396	+0.001499
	FIT $q = 5\%$	+0.2724	[-1.3665, +1.9112]	5.00181	+0.01000	+0.000951
	FIT $q = 10\%$	+0.0591	[-2.0584, +2.1767]	9.99602	+0.00216	+0.000198
	FIT $q = 25\%$	+0.2823	[-1.3173, +1.8820]	25.25734	+0.01041	-0.000168
	FIT $q = 100\%$	-0.5283	[-2.9541, +1.8975]	100.00000	-0.01903	+0.001522
	UNIV $\delta = 1$	-0.0628	[-2.1525, +2.0268]	17.62796	-0.00234	-0.000239
	UNIV $\delta = 5$	+2.7225	[+0.4696, +4.9754]	18.69553	+0.09961	-0.000245
	UNIV $\delta = 20$	+0.2670	[-2.3698, +2.9037]	22.32524	+0.00969	-0.000887
	UNIV $\delta = 60$	-1.5629	[-5.0172, +1.8913]	29.05235	-0.05775	-0.005361
	UNIV $\delta = 240$	+8.5468	[+2.4566, +14.6370]	52.23750	+0.31131	+0.005114
	TRAD $q = 1\%$	+0.0031	[-0.0001, +0.0062]	0.00117	+0.00011	+0.000000
	TRAD $q = 5\%$	+0.0081	[-0.0033, +0.0195]	0.00512	+0.00030	+0.000000
	TRAD $q = 10\%$	+0.0183	[-0.0076, +0.0441]	0.00848	+0.00068	+0.000000
	TRAD $q = 25\%$	+0.0309	[-0.0100, +0.0719]	0.02062	+0.00116	+0.000000
	TRAD $q = 100\%$	+0.1140	[-0.0501, +0.2781]	0.07605	+0.00427	+0.000000
	EXEC same close	-1.3362	[-9.7945, +7.1222]	95.17906	-0.06188	+0.000000
	EXEC same open	-24.4096	[-31.4239, -17.3953]	95.17906	-0.87409	+0.000000

TRAD and EXEC leave RankIC unchanged because they reuse clean scores. Their different return responses show why predictive diagnostics alone miss execution effects.

Crossing the practical reference region does not establish a monotone dose threshold. Table 7 records the first evaluated point outside the prespecified $[-2, 2]$-point region and any later re-entry, without interpolating unmeasured severities. For CSI 500 LightGBM UNIV, the five-session point lies outside, the twenty- and sixty-session points lie inside, and the 240-session point lies outside again. Execution clocks are reported separately as categories.

Only evaluated numerical levels enter this summary. A dash marks the absence of an observed exit or subsequent re-entry; no values between tested levels are interpolated.

Table 7: **Crossing the practical reference region need not be permanent.** Future-membership effects can leave and later re-enter the region as the requested lead increases.

Universe	Model	Axis	First exit	Prior level	ΔNet	Re-entry
300	Ridge	OBS	$L = 1$	Clean	+160.2273	–
		FIT	–	–	–	–
		UNIV	$\delta = 240$	$\delta = 60$	+5.8204	–
		TRAD	–	–	–	–
	LGBM	OBS	$L = 1$	Clean	+153.5336	–
		FIT	–	–	–	–
		UNIV	$\delta = 240$	$\delta = 60$	+6.3382	–
		TRAD	–	–	–	–
500	Ridge	OBS	$L = 1$	Clean	+160.9565	–
		FIT	–	–	–	–
		UNIV	$\delta = 60$	$\delta = 20$	+2.9061	–
		TRAD	–	–	–	–
	LGBM	OBS	$L = 1$	Clean	+158.0114	–
		FIT	–	–	–	–
		UNIV	$\delta = 5$	$\delta = 1$	+2.7225	$\delta = 20$
		TRAD	–	–	–	–

The CSI 500/LightGBM membership response leaves, re-enters, and then leaves the region again; a larger lead is not a monotone severity scale.

B.3 Annual decompositions of the full later-period grid

OBS and same-open execution retain opposite directions across the observed years. Figures 10 and 11 show the four numerical axes; Table 8 includes all 84 conditions, both execution alternatives and the four clean references.

Annual values decompose the paired daily effects within each year; TRAD averages the twenty mask-specific annual endpoints. The clean rows remain zero references. All 48 annual all-feature, joint-exposure OBS effects are positive; all twelve annual same-open effects are negative. These annual summaries describe the shared market paths.

Table 8: **The largest protocol effects retain their directions across years.** All-feature joint OBS raises returns and same-open EXEC lowers them in each year. Annual ΔNet (pp).

Universe / model	Intervention	2022	2023	2024
CSI 300 Ridge	Clean	+0.0000	+0.0000	+0.0000
	OBS $L = 1$	+173.7721	+160.0635	+146.8462
	OBS $L = 2$	+241.6788	+220.8891	+197.5540
	OBS $L = 5$	+423.0021	+375.7021	+346.9504
	OBS $L = 10$	+422.6434	+375.3442	+348.4820
CSI 300 Ridge	FIT $q = 1\%$	+0.0447	+0.0406	+0.0000
	FIT $q = 5\%$	+0.5432	-0.0166	-0.0079
	FIT $q = 10\%$	+0.4700	+0.1071	+0.0347
	FIT $q = 25\%$	+0.2453	-0.0416	-0.0912
	FIT $q = 100\%$	-0.0468	-0.1964	-0.0258
CSI 300 Ridge	UNIV $\delta = 1$	+0.3334	-0.2050	+0.0043
	UNIV $\delta = 5$	-0.3160	-0.2453	-0.0782
	UNIV $\delta = 20$	-0.0242	+0.6902	+2.4205
	UNIV $\delta = 60$	+1.5251	-1.5536	-0.1816
	UNIV $\delta = 240$	+4.4250	+3.9003	+9.1359

Continued on next page

Universe / model	Intervention	2022	2023	2024
CSI 300 Ridge	TRAD $q = 1\%$	+0.0000	+0.0000	+0.0000
	TRAD $q = 5\%$	+0.0000	+0.0000	+0.0009
	TRAD $q = 10\%$	+0.0000	-0.0007	+0.0026
	TRAD $q = 25\%$	+0.0000	-0.0022	+0.0070
	TRAD $q = 100\%$	+0.0000	-0.0073	+0.0174
CSI 300 Ridge	EXEC same close	-1.6978	+0.2745	-4.3485
	EXEC same open	-65.6037	-42.6072	-38.1264
CSI 300 LightGBM	Clean	+0.0000	+0.0000	+0.0000
	OBS $L = 1$	+166.3759	+151.1053	+143.1196
	OBS $L = 2$	+237.0493	+213.1477	+191.2609
	OBS $L = 5$	+422.4420	+369.0777	+345.1666
	OBS $L = 10$	+422.3543	+368.8099	+347.2378
CSI 300 LightGBM	FIT $q = 1\%$	-0.9335	+2.2061	+0.3789
	FIT $q = 5\%$	+0.4482	-1.5332	+0.5769
	FIT $q = 10\%$	-2.1857	-0.0675	+2.1361
	FIT $q = 25\%$	-1.3633	+1.0465	+1.1851
	FIT $q = 100\%$	-0.2029	+2.8324	-0.8264
CSI 300 LightGBM	UNIV $\delta = 1$	-3.2445	-2.0470	+1.8728
	UNIV $\delta = 5$	-4.1464	-0.0515	+1.4350
	UNIV $\delta = 20$	-1.6866	-1.9076	+4.2247
	UNIV $\delta = 60$	-2.2857	+2.5757	+5.3444
	UNIV $\delta = 240$	-0.0823	+7.4076	+11.6894
CSI 300 LightGBM	TRAD $q = 1\%$	-0.0008	+0.0000	+0.0000
	TRAD $q = 5\%$	-0.0264	+0.0063	-0.0129
	TRAD $q = 10\%$	-0.0403	+0.0205	-0.0376
	TRAD $q = 25\%$	-0.0655	+0.0492	-0.0737
	TRAD $q = 100\%$	-0.2617	+0.1637	-0.2298
CSI 300 LightGBM	EXEC same close	+0.0991	+1.6956	-1.7636
	EXEC same open	-31.0898	-6.7671	-9.4656
CSI 500 Ridge	Clean	+0.0000	+0.0000	+0.0000
	OBS $L = 1$	+174.1075	+156.7036	+152.0583
	OBS $L = 2$	+245.1266	+213.4544	+203.9658
	OBS $L = 5$	+447.7755	+384.0772	+368.8462
	OBS $L = 10$	+442.6375	+383.2124	+367.1034
CSI 500 Ridge	FIT $q = 1\%$	+0.0061	-0.0176	-0.1296
	FIT $q = 5\%$	-0.0789	-0.0758	-0.1806
	FIT $q = 10\%$	-0.3302	-0.0244	-0.0690
	FIT $q = 25\%$	-0.4768	+0.4260	-0.1509
	FIT $q = 100\%$	-0.3541	+0.1324	+0.2334
CSI 500 Ridge	UNIV $\delta = 1$	+0.1492	-0.3540	-0.0659
	UNIV $\delta = 5$	-0.8592	+0.5959	+0.0859
	UNIV $\delta = 20$	+0.6997	+1.0309	+1.2390
	UNIV $\delta = 60$	+0.6020	+5.4154	+2.7009
	UNIV $\delta = 240$	+2.0215	+8.3399	+11.7429
CSI 500 Ridge	TRAD $q = 1\%$	-0.0005	+0.0008	+0.0016
	TRAD $q = 5\%$	-0.0044	+0.0027	+0.0064
	TRAD $q = 10\%$	-0.0126	-0.0006	+0.0188
	TRAD $q = 25\%$	-0.0080	+0.0213	+0.0454
	TRAD $q = 100\%$	-0.0895	+0.1178	+0.1313
CSI 500 Ridge	EXEC same close	-0.2988	-0.2395	-4.0600
	EXEC same open	-66.1357	-42.5457	-30.2284

Continued on next page

Universe / model	Intervention	2022	2023	2024
CSI 500 LightGBM	Clean	+0.0000	+0.0000	+0.0000
	OBS $L = 1$	+175.1816	+145.4391	+153.4135
	OBS $L = 2$	+251.6467	+205.1388	+212.1992
	OBS $L = 5$	+454.9675	+375.2018	+375.7179
	OBS $L = 10$	+452.3303	+374.2497	+377.0604
CSI 500 LightGBM	FIT $q = 1\%$	-1.0134	+1.2070	+1.7606
	FIT $q = 5\%$	-1.0410	-0.4120	+2.2700
	FIT $q = 10\%$	-0.4420	-1.4440	+2.0634
	FIT $q = 25\%$	-1.3838	-0.2999	+2.5308
	FIT $q = 100\%$	-2.7636	-2.1504	+3.3290
CSI 500 LightGBM	UNIV $\delta = 1$	-1.1217	-1.7318	+2.6649
	UNIV $\delta = 5$	+1.2767	-0.6177	+7.5085
	UNIV $\delta = 20$	-3.1591	-0.6791	+4.6392
	UNIV $\delta = 60$	-3.1516	-0.3186	-1.2185
	UNIV $\delta = 240$	+5.9588	+5.3068	+14.3747
CSI 500 LightGBM	TRAD $q = 1\%$	+0.0051	+0.0012	+0.0029
	TRAD $q = 5\%$	+0.0033	+0.0042	+0.0168
	TRAD $q = 10\%$	+0.0067	+0.0060	+0.0420
	TRAD $q = 25\%$	+0.0294	+0.0105	+0.0528
	TRAD $q = 100\%$	+0.0845	+0.0254	+0.2323
CSI 500 LightGBM	EXEC same close	-0.1853	+0.9980	-4.8213
	EXEC same open	-44.6137	-11.7463	-16.8688

OBS and same-open EXEC retain opposite directions across years; the remaining axes display more variable annual responses.

B.4 Return levels and development-period anchors

The paired differences measure changes in arithmetic annualized mean daily return, not compound annual growth rates. Table 9 gives clean and two-session OBS levels, including Momentum. The large future-information levels are diagnostic counterfactuals, not attainable investment returns.

Table 9 reports Net for clean and $t + 2$ features, followed by their difference ΔNet. Comparisons of clean models are made within a period and stock universe.

Table 9: **Future observations can turn weak clean results into apparent profitability.** Clean and exposed Net (%) show how much the economic evidence changes.

Period	Universe	Model	Days	Clean ↑	OBS $t + 2$	ΔNet (pp)
2016–2021	CSI 300	Ridge	1,461	−3.0643	+216.3255	+219.3898
		LightGBM	1,461	**+14.9565**	+215.8994	+200.9429
2022–2024	CSI 300	Ridge	726	−8.5745	+211.4661	+220.0406
		LightGBM	726	**−5.0055**	+208.8138	+213.8193
	CSI 500	Ridge	726	**−0.4884**	+220.3606	+220.8489
		LightGBM	726	−1.5457	+221.4493	+222.9949
	CSI 300	Momentum	726	−13.4423	+67.6897	+81.1320
	CSI 500	Momentum	726	−13.2267	+74.8927	+88.1194

The 2016–2021 CSI 300 development comparison uses the five original anchors: two-session future observations, 5% preprocessing admission, five-session future membership, 5% future-fillability exposure and same-close execution. Table 10 gives pooled effects, Table 11 annual decompositions, and Table 12 descriptive lag-20 intervals.

The development anchors show large OBS gains but much smaller responses to the other departures. OBS raises annualized return by 219.39 points for Ridge and 200.94 for LightGBM; FIT and UNIV

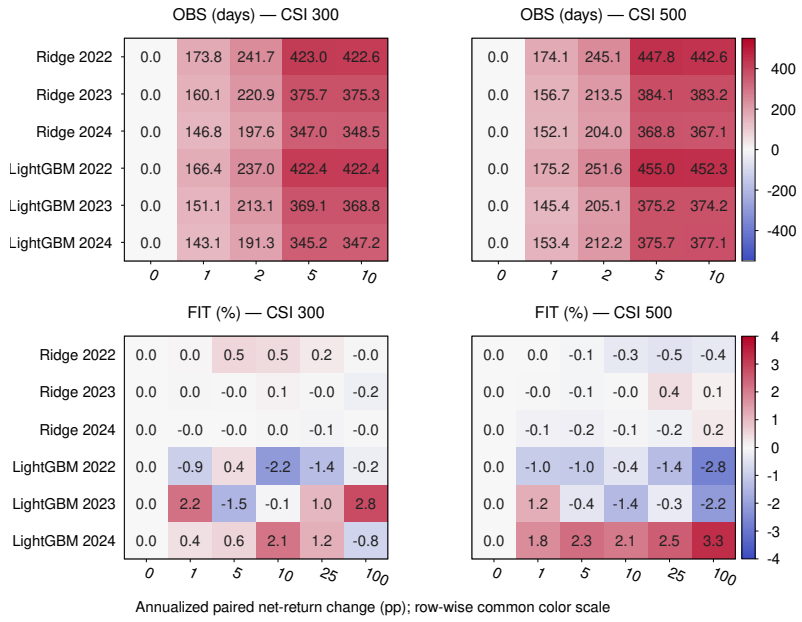


Figure 10: Year-level intervention-minus-clean net-return effects for observation timing and preprocessing scope, in annualized percentage points. Color scales are shared across universes within an intervention and differ between interventions. Clean references are zero.

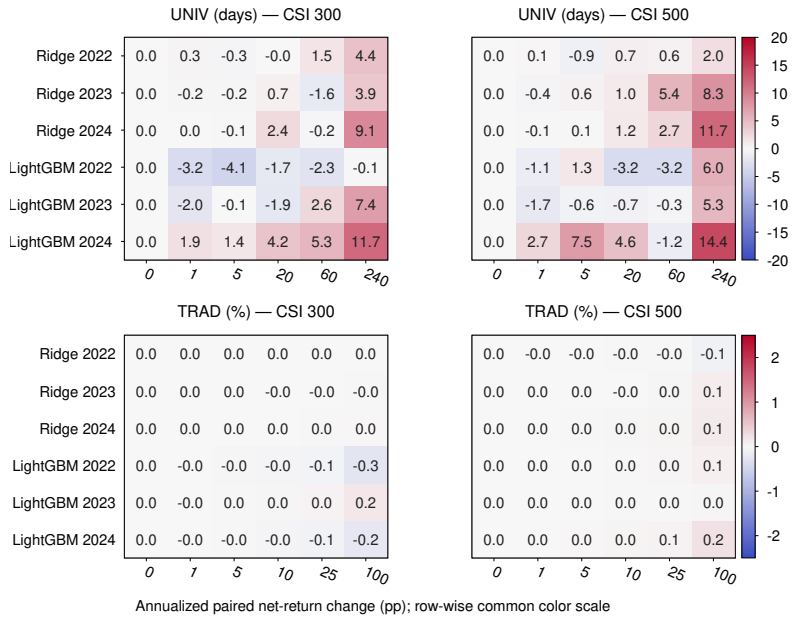


Figure 11: Year-level intervention-minus-clean net-return effects for future membership and future fillability, in annualized percentage points. Scales are shared across universes within an intervention and differ between interventions. Categorical execution results are retained in Table 8.

range from -0.13 to $+1.31$, TRAD is near zero, and same-close execution lowers return by 2.98 and 0.28 points. Annual columns decompose these same paired paths.

Table 10: **Future observations also dominate the earlier development comparison.** The other four departures have much smaller effects on CSI 300. ΔNet (pp), 2016–2021.

Intervention	Ridge	LightGBM
Features at $t+2$	$+219.3898$	$+200.9429$
Preprocessing scope (5%)	-0.1273	$+1.3129$
Membership at $t+5$	$+0.0888$	$+1.2582$
Future fillability (5%)	$+0.0011$	$+0.0020$
Execution at close t	-2.9816	-0.2822

Table 11: **The development averages conceal substantial annual variation.** The annual ΔNet (pp) breakdown retains every intervention and year.

Model	Intervention	2016	2017	2018	2019	2020	2021
Ridge	Features at $t+2$	$+181.1361$	$+184.8106$	$+194.8590$	$+223.7465$	$+286.1767$	$+245.8916$
	Preprocessing (5%)	-0.7061	$+0.2586$	-0.0107	-0.2606	$+0.0062$	-0.0501
	Membership at $t+5$	-0.1700	$+0.6936$	$+0.0468$	$+0.1522$	-0.7214	$+0.5300$
	Future fillability (5%)	$+0.0261$	-0.0101	-0.0015	-0.0065	$+0.0198$	-0.0210
	Execution at close t	-5.8219	-2.6488	-4.4759	-0.6072	-0.9662	-3.3689
LightGBM	Features at $t+2$	$+155.3271$	$+166.7797$	$+180.9375$	$+209.6059$	$+254.2563$	$+239.0435$
	Preprocessing (5%)	-0.8646	$+1.8035$	$+1.2456$	$+1.0340$	$+1.3222$	$+3.3446$
	Membership at $t+5$	$+3.2299$	$+0.6009$	-2.5946	$+2.2462$	$+1.2423$	$+2.8151$
	Future fillability (5%)	-0.0010	-0.0027	$+0.0041$	$+0.0099$	$+0.0037$	-0.0021
	Execution at close t	-1.9914	-0.9070	$+0.0471$	$+1.9923$	$+1.4178$	-2.2517

Annual results show how each intervention's effect varies over time.

Table 12: **The large OBS response remains clear after accounting for temporal variation.** Earlier-period ΔNet (pp) for CSI 300.

Model	Intervention	ΔNet (pp)	CI
Ridge	Features at $t+2$	$+219.390$	$[+202.055, +236.725]$
	Preprocessing (5%)	-0.127	$[-0.380, +0.125]$
	Membership at $t+5$	$+0.089$	$[-0.452, +0.630]$
	Future fillability (5%)	$+0.001$	$[-0.011, +0.014]$
	Execution at close t	-2.982	$[-7.966, +2.003]$
LightGBM	Features at $t+2$	$+200.943$	$[+184.321, +217.565]$
	Preprocessing (5%)	$+1.313$	$[+0.006, +2.619]$
	Membership at $t+5$	$+1.258$	$[-0.109, +2.625]$
	Future fillability (5%)	$+0.002$	$[-0.004, +0.008]$
	Execution at close t	-0.282	$[-5.817, +5.253]$

The smaller departures have different signs and magnitudes from OBS.

B.5 Whole-vector fallback under future observation requests

For each positive OBS lead, we request the same stock's complete future vector. A missing stock–date key or non-finite component triggers replacement by the entire clean vector; no observation is dropped. Feature-group and sparse interventions then select their designated columns or rows from this fallback-resolved vector. The rule was fixed after inspecting later-period anchors, so the 2022–2024 severity and mechanism studies remain retrospective (Appendix K).

At the reported two-session anchor, whole-vector fallback is rare and comes only from non-finite values. Table 13 gives CSI 300 counts at $L = 2$ for both fitted models; the largest fold–split rate is 0.025656%. For CSI 500, the maximum is 0.044645%: 52 of 116,474 validation requests in the 2023 fold. Neither universe has a missing requested stock–date key. The Momentum control

uses the same rule for 34 of 216,508 CSI 300 test requests (0.0157%) and 101 of 359,704 CSI 500 requests (0.0281%).

Table 13: **Unavailable future vectors are rare and do not remove observations.** Almost every requested vector is applied; the remaining rows retain their clean inputs.

Year	Split	Requested	Applied	Fallback	Fallback (%)
2022	train	353,193	353,193	0	0.000000
	validation	70,985	70,979	6	0.008452
	test	71,948	71,930	18	0.025018
2023	train	356,122	356,116	6	0.001685
	validation	70,160	70,142	18	0.025656
	test	72,476	72,476	0	0.000000
2024	train	358,757	358,733	24	0.006690
	validation	70,676	70,676	0	0.000000
	test	72,084	72,068	16	0.022196

B.6 An unfitted ranking control and terminal positions

The Momentum rule ranks the trailing twenty-session return directly (Jegadeesh & Titman, 1993), with no fitted predictor or preprocessor. Applicable interventions retain the same sleeve, cost and execution rules (Table 14).

Future observations raise returns even without a fitted prediction map. Advancing observations to $t+2$ raises return by 81.132 and 88.119 points in CSI 300 and CSI 500, positive in each of the three years. Future fillability at 5% changes return by only -0.00539 and $+0.00552$ points. Membership and same-close timing produce small negative effects. FIT is inapplicable and reported as N/A rather than as a zero-effect estimate.

Table 14: **Look-ahead can inflate returns without a fitted predictor.** The Momentum control also shows a large OBS effect and much smaller effects from the other applicable departures. ΔNet (pp).

Intervention	CSI 300	CSI 500
Features at $t+2$	$+81.132$	$+88.119$
Preprocessing scope (5%)	N/A	N/A
Membership at $t+5$	-0.502	-0.070
Future fillability (5%)	-0.00539	$+0.00552$
Execution at close t	-0.512	-0.619

All original 2022–2024 severity paths and both Momentum paths finish with zero holdings. At the development endpoint, unresolved lots of `sh.600005` remain; their number depends on the model and intervention. These holdings are marked without a synthetic sale or cash receipt. We therefore report development and later-period levels separately, preserving their different terminal states.

C Information content, learning entry, and sparse exposure

This appendix gives the complete comparisons of information content, training and scoring exposure, and exposure amount: feature–target overlap, all representations and leads, all training–scoring cells, sparse doses and single-feature replacements. Paired uncertainty and factorial interactions follow. Every comparison retains its matching clean reference; shared endpoints are reported once within each comparison.

C.1 From future feature content to portfolio outcomes

The observation experiment separates what future information a model receives, how training and scoring turn it into ranks, and how the ranks change the portfolio. The complete comparisons use the severity study's two universes, fixed model configurations, annual folds and simulator.

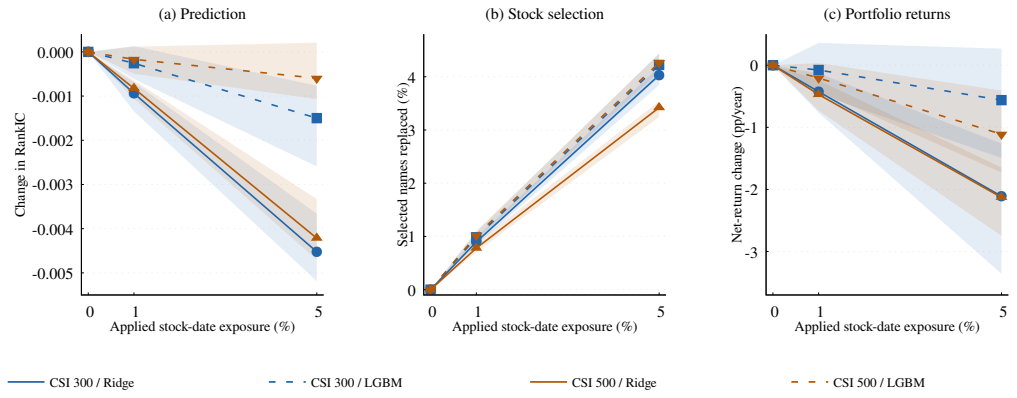


Figure 12: **Small prediction changes accompany measurable decision and portfolio effects.** One-session future exposure across stock–date records under the frozen clean All-feature model. Panels show RankIC, selected-name replacement and annualized arithmetic net-return change for the same interventions. Horizontal coordinates are actually exposed scoring rows; lines weight twenty fixed masks equally. Shaded mask min–max ranges describe exposure-location variation, not confidence intervals. Both models and both stock universes are shown.

C.1.1 Temporal overlap and the close-return proxy

Write O_u and C_u for the same stock's HFQ-adjusted open and close on market date u. The decision-t target and its log transform are

$$y_t = \frac{O_{t+6}}{O_{t+1}} - 1, \qquad \ell_t = \log(1 + y_t) = \log O_{t+6} - \log O_{t+1}. \tag{9}$$

Advancing a feature endpoint to $s = t + L$ changes the information supplied to the model, while its row, label, universe membership, and trading clock remain attached to t. Table 15 records every feature's information span and its intersection with the label interval $[t+1, t+6]$. Even $L = 1$ admits information from the first holding-day bar. Temporal overlap has different content across features: a signed return describes price direction, whereas a range or volume statistic describes future market state.

The table expresses spans relative to decision date t, so the target interval is $[1, 6]$. A return span joins its primitive close observations; overlap does not mean every interior price enters the statistic. Trading value contains a price component and is excluded from the market-state group.

Table 15: **Equal leads expose different kinds of future information.** Return windows contain price direction, while ranges and volume describe future market conditions.

Feature	Input span	Content	$L=1$	$L=2$	$L=5$	$L=10$
`ret_1`	$[L-1, L]$	Signed return	$\{1\}$	$[1,2]$	$[4,5]$	—
`ret_5`	$[L-5, L]$	Signed return	$\{1\}$	$[1,2]$	$[1,5]$	$[5,6]$
`ret_10`	$[L-10, L]$	Signed return	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
`ret_20`	$[L-20, L]$	Signed return	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
`ret_60`	$[L-60, L]$	Signed return	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
`vol_5`	$[L-5, L]$	Dispersion	$\{1\}$	$[1,2]$	$[1,5]$	$[5,6]$
`vol_20`	$[L-20, L]$	Dispersion	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
`vol_60`	$[L-60, L]$	Dispersion	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
`hl_range_1`	$\{L\}$	Range	$\{1\}$	$\{2\}$	$\{5\}$	—
`hl_range_5_mean`	$[L-4, L]$	Range	$\{1\}$	$[1,2]$	$[1,5]$	$\{6\}$
`oc_gap_1`	$\{L\}$	Intraday return	$\{1\}$	$\{2\}$	$\{5\}$	—
`close_ma5_ratio`	$[L-4, L]$	Price level	$\{1\}$	$[1,2]$	$[1,5]$	$\{6\}$
`close_ma20_ratio`	$[L-19, L]$	Price level	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$

Continued on next page

Feature	Input span	Content	$L=1$	$L=2$	$L=5$	$L=10$
close_ma60_ratio	$[L-59, L]$	Price level	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
log_volume	$\{L\}$	Activity	$\{1\}$	$\{2\}$	$\{5\}$	—
vol_ratio_5	$[L-4, L]$	Activity	$\{1\}$	$[1,2]$	$[1,5]$	$\{6\}$
vol_ratio_20	$[L-19, L]$	Activity	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$
log_dollar_volume	$\{L\}$	Trading value	$\{1\}$	$\{2\}$	$\{5\}$	—
up_days_ratio_20	$[L-20, L]$	Return signs	$\{1\}$	$[1,2]$	$[1,5]$	$[1,6]$

The nested return windows make the horizon-aligned mechanism explicit. Define the five-session close-return proxy q_t by

$$q_t = \log C_{t+5} - \log C_t = \mathrm{ret}_5(t+5), \tag{10}$$

$$= \mathrm{ret}_{10}(t+10) - \mathrm{ret}_5(t+10). \tag{11}$$

Thus $L=5$ exposes the proxy directly and $L=10$ exposes the same proxy as a linear combination. Its relationship to the actual open-to-open label is

$$\ell_t = q_t + g_{t+6} - g_{t+1}, \qquad g_u = \log O_u - \log C_{u-1}. \tag{12}$$

Here g_u is the overnight close-to-open change, distinct from the registered oc_gap_1 feature, which is the same-day open-to-close change. The boundary gaps explain why the close proxy closely tracks, but is not identical to, the target. Table 16 measures this alignment before preprocessing or model fitting. Each expression is evaluated on the same common-eligible stock–date set; their pooled mean daily RankICs are approximately 0.9660 in CSI 300 and 0.9629 in CSI 500. This model-free comparison provides the reference against which the learned high-lead responses can be interpreted.

Each alignment cell lists the $L=5$ and $L=10$ constructions in that order. RankIC averages daily cross-sectional correlations; Pearson correlation (Pearson) pools stock–date observations. Both use the original simple-return target on the same eligible sample.

The signed-return groups approach target reconstruction, while market-state groups remain much less predictive. At $L=5$, jointly exposed All models reach absolute RankIC 0.960–0.966 and Returns models reach 0.961–0.965. Market-state models reach 0.181–0.277; Volume-only and Range-only models reach 0.215–0.291 and 0.059–0.096, respectively. The complete cells below separate information about future market state from the signed-return identity that enables near-reconstruction.

Table 16: **The five- and ten-session constructions recover the same close-return proxy.** Their identical correlations show that the proxy closely tracks, but does not equal, the open-to-open target.

Universe	Period	Rows	Dates	RankIC	Pearson
300	2022	71,858	242	0.96565 / 0.96565	0.96695 / 0.96695
	2023	72,476	242	0.96830 / 0.96830	0.96614 / 0.96614
	2024	72,000	242	0.96414 / 0.96414	0.95597 / 0.95597
	2022–2024	216,334	726	0.96603 / 0.96603	0.96034 / 0.96034
500	2022	119,140	242	0.96023 / 0.96023	0.96720 / 0.96720
	2023	119,889	242	0.96596 / 0.96596	0.96691 / 0.96691
	2024	120,175	242	0.96249 / 0.96249	0.95802 / 0.95802
	2022–2024	359,204	726	0.96289 / 0.96289	0.96140 / 0.96140

C.1.2 Feature content crossed with training and scoring

We cross six feature groups with the same leads $L \in \{1,2,5,10\}$. *All* contains all 19 features; *Returns*, the five signed returns. *State* contains three return-volatility statistics, two high–low-range statistics and three volume statistics (log_volume, vol_ratio_5 and vol_ratio_20). It excludes signed returns, same-day direction, close-to-average price ratios, price-bearing trading value and the up-day fraction. *Volatility*, *Volume* and *Range* retain the corresponding three, three and two State features. State therefore preserves future information without the signed price-path basis in Equations (10)–(11).

Each group has its own clean preprocessor and fitted model. To see whether the model learns to use exposed information, we compare clean training and scoring (CC), clean training with exposed scoring (CL), exposed training with clean scoring (LC), and exposure at both stages (LL). The first letter always refers to training. Training exposure affects winsorization, standardization and model fitting; configurations and labels remain fixed. Test-only exposure and frozen-model evaluation on future inputs describe the same CL cell.

For a scalar endpoint M, the four cells yield the score-time contrast $M_{CL} - M_{CC}$, training-only contrast $M_{LC} - M_{CC}$, refitting increment $M_{LL} - M_{CL}$, and interaction $M_{LL} - M_{LC} - M_{CL} + M_{CC}$. Table 17 supplies all four cell values for both predictive and portfolio endpoints, so each contrast can be recovered within the same feature group and lead. Its return-change unit is

$$\Delta\mathrm{Net}_{ab} = 100 \times 252 \ \mathrm{mean}_t(r^{\mathrm{net}}_{ab,t} - r^{\mathrm{net}}_{CC,t}) \quad \text{percentage points per year.} \tag{13}$$

RankIC changes subtract the clean equal-date mean daily Spearman correlation.

The dense table contains all 384 cells across 96 settings. CC reports RankIC / Net levels; CL, LC, and LL report ΔRankIC / ΔNet against that row's CC. Keeping this order lets the reader follow a predictive change into its economic consequence. Net levels are percent and return changes are pp.

Changing the training inputs can reverse the effect of future scoring inputs. At $L = 1$, All-feature CL lowers net return in all four model–universe combinations, whereas LL raises it in all four. Thus the same representation has different economic effects under a clean model and a model fitted to that representation. Removing the signed-return basis still leaves a substantial response: at $L = 5$, State LL changes range from 0.1922 to 0.2950 in RankIC and from 125.23 to 199.98 annualized net-return percentage points. Future market-state information can therefore matter without providing the signed-return identity.

Table 17: **Future inputs can hurt a clean model yet help a model trained on them.** The complete grid separates information content, lead, and learning entry point.

RankIC / Net; exposed columns show ΔRankIC / ΔNet						
Universe / model	Features	L	CC	CL	LC	LL
CSI 300 Ridge	All	1	-0.0126 / -8.57	-0.0918 / -47.93	+0.0125 / -6.42	+0.3968 / +160.23
		2	-0.0126 / -8.57	-0.1193 / -61.02	+0.0127 / -6.70	+0.5200 / +220.04
		5	-0.0126 / -8.57	-0.3030 / -156.05	+0.0164 / -7.49	+0.9765 / +381.88
		10	-0.0126 / -8.57	+0.1107 / +40.50	+0.0041 / -5.52	+0.9722 / +382.16
CSI 300 Ridge	Returns	1	+0.0039 / -10.16	-0.1416 / -51.85	-0.0011 / -3.25	+0.3605 / +154.13
		2	+0.0039 / -10.16	-0.2715 / -103.16	+0.0013 / -3.82	+0.4373 / +195.80
		5	+0.0039 / -10.16	-0.6774 / -252.01	+0.0014 / -6.87	+0.9611 / +384.13
		10	+0.0039 / -10.16	+0.2417 / +129.96	-0.0101 / -3.46	+0.9564 / +383.36
CSI 300 Ridge	State	1	-0.0109 / -11.42	+0.0281 / +25.52	+0.0094 / +4.42	+0.0626 / +47.75
		2	-0.0109 / -11.42	+0.0585 / +47.86	+0.0089 / +4.62	+0.1136 / +72.60
		5	-0.0109 / -11.42	+0.1684 / +107.82	+0.0094 / +8.14	+0.1922 / +127.16
		10	-0.0109 / -11.42	+0.0879 / +44.59	+0.0132 / +7.44	+0.1764 / +99.79
CSI 300 Ridge	Volatility	1	+0.0242 / -4.82	+0.0242 / +16.96	-0.0117 / -1.35	+0.0154 / +17.74
		2	+0.0242 / -4.82	+0.0552 / +35.09	-0.0171 / -2.71	+0.0461 / +38.91
		5	+0.0242 / -4.82	+0.1626 / +90.34	-0.0236 / -3.03	+0.1594 / +104.51
		10	+0.0242 / -4.82	+0.0340 / +28.50	-0.0326 / -0.26	+0.0441 / +46.49
CSI 300 Ridge	Volume	1	-0.0144 / -5.11	+0.0198 / +17.81	+0.0182 / +1.08	+0.0730 / +39.89
		2	-0.0144 / -5.11	+0.0485 / +40.85	+0.0228 / +1.58	+0.1373 / +75.35
		5	-0.0144 / -5.11	+0.1461 / +89.21	+0.0260 / +3.40	+0.2423 / +129.06
		10	-0.0144 / -5.11	+0.0653 / +23.77	+0.0275 / +4.87	+0.1444 / +69.02
CSI 300 Ridge	Range	1	-0.0692 / -10.05	+0.0290 / +37.95	+0.0368 / -0.59	+0.0762 / +48.83
		2	-0.0692 / -10.05	+0.0570 / +59.95	+0.0187 / -0.17	+0.0833 / +67.40
		5	-0.0692 / -10.05	+0.1260 / +109.89	-0.0042 / -0.95	+0.1278 / +118.41
		10	-0.0692 / -10.05	+0.0928 / +64.63	-0.0068 / -0.47	+0.0955 / +78.35

Continued on next page

RankIC / Net; exposed columns show ΔRankIC / ΔNet

Universe / model	Features	L	CC	CL	LC	LL
CSI 300 LightGBM	All	1	+0.0083 / -5.01	-0.0307 / -14.64	-0.0036 / -7.70	+0.3622 / +153.53
		2	+0.0083 / -5.01	-0.0375 / -24.06	-0.0061 / -9.47	+0.4936 / +213.82
		5	+0.0083 / -5.01	-0.1318 / -70.82	-0.0030 / -10.99	+0.9576 / +378.90
		10	+0.0083 / -5.01	+0.0351 / -1.00	-0.0146 / -10.69	+0.9529 / +379.47
CSI 300 LightGBM	Returns	1	+0.0064 / -2.34	-0.0116 / -18.43	-0.0011 / -9.58	+0.3420 / +140.99
		2	+0.0064 / -2.34	-0.0397 / -38.47	+0.0009 / -12.05	+0.4223 / +184.42
		5	+0.0064 / -2.34	-0.1376 / -101.90	-0.0011 / -14.29	+0.9586 / +376.43
		10	+0.0064 / -2.34	+0.1058 / -1.43	-0.0127 / -12.12	+0.9543 / +376.65
CSI 300 LightGBM	State	1	-0.0078 / -4.48	+0.0128 / +8.82	+0.0040 / -1.55	+0.0552 / +39.40
		2	-0.0078 / -4.48	+0.0226 / +14.76	+0.0093 / -2.36	+0.1194 / +71.82
		5	-0.0078 / -4.48	+0.0533 / +28.52	+0.0112 / -5.44	+0.2062 / +125.23
		10	-0.0078 / -4.48	+0.0348 / +21.45	+0.0045 / -7.26	+0.1637 / +95.76
CSI 300 LightGBM	Volatility	1	+0.0016 / -2.99	+0.0082 / +4.92	-0.0021 / -3.42	+0.0133 / +5.48
		2	+0.0016 / -2.99	+0.0180 / +10.24	-0.0017 / -5.39	+0.0416 / +22.60
		5	+0.0016 / -2.99	+0.0620 / +26.39	-0.0091 / -7.33	+0.1366 / +73.67
		10	+0.0016 / -2.99	+0.0112 / +9.93	-0.0165 / -6.44	+0.0395 / +32.02
CSI 300 LightGBM	Volume	1	-0.0053 / -5.35	+0.0144 / +11.51	+0.0009 / +0.59	+0.0513 / +37.44
		2	-0.0053 / -5.35	+0.0356 / +26.80	+0.0041 / +2.02	+0.1109 / +75.35
		5	-0.0053 / -5.35	+0.0766 / +51.97	+0.0065 / +4.94	+0.2202 / +127.34
		10	-0.0053 / -5.35	+0.0268 / +12.78	+0.0127 / +3.11	+0.1422 / +74.03
CSI 300 LightGBM	Range	1	-0.0475 / -12.52	+0.0128 / +12.40	+0.0040 / +0.41	+0.0425 / +43.45
		2	-0.0475 / -12.52	+0.0266 / +19.85	-0.0013 / +0.73	+0.0517 / +57.55
		5	-0.0475 / -12.52	+0.0585 / +38.37	-0.0026 / +1.76	+0.1181 / +110.68
		10	-0.0475 / -12.52	+0.0494 / +26.42	-0.0082 / -0.87	+0.0768 / +70.33
CSI 500 Ridge	All	1	+0.0045 / -0.49	-0.0876 / -45.51	-0.0072 / -13.53	+0.3868 / +160.96
		2	+0.0045 / -0.49	-0.0924 / -49.48	-0.0103 / -15.23	+0.4986 / +220.85
		5	+0.0045 / -0.49	-0.1660 / -87.66	-0.0155 / -15.83	+0.9552 / +400.23
		10	+0.0045 / -0.49	+0.1624 / +92.46	-0.0163 / -9.49	+0.9509 / +397.65
CSI 500 Ridge	Returns	1	+0.0121 / -8.63	-0.1628 / -58.04	-0.0125 / -4.58	+0.3561 / +159.68
		2	+0.0121 / -8.63	-0.2672 / -97.61	-0.0164 / -7.81	+0.4222 / +204.68
		5	+0.0121 / -8.63	-0.5830 / -214.49	-0.0217 / -7.73	+0.9490 / +408.65
		10	+0.0121 / -8.63	+0.2206 / +122.40	-0.0220 / -0.42	+0.9438 / +406.11
CSI 500 Ridge	State	1	-0.0137 / -6.31	+0.0121 / +12.79	-0.0015 / -3.51	+0.0666 / +47.14
		2	-0.0137 / -6.31	+0.0311 / +27.25	-0.0008 / -1.43	+0.1409 / +91.00
		5	-0.0137 / -6.31	+0.1367 / +106.19	+0.0001 / -0.74	+0.2451 / +172.47
		10	-0.0137 / -6.31	+0.1171 / +82.27	+0.0038 / +2.10	+0.2335 / +138.32
CSI 500 Ridge	Volatility	1	+0.0033 / -9.75	+0.0199 / +16.30	-0.0069 / -1.61	+0.0156 / +16.69
		2	+0.0033 / -9.75	+0.0477 / +38.94	-0.0103 / -2.08	+0.0449 / +41.56
		5	+0.0033 / -9.75	+0.1578 / +118.26	-0.0136 / +0.70	+0.1303 / +101.60
		10	+0.0033 / -9.75	+0.0981 / +84.72	-0.0283 / -1.66	+0.0823 / +89.07
CSI 500 Ridge	Volume	1	-0.0043 / -5.26	-0.0041 / +9.17	-0.0001 / -2.29	+0.0749 / +52.03
		2	-0.0043 / -5.26	+0.0113 / +29.95	-0.0007 / -2.03	+0.1578 / +103.71
		5	-0.0043 / -5.26	+0.1481 / +113.37	-0.0032 / -1.99	+0.2952 / +187.42
		10	-0.0043 / -5.26	+0.1324 / +54.23	-0.0033 / -3.58	+0.1995 / +115.86
CSI 500 Ridge	Range	1	-0.0781 / -10.82	+0.0146 / +14.22	+0.0215 / -1.43	+0.0656 / +52.86
		2	-0.0781 / -10.82	+0.0398 / +36.47	+0.0126 / -1.20	+0.0933 / +78.03
		5	-0.0781 / -10.82	+0.1540 / +136.38	+0.0024 / -1.05	+0.1746 / +150.85
		10	-0.0781 / -10.82	+0.1463 / +118.71	-0.0004 / -0.71	+0.1444 / +114.16
CSI 500 LightGBM	All	1	+0.0061 / -1.55	-0.0111 / -23.78	-0.0044 / -12.32	+0.3737 / +158.01
		2	+0.0061 / -1.55	+0.0114 / -16.10	-0.0085 / -12.47	+0.4947 / +222.99
		5	+0.0061 / -1.55	+0.0232 / -21.26	-0.0153 / -14.56	+0.9566 / +401.96
		10	+0.0061 / -1.55	+0.1374 / +52.63	-0.0144 / -8.94	+0.9517 / +401.21
CSI 500 LightGBM	Returns	1	+0.0013 / -3.55	+0.0066 / -15.60	-0.0007 / -8.13	+0.3510 / +150.61
		2	+0.0013 / -3.55	+0.0019 / -26.09	-0.0060 / -11.24	+0.4199 / +195.56
		5	+0.0013 / -3.55	-0.0154 / -44.66	-0.0107 / -12.76	+0.9604 / +403.98
		10	+0.0013 / -3.55	+0.1816 / +47.54	-0.0099 / -6.71	+0.9555 / +402.32

Continued on next page

RankIC / Net; exposed columns show ΔRankIC / ΔNet						
Universe / model	Features	L	CC	CL	LC	LL
CSI 500 LightGBM	State	1	-0.0176 / -7.78	+0.0151 / +6.84	+0.0044 / -2.63	+0.0734 / +51.46
		2	-0.0176 / -7.78	+0.0384 / +18.36	+0.0125 / -2.99	+0.1687 / +108.66
		5	-0.0176 / -7.78	+0.1115 / +46.19	+0.0092 / -2.87	+0.2950 / +199.98
		10	-0.0176 / -7.78	+0.0842 / +41.72	+0.0149 / +2.10	+0.2556 / +160.94
CSI 500 LightGBM	Volatility	1	-0.0159 / -8.07	+0.0103 / +6.18	-0.0011 / -3.50	+0.0120 / +5.67
		2	-0.0159 / -8.07	+0.0274 / +15.56	+0.0002 / -4.08	+0.0400 / +24.24
		5	-0.0159 / -8.07	+0.1005 / +52.97	+0.0071 / -4.63	+0.1317 / +82.99
		10	-0.0159 / -8.07	+0.0492 / +39.89	+0.0134 / -1.34	+0.0817 / +73.77
CSI 500 LightGBM	Volume	1	+0.0067 / -7.07	+0.0123 / +11.47	-0.0123 / -0.52	+0.0541 / +48.53
		2	+0.0067 / -7.07	+0.0393 / +31.16	-0.0142 / +0.15	+0.1421 / +104.83
		5	+0.0067 / -7.07	+0.1311 / +79.34	-0.0243 / -1.70	+0.2806 / +188.82
		10	+0.0067 / -7.07	+0.0837 / +38.79	-0.0165 / -1.35	+0.2105 / +124.20
CSI 500 LightGBM	Range	1	-0.0585 / -8.40	+0.0089 / +8.86	+0.0008 / -1.70	+0.0345 / +40.75
		2	-0.0585 / -8.40	+0.0287 / +21.98	+0.0014 / -2.34	+0.0517 / +61.42
		5	-0.0585 / -8.40	+0.0907 / +52.13	+0.0234 / -3.40	+0.1498 / +134.65
		10	-0.0585 / -8.40	+0.0839 / +38.90	+0.0030 / -3.64	+0.1165 / +106.52

Appendix C.2 gives pointwise intervals for every non-reference cell, every single-feature comparison and the equal-mask sparse-exposure means.

C.1.3 Sparse reach and the selected-name boundary

The sparse study advances test inputs by one session while keeping the clean All-feature fit. Date and stock–date exposure use $p \in \{0, 0.01, 0.05, 0.10, 0.25, 1\}$ and 20 fixed, outcome-independent nested masks per axis. Date exposure changes the whole cross-section on selected dates; stock–date exposure disperses changes across stocks and dates. Feature exposure uses $k \in \{0, 1, 2, 4, 8, 19\}$ and 19 cyclic rotations of a fixed outcome-independent order. Each feature appears in exactly k rotations at any intermediate k.

Partial replacement requires the complete future 19-feature vector to be available and finite; otherwise the whole row remains clean. Eligible rows combine the masked future components with the remaining contemporaneous components. All zero-dose endpoints equal CC; all three full-exposure endpoints equal All-feature CL at $L = 1$.

Realized exposure (Reach) counts information actually supplied to the model: ρ_{row} is the fraction of scoring rows receiving a future feature, and ρ_{cell} is the fraction of the $19N$ scoring feature cells sourced from $t+1$. These coordinates cover all 726 scoring dates. Portfolio transmission is measured on sleeve-opening dates through

$$\text{Rep}_{ab} = 100 \, \frac{\sum_t |S_{CC,t} \setminus S_{ab,t}|}{\sum_t |S_{CC,t}|}, \tag{14}$$

where $S_{ab,t}$ is the new sleeve's selected set. Selected-name replacement (Rep.) is reported as a percentage of clean slots, weighted by their counts across dates; it is a set distance from CC rather than an additive factorial contrast. Reach uses rows for date and stock–date exposure, and feature cells for feature exposure. Table entries average the twenty masks or nineteen rotations equally.

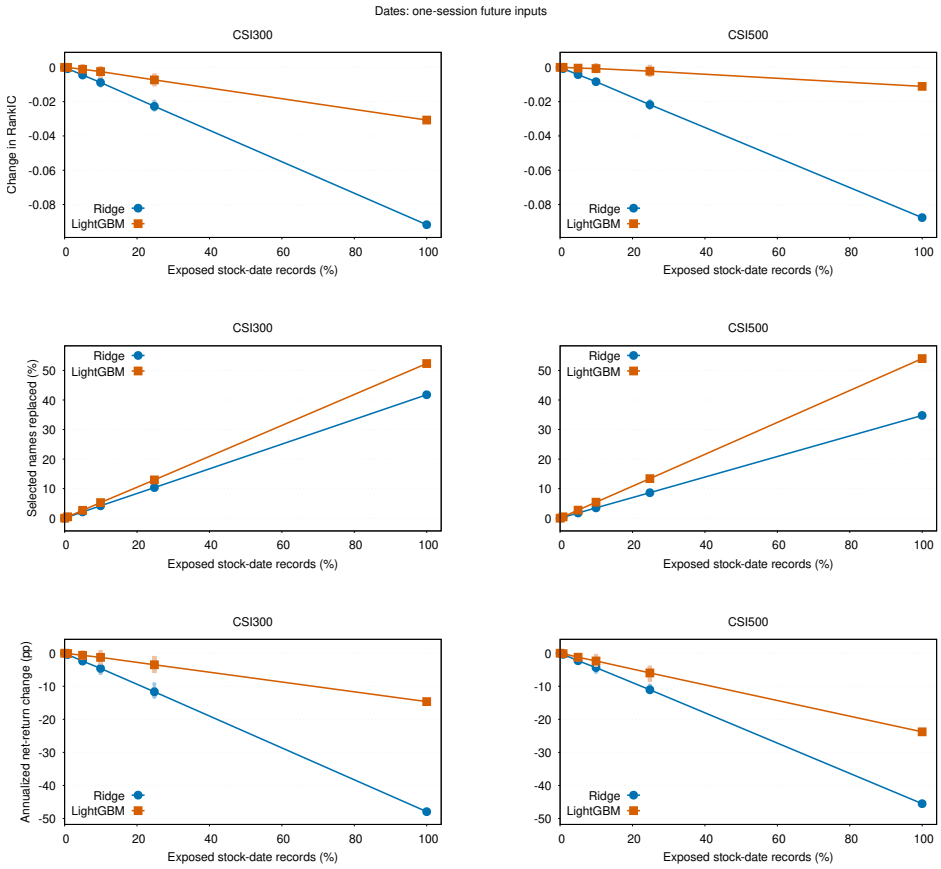


Figure 13: Complete date-exposure curves at $L = 1$, with the clean model frozen. Columns show CSI 300 and CSI 500; both Ridge and LightGBM are retained in every panel. The horizontal coordinate is realized exposure among all scoring rows. Rows trace ΔRankIC, clean selected-name replacement (percent), and ΔNet (annualized arithmetic percentage points). Individual points show all 20 fixed date masks; lines connect their equal-mask means at the evaluated doses. Mask spread measures variation in the exposed dates, not a confidence interval.

Sparse exposure links small perturbations to the same full-exposure endpoint, with responses that can be nonmonotone. Table 18 reports realized reach, RankIC, replacement and net return at every dose, including LightGBM's nonmonotone feature-subset response. Whole-date and dispersed stock–date exposure remain separate: only the latter directly tests where contaminated records fall within a cross-sectional ranking. Figures 13–15 show all doses for both models and universes. Points retain all masks or rotations; their spread describes exposure-location variation, while lines connect equal-design means.

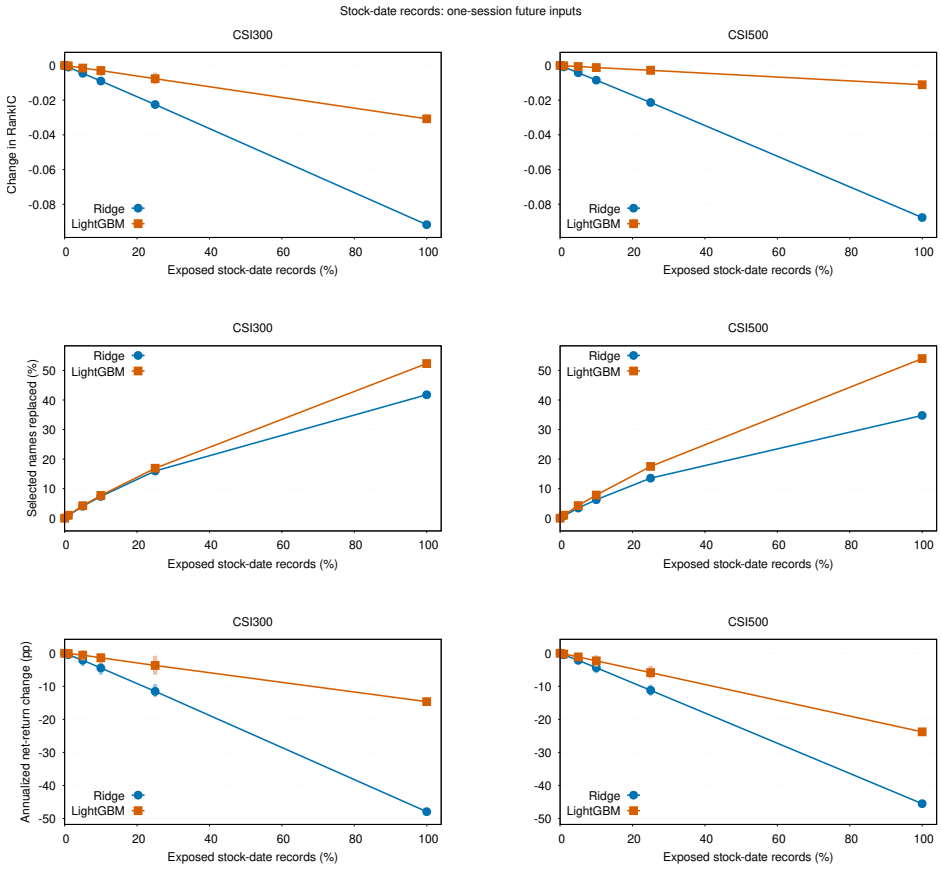


Figure 14: Complete stock–date-exposure curves at $L = 1$ for both models and universes. Exposure is dispersed across individual scoring records rather than synchronized across dates. Axes and units match Figure 13. All 20 fixed stock–date masks are shown, together with their equal-mask means. The full-exposure endpoint is exactly the same CL condition as in the date and feature curves.

Table 18: **Limited exposure can reach the selection boundary and change returns.** Increasing the exposed fraction does not produce a uniform predictive or economic response.

			Ridge			LightGBM		
Exposure	Dose	Reach (%)	ΔRankIC	Rep. (%)	ΔNet	ΔRankIC	Rep. (%)	ΔNet
CSI 300								
Dates	0%	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1%	0.826	-0.0008	0.36	-0.41	-0.0001	0.45	-0.09
	5%	4.959	-0.0045	2.11	-2.37	-0.0011	2.65	-0.64
	10%	9.918	-0.0088	4.18	-4.59	-0.0025	5.27	-1.25
	25%	24.792	-0.0227	10.35	-11.66	-0.0073	12.96	-3.48
	100%	99.992	-0.0918	41.78	-47.93	-0.0307	52.33	-14.64
Stock-dates	0%	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1%	0.999	-0.0009	0.90	-0.43	-0.0003	0.98	-0.08
	5%	4.999	-0.0045	4.03	-2.11	-0.0015	4.22	-0.56
	10%	9.998	-0.0090	7.37	-4.45	-0.0029	7.69	-1.36
	25%	24.998	-0.0225	15.97	-11.52	-0.0076	16.90	-3.69
	100%	99.992	-0.0918	41.78	-47.93	-0.0307	52.33	-14.64

Continued on next page

			Ridge			LightGBM		
Exposure	Dose	Reach (%)	ΔRankIC	Rep. (%)	ΔNet	ΔRankIC	Rep. (%)	ΔNet
Features	0	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1	5.263	-0.0049	11.90	-0.83	-0.0013	11.59	-0.68
	2	10.525	-0.0106	17.20	-3.78	-0.0011	17.78	-0.09
	4	21.051	-0.0200	25.66	-8.74	-0.0016	26.66	+0.65
	8	42.102	-0.0346	36.14	-15.92	-0.0041	37.93	-0.44
	19	99.992	-0.0918	41.78	-47.93	-0.0307	52.33	-14.64
CSI 500								
Dates	0%	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1%	0.827	-0.0007	0.30	-0.35	0.0000	0.46	-0.12
	5%	4.959	-0.0043	1.75	-2.28	-0.0005	2.72	-1.21
	10%	9.919	-0.0084	3.48	-4.35	-0.0007	5.41	-2.34
	25%	24.793	-0.0218	8.62	-11.05	-0.0022	13.39	-5.95
	100%	99.986	-0.0876	34.77	-45.51	-0.0111	54.01	-23.78
Stock-dates	0%	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1%	1.000	-0.0008	0.78	-0.47	-0.0002	1.02	-0.21
	5%	4.999	-0.0042	3.42	-2.13	-0.0006	4.26	-1.11
	10%	9.998	-0.0085	6.25	-4.37	-0.0012	7.83	-2.29
	25%	24.996	-0.0213	13.54	-11.21	-0.0028	17.52	-5.86
	100%	99.986	-0.0876	34.77	-45.51	-0.0111	54.01	-23.78
Features	0	0.000	0.0000	0.00	0.00	0.0000	0.00	0.00
	1	5.262	-0.0050	7.74	-1.54	+0.0003	11.31	-0.15
	2	10.525	-0.0100	11.67	-3.62	+0.0014	17.32	+0.07
	4	21.050	-0.0197	16.84	-7.53	+0.0019	26.10	0.00
	8	42.099	-0.0377	23.05	-15.24	+0.0032	37.96	-2.67
	19	99.986	-0.0876	34.77	-45.51	-0.0111	54.01	-23.78

C.1.4 WHICH INDIVIDUAL FEATURE IS EXPOSED?

An average over exposed features can conceal substantial effects of opposite signs. Table 19 gives the 19 single-feature responses whose mean is reported at $k = 1$, ordered as in Table 2 and accompanied by their fixed-cycle positions. Every model–universe combination contains positive and negative portfolio responses. The $k = 2, 4, 8$ curves use joint replacements under the same balanced rotations, not sums of single-feature effects.

Rotation records the fixed feature order from 1 to 19. A two-feature rotation pairs a feature with its successor, wrapping from 19 to 1. The table retains the original feature-definition order; variation across rotations describes feature content, not uncertainty. Thus both feature identity and the fraction of exposed cells matter.

Table 19: **The identity of an exposed feature can reverse its effect.** Ten-session return and intraday return illustrate opposite predictive and economic responses at the same one-feature dose.

ΔRankIC / ΔNet (pp)		CSI 300		CSI 500	
Feature	Rotation	Ridge	LightGBM	Ridge	LightGBM
`ret_1`	19	+0.0833 / +31.77	+0.0636 / +12.78	+0.0293 / +11.54	+0.0275 / -7.11
`ret_5`	11	-0.0753 / -29.56	-0.0244 / -6.99	-0.0720 / -28.69	-0.0143 / -2.01
`ret_10`	1	+0.0770 / +30.75	+0.0282 / +7.97	+0.0438 / +18.40	+0.0316 / +14.20
`ret_20`	2	+0.0040 / +0.22	+0.0094 / +3.58	-0.0255 / -11.71	+0.0076 / +2.04
`ret_60`	16	-0.0179 / -6.97	-0.0020 / -1.62	-0.0188 / -8.70	-0.0029 / -1.46
`vol_5`	4	+0.0002 / -0.17	-0.0039 / -0.76	+0.0024 / +2.38	-0.0012 / -0.22
`vol_20`	7	-0.0010 / +0.13	-0.0008 / +0.07	+0.0031 / +3.83	-0.0003 / +1.44
`vol_60`	9	-0.0019 / -1.44	-0.0010 / -0.15	-0.0032 / -2.27	-0.0007 / -0.42
`hl_range_1`	12	+0.0104 / +6.25	+0.0060 / +2.95	+0.0020 / +1.17	+0.0044 / +3.59

Continued on next page

ΔRankIC / ΔNet (pp)		CSI 300		CSI 500	
Feature	Rotation	Ridge	LightGBM	Ridge	LightGBM
hl_range_5_mean	10	+0.0116 / +7.85	+0.0033 / +1.06	+0.0151 / +11.48	+0.0063 / +2.35
oc_gap_1	18	-0.1443 / -55.23	-0.0834 / -21.91	-0.1115 / -47.16	-0.0548 / -24.31
close_ma5_ratio	3	+0.0460 / +17.56	+0.0150 / +1.89	+0.0345 / +13.38	+0.0202 / +9.28
close_ma20_ratio	14	-0.1146 / -39.02	-0.0336 / -10.49	-0.0022 / +3.46	-0.0091 / -0.67
close_ma60_ratio	6	+0.0055 / +3.35	-0.0048 / -1.76	-0.0074 / +0.11	-0.0043 / +1.84
log_volume	5	-0.0059 / -0.91	-0.0036 / -0.87	-0.0021 / -0.79	+0.0002 / +0.81
vol_ratio_5	8	-0.0151 / -3.06	-0.0023 / +0.16	+0.0078 / +2.42	-0.0005 / +0.62
vol_ratio_20	17	+0.0302 / +17.33	+0.0096 / +1.55	+0.0036 / +1.11	+0.0058 / +1.00
log_dollar_volume	13	+0.0064 / +2.50	+0.0022 / +1.51	-0.0093 / -3.59	-0.0069 / -3.04
up_days_ratio_20	15	+0.0080 / +2.79	-0.0029 / -1.82	+0.0163 / +4.47	-0.0031 / -0.77

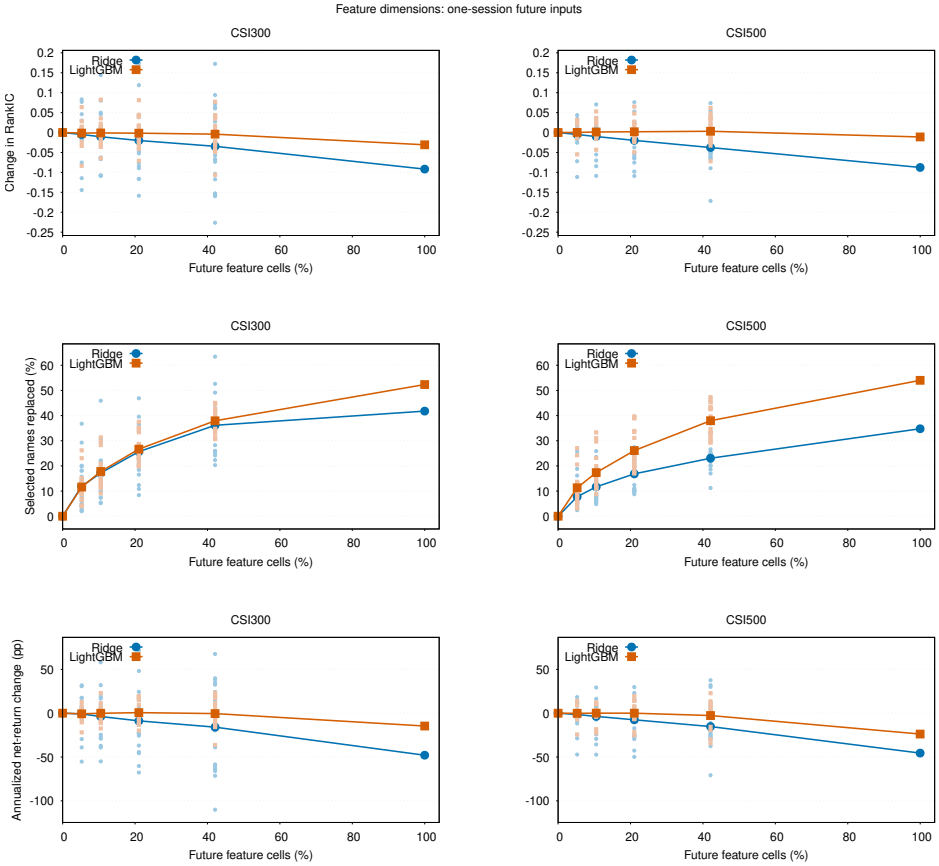


Figure 15: Complete feature-exposure curves at $L = 1$ for both models and universes. The horizontal coordinate is realized exposure among all scoring feature cells, not the proportion of rows touched. Points retain all 19 balanced rotations at each feature count; lines connect the equal-rotation means. Vertical axes match Figure 13. Rotation spread describes the effect of feature identity; each feature enters exactly k of the 19 rotations at count k.

C.1.5 ANNUAL DECOMPOSITION

Annual results retain the sign and magnitude differences hidden by pooling. Table 20 reports net-return responses for every feature group and lead, separating CL from LL. Table 21 reports RankIC, selected-name replacement and net return at every dose for all three sparse-exposure designs. Each

year has 242 scoring and return dates; sleeve-opening dates number 242, 242 and 201. The final 41 scoring dates of 2024 therefore contribute to prediction diagnostics and valuation without opening new sleeves. Every annual effect uses that year's group-specific clean reference.

Dense annual columns show ΔNet. Sparse annual cells keep the order ΔRankIC / Rep. / ΔNet, exposing the prediction–selection–return sequence at each dose. Rep. is percent and ΔNet is pp; zero and full-exposure endpoints are retained.

Table 20: **Annual results reveal how the learning entry point changes portfolio effects.** CL and LL are compared within each feature group, lead, and year. ΔNet (pp).

			2022		2023		2024	
Universe / model	Features	L	CL	LL	CL	LL	CL	LL
CSI 300 Ridge	All	1	-64.15	+173.77	-42.59	+160.06	-37.05	+146.85
		2	-82.16	+241.68	-48.75	+220.89	-52.13	+197.55
		5	-211.74	+423.00	-122.91	+375.70	-133.49	+346.95
		10	+79.34	+422.64	+9.28	+375.34	+32.87	+348.48
CSI 300 Ridge	Returns	1	-55.04	+153.98	-49.75	+172.91	-50.75	+135.49
		2	-105.98	+199.54	-101.86	+213.78	-101.63	+174.08
		5	-259.15	+413.06	-250.51	+392.87	-246.38	+346.47
		10	+209.40	+412.29	+111.94	+392.47	+68.53	+345.31
CSI 300 Ridge	State	1	+20.91	+45.04	+28.03	+54.89	+27.63	+43.32
		2	+36.34	+67.32	+49.70	+81.28	+57.56	+69.22
		5	+75.73	+118.96	+124.99	+133.14	+122.73	+129.38
		10	+33.48	+89.23	+49.38	+111.64	+50.91	+98.48
CSI 300 Ridge	Volatility	1	+11.37	+13.42	+18.30	+17.93	+21.19	+21.88
		2	+23.30	+33.33	+36.24	+36.39	+45.73	+47.02
		5	+59.50	+96.61	+109.66	+114.53	+101.86	+102.39
		10	+12.91	+23.26	+34.71	+71.27	+37.87	+44.94
CSI 300 Ridge	Volume	1	+10.10	+47.00	+24.32	+39.96	+19.02	+32.71
		2	+24.57	+86.76	+49.47	+71.23	+48.49	+68.07
		5	+55.54	+133.46	+98.41	+125.00	+113.69	+128.71
		10	+3.91	+55.30	+18.30	+68.86	+49.12	+82.91
CSI 300 Ridge	Range	1	+50.38	+56.42	+36.20	+48.74	+27.28	+41.33
		2	+73.49	+76.09	+56.41	+67.03	+49.96	+59.08
		5	+104.80	+126.41	+114.70	+116.69	+110.19	+112.13
		10	+45.27	+69.12	+76.17	+86.80	+72.45	+79.12
CSI 300 LightGBM	All	1	-30.30	+166.38	-6.31	+151.11	-7.30	+143.12
		2	-53.41	+237.05	-10.99	+213.15	-7.78	+191.26
		5	-125.88	+422.44	-53.49	+369.08	-33.10	+345.17
		10	+1.37	+422.35	-14.81	+368.81	+10.43	+347.24
CSI 300 LightGBM	Returns	1	-25.57	+146.74	-15.40	+146.36	-14.32	+129.89
		2	-50.41	+194.13	-35.06	+188.49	-29.95	+170.65
		5	-124.79	+416.11	-100.59	+369.04	-80.33	+344.15
		10	+17.47	+415.83	-19.96	+368.80	-1.78	+345.31
CSI 300 LightGBM	State	1	+5.98	+44.09	+13.92	+44.17	+6.56	+29.95
		2	+6.33	+80.60	+25.94	+75.90	+12.02	+58.95
		5	+14.23	+133.56	+47.85	+132.91	+23.47	+109.23
		10	+29.31	+91.00	+18.49	+111.80	+16.54	+84.47
CSI 300 LightGBM	Volatility	1	+1.54	-1.31	+8.36	+12.36	+4.84	+5.38
		2	+4.14	+13.60	+17.91	+34.78	+8.67	+19.43
		5	+7.99	+53.55	+53.08	+107.18	+18.08	+60.28
		10	+6.32	+17.94	+20.72	+52.28	+2.74	+25.85
CSI 300 LightGBM	Volume	1	+11.49	+42.36	+12.51	+43.61	+10.52	+26.34
		2	+21.96	+82.15	+32.31	+79.05	+26.14	+64.86
		5	+33.74	+130.42	+63.49	+135.64	+58.69	+115.94
		10	+2.29	+66.46	+15.34	+79.75	+20.72	+75.88

Continued on next page

			2022		2023		2024	
Universe / model	Features	L	CL	LL	CL	LL	CL	LL
CSI 300 LightGBM	Range	1	+8.89	+46.94	+15.73	+46.59	+12.57	+36.83
		2	+14.58	+60.10	+24.04	+59.93	+20.94	+52.63
		5	+28.04	+106.90	+50.03	+120.39	+37.04	+104.75
		10	+25.65	+58.99	+27.55	+82.25	+26.05	+69.74
CSI 500 Ridge	All	1	-65.20	+174.11	-42.28	+156.70	-29.06	+152.06
		2	-68.06	+245.13	-43.72	+213.45	-36.66	+203.97
		5	-122.14	+447.78	-57.57	+384.08	-83.28	+368.85
		10	+146.35	+442.64	+76.25	+383.21	+54.77	+367.10
CSI 500 Ridge	Returns	1	-57.05	+168.76	-66.75	+163.67	-50.32	+146.61
		2	-92.24	+219.32	-107.71	+204.80	-92.86	+189.91
		5	-193.50	+451.96	-226.95	+399.48	-223.03	+374.50
		10	+123.31	+446.99	+145.49	+398.60	+98.41	+372.74
CSI 500 Ridge	State	1	+9.37	+46.61	+13.52	+50.26	+15.47	+44.54
		2	+14.95	+104.42	+30.46	+88.12	+36.32	+80.45
		5	+77.03	+201.17	+128.02	+163.43	+113.53	+152.81
		10	+83.83	+160.13	+84.93	+130.55	+78.05	+124.27
CSI 500 Ridge	Volatility	1	+13.57	+14.24	+17.00	+14.98	+18.32	+20.86
		2	+35.27	+42.80	+39.17	+36.18	+42.36	+45.70
		5	+117.38	+63.72	+129.65	+128.50	+107.75	+112.58
		10	+107.54	+97.17	+76.87	+92.17	+69.75	+77.87
CSI 500 Ridge	Volume	1	-2.57	+60.78	+17.73	+53.48	+12.37	+41.83
		2	+5.19	+126.55	+47.63	+98.72	+37.05	+85.87
		5	+63.66	+234.09	+147.79	+170.31	+128.67	+157.85
		10	+24.31	+136.96	+67.90	+105.77	+70.49	+104.83
CSI 500 Ridge	Range	1	+15.60	+65.11	+14.46	+48.96	+12.61	+44.51
		2	+41.73	+94.60	+35.19	+73.26	+32.48	+66.22
		5	+164.85	+181.09	+130.61	+138.26	+113.69	+133.19
		10	+147.03	+136.46	+104.19	+104.01	+104.93	+102.01
CSI 500 LightGBM	All	1	-43.65	+175.18	-11.84	+145.44	-15.85	+153.41
		2	-40.46	+251.65	+3.02	+205.14	-10.85	+212.20
		5	-61.97	+454.97	+10.23	+375.20	-12.05	+375.72
		10	+49.05	+452.33	+48.22	+374.25	+60.63	+377.06
CSI 500 LightGBM	Returns	1	-24.28	+164.44	-6.41	+144.53	-16.12	+142.86
		2	-42.86	+214.67	-8.51	+184.95	-26.90	+187.05
		5	-76.44	+452.38	-9.07	+383.96	-48.46	+375.61
		10	+37.25	+449.23	+67.29	+382.38	+38.09	+375.34
CSI 500 LightGBM	State	1	+3.42	+57.20	+10.58	+53.52	+6.51	+43.67
		2	+12.26	+131.95	+27.98	+104.25	+14.86	+89.79
		5	+32.64	+239.90	+67.15	+190.42	+38.79	+169.61
		10	+51.28	+191.00	+42.29	+153.90	+31.59	+137.92
CSI 500 LightGBM	Volatility	1	+6.69	+5.23	+8.29	+6.73	+3.55	+5.06
		2	+16.18	+19.71	+19.61	+25.90	+10.88	+27.09
		5	+55.36	+67.60	+64.64	+102.46	+38.89	+78.91
		10	+49.79	+73.27	+43.20	+81.43	+26.68	+66.62
CSI 500 LightGBM	Volume	1	+13.38	+57.92	+11.53	+47.92	+9.50	+39.75
		2	+35.51	+130.62	+30.63	+100.05	+27.33	+83.82
		5	+82.96	+236.88	+81.52	+177.55	+73.55	+152.03
		10	+39.96	+146.50	+41.48	+118.12	+34.93	+107.99
CSI 500 LightGBM	Range	1	+8.01	+49.04	+11.85	+37.48	+6.73	+35.74
		2	+21.21	+72.19	+25.63	+58.67	+19.10	+53.41
		5	+40.48	+147.08	+72.41	+130.98	+43.51	+125.88
		10	+29.43	+131.17	+47.08	+97.17	+40.20	+91.21

Table 21: **Sparse exposure changes selections even when annual predictive shifts are small.** The complete annual breakdown retains both the recurring patterns and the exceptions.

ΔRankIC / Rep. (%) / ΔNet (pp)					
Universe / model	Exposure	Dose	2022	2023	2024
CSI 300 Ridge	Dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0011 / 0.4 / -0.52	-0.0006 / 0.3 / -0.43	-0.0006 / 0.4 / -0.27
		5%	-0.0052 / 2.2 / -3.09	-0.0039 / 2.1 / -2.28	-0.0042 / 2.1 / -1.74
		10%	-0.0098 / 4.2 / -5.89	-0.0076 / 4.2 / -4.35	-0.0091 / 4.1 / -3.53
		25%	-0.0256 / 10.6 / -15.23	-0.0193 / 10.4 / -10.58	-0.0233 / 10.0 / -9.18
		100%	-0.1067 / 42.6 / -64.15	-0.0762 / 42.0 / -42.59	-0.0924 / 40.5 / -37.05
CSI 300 Ridge	Stock-dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0010 / 0.9 / -0.58	-0.0009 / 0.9 / -0.37	-0.0009 / 0.9 / -0.32
		5%	-0.0050 / 4.2 / -2.93	-0.0040 / 3.8 / -2.05	-0.0045 / 4.1 / -1.35
		10%	-0.0101 / 7.7 / -6.07	-0.0077 / 6.9 / -4.30	-0.0091 / 7.5 / -2.98
		25%	-0.0256 / 16.8 / -15.21	-0.0188 / 15.1 / -10.47	-0.0232 / 16.1 / -8.89
		100%	-0.1067 / 42.6 / -64.15	-0.0762 / 42.0 / -42.59	-0.0924 / 40.5 / -37.05
CSI 300 Ridge	Features	0	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1	-0.0055 / 13.1 / -2.70	-0.0039 / 12.1 / +0.11	-0.0054 / 10.2 / +0.09
		2	-0.0118 / 18.3 / -7.27	-0.0096 / 17.2 / -3.12	-0.0104 / 15.9 / -0.94
		4	-0.0220 / 26.8 / -15.58	-0.0182 / 25.4 / -7.76	-0.0197 / 24.7 / -2.89
		8	-0.0386 / 36.8 / -25.70	-0.0291 / 35.6 / -13.50	-0.0359 / 35.9 / -8.54
		19	-0.1067 / 42.6 / -64.15	-0.0762 / 42.0 / -42.59	-0.0924 / 40.5 / -37.05
CSI 300 LightGBM	Dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0006 / 0.4 / -0.34	+0.0002 / 0.4 / 0.00	+0.0002 / 0.5 / +0.06
		5%	-0.0028 / 2.5 / -1.68	-0.0001 / 2.6 / -0.19	-0.0006 / 2.9 / -0.04
		10%	-0.0056 / 5.0 / -2.92	+0.0001 / 5.2 / -0.44	-0.0019 / 5.7 / -0.40
		25%	-0.0151 / 12.5 / -7.71	-0.0003 / 12.8 / -1.05	-0.0066 / 13.7 / -1.69
		100%	-0.0631 / 50.3 / -30.30	-0.0023 / 51.6 / -6.31	-0.0269 / 55.6 / -7.30
CSI 300 LightGBM	Stock-dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0005 / 0.9 / -0.13	0.0000 / 1.0 / +0.02	-0.0003 / 1.1 / -0.12
		5%	-0.0027 / 4.1 / -0.95	-0.0003 / 4.0 / -0.56	-0.0015 / 4.5 / -0.18
		10%	-0.0057 / 7.5 / -2.35	-0.0004 / 7.4 / -1.13	-0.0027 / 8.3 / -0.60
		25%	-0.0153 / 16.4 / -7.03	-0.0009 / 16.4 / -2.07	-0.0067 / 18.1 / -1.96
		100%	-0.0631 / 50.3 / -30.30	-0.0023 / 51.6 / -6.31	-0.0269 / 55.6 / -7.30
CSI 300 LightGBM	Features	0	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1	-0.0033 / 11.8 / -0.91	+0.0005 / 11.4 / -0.66	-0.0012 / 11.5 / -0.46
		2	-0.0047 / 18.0 / -0.78	+0.0026 / 17.5 / +0.14	-0.0010 / 17.8 / +0.37
		4	-0.0080 / 26.7 / -1.29	+0.0050 / 26.4 / +1.63	-0.0019 / 27.0 / +1.60
		8	-0.0159 / 37.2 / -5.57	+0.0073 / 37.8 / +1.02	-0.0036 / 39.0 / +3.22
		19	-0.0631 / 50.3 / -30.30	-0.0023 / 51.6 / -6.31	-0.0269 / 55.6 / -7.30
CSI 500 Ridge	Dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0010 / 0.3 / -0.51	-0.0007 / 0.3 / -0.34	-0.0005 / 0.3 / -0.18
		5%	-0.0060 / 2.0 / -3.22	-0.0038 / 1.6 / -2.28	-0.0031 / 1.6 / -1.34
		10%	-0.0114 / 4.0 / -6.04	-0.0071 / 3.2 / -4.24	-0.0066 / 3.2 / -2.78
		25%	-0.0299 / 9.9 / -15.47	-0.0185 / 8.2 / -10.83	-0.0169 / 7.6 / -6.85
		100%	-0.1227 / 39.7 / -65.20	-0.0713 / 33.1 / -42.28	-0.0688 / 30.8 / -29.06
CSI 500 Ridge	Stock-dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0011 / 0.9 / -0.67	-0.0007 / 0.7 / -0.45	-0.0007 / 0.8 / -0.30
		5%	-0.0060 / 3.9 / -3.28	-0.0034 / 2.9 / -2.02	-0.0033 / 3.4 / -1.09
		10%	-0.0117 / 7.2 / -6.56	-0.0070 / 5.3 / -4.26	-0.0067 / 6.2 / -2.29
		25%	-0.0298 / 15.7 / -16.38	-0.0175 / 11.7 / -10.83	-0.0167 / 13.2 / -6.41
		100%	-0.1227 / 39.7 / -65.20	-0.0713 / 33.1 / -42.28	-0.0688 / 30.8 / -29.06

Continued on next page

ΔRankIC / Rep. (%) / ΔNet (pp)					
Universe / model	Exposure	Dose	2022	2023	2024
CSI 500 Ridge	Features	0	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1	-0.0067 / 8.6 / -2.44	-0.0042 / 7.0 / -0.98	-0.0039 / 7.6 / -1.19
		2	-0.0138 / 13.4 / -5.92	-0.0087 / 9.9 / -2.49	-0.0076 / 11.7 / -2.44
		4	-0.0263 / 19.5 / -11.15	-0.0175 / 14.7 / -6.27	-0.0151 / 16.3 / -5.16
		8	-0.0522 / 25.1 / -22.46	-0.0320 / 21.4 / -13.50	-0.0289 / 22.5 / -9.76
		19	-0.1227 / 39.7 / -65.20	-0.0713 / 33.1 / -42.28	-0.0688 / 30.8 / -29.06
CSI 500 LightGBM	Dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0004 / 0.5 / -0.21	+0.0004 / 0.4 / -0.06	+0.0001 / 0.5 / -0.10
		5%	-0.0026 / 2.9 / -2.26	+0.0014 / 2.7 / -0.63	-0.0002 / 2.6 / -0.74
		10%	-0.0048 / 5.7 / -4.23	+0.0032 / 5.3 / -1.21	-0.0005 / 5.3 / -1.58
		25%	-0.0128 / 14.2 / -11.31	+0.0083 / 13.2 / -2.66	-0.0022 / 12.7 / -3.89
		100%	-0.0536 / 57.1 / -43.65	+0.0317 / 52.9 / -11.84	-0.0114 / 51.7 / -15.85
CSI 500 LightGBM	Stock-dates	0%	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1%	-0.0004 / 1.1 / -0.27	+0.0001 / 1.0 / -0.28	-0.0002 / 1.0 / -0.07
		5%	-0.0026 / 4.6 / -2.09	+0.0014 / 4.0 / -0.62	-0.0006 / 4.1 / -0.62
		10%	-0.0053 / 8.4 / -4.56	+0.0030 / 7.3 / -1.09	-0.0014 / 7.7 / -1.22
		25%	-0.0132 / 18.9 / -11.02	+0.0080 / 16.5 / -3.28	-0.0032 / 17.2 / -3.28
		100%	-0.0536 / 57.1 / -43.65	+0.0317 / 52.9 / -11.84	-0.0114 / 51.7 / -15.85
CSI 500 LightGBM	Features	0	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00	0.0000 / 0.0 / 0.00
		1	-0.0023 / 11.7 / -1.67	+0.0027 / 10.9 / 0.00	+0.0004 / 11.3 / +1.22
		2	-0.0042 / 18.1 / -2.64	+0.0065 / 16.7 / +0.86	+0.0018 / 17.1 / +1.98
		4	-0.0085 / 27.3 / -5.42	+0.0118 / 25.3 / +1.72	+0.0025 / 25.6 / +3.69
		8	-0.0171 / 39.7 / -12.94	+0.0224 / 37.2 / +1.32	+0.0043 / 36.8 / +3.62
		19	-0.0536 / 57.1 / -43.65	+0.0317 / 52.9 / -11.84	-0.0114 / 51.7 / -15.85

C.2 Uncertainty across the complete observation grid

Feature groups, leads, training/scoring cells, sparse doses and masks were fixed before computing the mechanism experiment. All and State at $L = 1$ were chosen for emphasis after observing the grid (Table 29); Appendix C.1 retains the other comparisons. Here we report the previously computed pointwise descriptive 95% HAC(20) intervals for the full dense grid and all single-feature replacements. They describe temporal variability, without simultaneous coverage or adjustment for selecting main-text examples after observing results.

Interval construction. For a paired daily difference d_t, let $n = 726$, $\bar{d} = n^{-1}\sum_t d_t$, and $\widehat{\gamma}_j = n^{-1}\sum_{t=j+1}^{n}(d_t - \bar{d})(d_{t-j} - \bar{d})$. The Newey–West estimator (Newey & West, 1987) gives

$$\widehat{\mathrm{se}}_b = a\left[\frac{1}{n}\left\{\widehat{\gamma}_0 + 2\sum_{j=1}^{b}\left(1 - \frac{j}{b+1}\right)\widehat{\gamma}_j\right\}\right]^{1/2}, \qquad a\bar{d} \pm 1.96\,\widehat{\mathrm{se}}_b. \tag{15}$$

For net returns, d_t is the intervention-minus-clean daily net return and $a = 252$; the displayed effect and bounds are multiplied by 100 to express annualized arithmetic percentage points. For RankIC, d_t is the paired daily cross-sectional Spearman difference and $a = 1$. The original summaries use $b = 20$ and the full daily calendar, rather than treating the three years as independent observations.

Complete condition-level intervals. Tables 22 and 23 cover every group, lead and non-baseline training/scoring cell for each model and universe. Tables 24 and 25 retain all 19 single-feature replacements in the original feature order. Point estimates match the preceding mechanism tables.

Each cell places the point estimate above its 95% CI; CC would have identically zero paired change and interval. This common layout applies to the dense, single-feature, sparse, and interaction tables. The single-feature comparison advances only the named feature by one session under the frozen All-feature model.

Table 22: **Economic effects depend on both feature content and learning entry.** The full ΔNet (pp) grid retains positive, negative, and unresolved responses.

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
All	1	CL	−47.93 [−54.06, −41.80]	−14.64 [−20.03, −9.25]	−45.51 [−51.97, −39.05]	−23.78 [−30.55, −17.01]
		LC	−6.42 [−21.01, +8.18]	−7.70 [−18.13, +2.72]	−13.53 [−28.26, +1.20]	−12.32 [−21.29, −3.34]
		LL	+160.23 [+143.59, +176.87]	+153.53 [+137.69, +169.38]	+160.96 [+143.09, +178.82]	+158.01 [+142.39, +173.64]
All	2	CL	−61.02 [−70.66, −51.37]	−24.06 [−34.24, −13.88]	−49.48 [−58.31, −40.66]	−16.10 [−25.85, −6.35]
		LC	−6.70 [−23.07, +9.67]	−9.47 [−22.09, +3.16]	−15.23 [−31.80, +1.34]	−12.47 [−22.98, −1.95]
		LL	+220.04 [+198.64, +241.44]	+213.82 [+192.97, +234.67]	+220.85 [+197.78, +243.91]	+222.99 [+201.66, +244.33]
All	5	CL	−156.05 [−181.93, −130.17]	−70.82 [−92.68, −48.97]	−87.66 [−108.70, −66.63]	−21.26 [−41.25, −1.27]
		LC	−7.49 [−26.41, +11.43]	−10.99 [−26.12, +4.13]	−15.83 [−34.58, +2.93]	−14.56 [−27.00, −2.12]
		LL	+381.88 [+348.77, +415.00]	+378.90 [+344.37, +413.42]	+400.23 [+365.06, +435.40]	+401.96 [+366.24, +437.69]
All	10	CL	+40.50 [+25.13, +55.86]	−1.00 [−18.25, +16.24]	+92.46 [+72.62, +112.29]	+52.63 [+31.02, +74.25]
		LC	−5.52 [−19.60, +8.56]	−10.69 [−23.59, +2.21]	−9.49 [−23.61, +4.63]	−8.94 [−18.55, +0.68]
		LL	+382.16 [+348.69, +415.62]	+379.47 [+343.81, +415.12]	+397.65 [+362.94, +432.36]	+401.21 [+364.90, +437.53]
Returns	1	CL	−51.85 [−57.96, −45.74]	−18.43 [−23.55, −13.32]	−58.04 [−64.18, −51.90]	−15.60 [−21.09, −10.12]
		LC	−3.25 [−20.48, +13.97]	−9.58 [−21.73, +2.57]	−4.58 [−21.26, +12.11]	−8.13 [−19.21, +2.96]
		LL	+154.13 [+136.77, +171.49]	+140.99 [+126.93, +155.06]	+159.68 [+140.17, +179.20]	+150.61 [+134.38, +166.85]
Returns	2	CL	−103.16 [−115.51, −90.81]	−38.47 [−46.18, −30.77]	−97.61 [−108.80, −86.41]	−26.09 [−35.11, −17.07]
		LC	−3.82 [−23.36, +15.73]	−12.05 [−26.03, +1.92]	−7.81 [−27.07, +11.44]	−11.24 [−24.46, +1.98]
		LL	+195.80 [+175.69, +215.91]	+184.42 [+166.21, +202.63]	+204.68 [+181.23, +228.12]	+195.56 [+174.94, +216.17]
Returns	5	CL	−252.01 [−280.84, −223.19]	−101.90 [−120.21, −83.59]	−214.49 [−239.37, −189.62]	−44.66 [−63.36, −25.95]
		LC	−6.87 [−29.67, +15.93]	−14.29 [−31.15, +2.56]	−7.73 [−29.39, +13.93]	−12.76 [−28.19, +2.68]
		LL	+384.13 [+351.66, +416.61]	+376.43 [+343.77, +409.09]	+408.65 [+371.88, +445.41]	+403.98 [+367.32, +440.64]
Returns	10	CL	+129.96 [+105.85, +154.06]	−1.43 [−24.30, +21.45]	+122.40 [+107.54, +137.26]	+47.54 [+23.09, +72.00]
		LC	−3.46 [−18.99, +12.07]	−12.12 [−25.10, +0.86]	−0.42 [−17.05, +16.21]	−6.71 [−18.60, +5.18]
		LL	+383.36 [+350.96, +415.75]	+376.65 [+343.40, +409.89]	+406.11 [+369.87, +442.35]	+402.32 [+365.56, +439.07]
State	1	CL	+25.52 [+19.31, +31.73]	+8.82 [+5.53, +12.12]	+12.79 [+9.25, +16.32]	+6.84 [+4.18, +9.50]
		LC	+4.42 [−1.92, +10.75]	−1.55 [−7.01, +3.91]	−3.51 [−12.99, +5.98]	−2.63 [−8.38, +3.12]
		LL	+47.75 [+36.53, +58.96]	+39.40 [+30.13, +48.68]	+47.14 [+35.74, +58.53]	+51.46 [+41.97, +60.95]

Continued on next page

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
State	2	CL	+47.86 [+36.09, +59.64]	+14.76 [+8.57, +20.95]	+27.25 [+20.08, +34.42]	+18.36 [+12.68, +24.05]
		LC	+4.62 [−3.38, +12.61]	−2.36 [−8.21, +3.50]	−1.43 [−13.10, +10.24]	−2.99 [−10.79, +4.81]
		LL	+72.60 [+57.30, +87.91]	+71.82 [+57.68, +85.96]	+91.00 [+75.59, +106.41]	+108.66 [+93.12, +124.20]
State	5	CL	+107.82 [+80.99, +134.64]	+28.52 [+14.61, +42.42]	+106.19 [+86.10, +126.29]	+46.19 [+34.08, +58.30]
		LC	+8.14 [−0.52, +16.80]	−5.44 [−13.82, +2.93]	−0.74 [−13.01, +11.53]	−2.87 [−12.38, +6.63]
		LL	+127.16 [+101.40, +152.92]	+125.23 [+100.07, +150.40]	+172.47 [+148.74, +196.20]	+199.98 [+173.78, +226.18]
State	10	CL	+44.59 [+29.38, +59.80]	+21.45 [+11.15, +31.75]	+82.27 [+65.98, +98.55]	+41.72 [+32.80, +50.64]
		LC	+7.44 [−3.44, +18.33]	−7.26 [−16.96, +2.44]	+2.10 [−8.18, +12.38]	+2.10 [−7.52, +11.72]
		LL	+99.79 [+78.93, +120.64]	+95.76 [+73.62, +117.89]	+138.32 [+117.91, +158.73]	+160.94 [+137.06, +184.82]
Volatility	1	CL	+16.96 [+11.42, +22.49]	+4.92 [+1.45, +8.38]	+16.30 [+11.52, +21.07]	+6.18 [+3.76, +8.60]
		LC	−1.35 [−5.60, +2.90]	−3.42 [−6.92, +0.09]	−1.61 [−3.81, +0.60]	−3.50 [−6.20, −0.80]
		LL	+17.74 [+10.61, +24.88]	+5.48 [+0.33, +10.63]	+16.69 [+10.82, +22.56]	+5.67 [+2.01, +9.33]
Volatility	2	CL	+35.09 [+24.17, +46.01]	+10.24 [+4.45, +16.02]	+38.94 [+29.25, +48.62]	+15.56 [+10.73, +20.39]
		LC	−2.71 [−7.79, +2.37]	−5.39 [−10.92, +0.13]	−2.08 [−4.68, +0.52]	−4.08 [−7.88, −0.28]
		LL	+38.91 [+26.44, +51.38]	+22.60 [+14.13, +31.08]	+41.56 [+30.81, +52.32]	+24.24 [+16.86, +31.61]
Volatility	5	CL	+90.34 [+67.09, +113.60]	+26.39 [+15.60, +37.17]	+118.26 [+99.93, +136.59]	+52.97 [+42.56, +63.37]
		LC	−3.03 [−8.86, +2.80]	−7.33 [−14.86, +0.20]	+0.70 [−2.70, +4.09]	−4.63 [−10.01, +0.75]
		LL	+104.51 [+79.55, +129.47]	+73.67 [+56.43, +90.91]	+101.60 [+81.99, +121.21]	+82.99 [+68.49, +97.50]
Volatility	10	CL	+28.50 [+10.76, +46.23]	+9.93 [−0.56, +20.42]	+84.72 [+70.33, +99.10]	+39.89 [+31.28, +48.51]
		LC	−0.26 [−12.16, +11.64]	−6.44 [−14.49, +1.60]	−1.66 [−7.84, +4.51]	−1.34 [−8.18, +5.49]
		LL	+46.49 [+26.11, +66.86]	+32.02 [+17.17, +46.88]	+89.07 [+74.07, +104.07]	+73.77 [+58.76, +88.78]
Volume	1	CL	+17.81 [+12.88, +22.74]	+11.51 [+8.12, +14.89]	+9.17 [+5.38, +12.97]	+11.47 [+9.09, +13.84]
		LC	+1.08 [−4.80, +6.96]	+0.59 [−4.15, +5.33]	−2.29 [−11.23, +6.64]	−0.52 [−4.87, +3.83]
		LL	+39.89 [+30.11, +49.66]	+37.44 [+28.81, +46.06]	+52.03 [+40.48, +63.58]	+48.53 [+40.25, +56.81]
Volume	2	CL	+40.85 [+30.35, +51.34]	+26.80 [+20.04, +33.57]	+29.95 [+21.48, +38.43]	+31.16 [+26.35, +35.96]
		LC	+1.58 [−5.01, +8.16]	+2.02 [−3.72, +7.75]	−2.03 [−11.23, +7.17]	+0.15 [−5.25, +5.54]
		LL	+75.35 [+60.84, +89.87]	+75.35 [+61.53, +89.17]	+103.71 [+87.16, +120.27]	+104.83 [+90.59, +119.07]
Volume	5	CL	+89.21 [+66.01, +112.42]	+51.97 [+38.48, +65.46]	+113.37 [+90.82, +135.93]	+79.34 [+68.37, +90.31]
		LC	+3.40 [−3.19, +9.99]	+4.94 [−2.85, +12.73]	−1.99 [−10.72, +6.75]	−1.70 [−8.90, +5.49]
		LL	+129.06 [+104.16, +153.95]	+127.34 [+102.58, +152.09]	+187.42 [+160.90, +213.93]	+188.82 [+164.34, +213.31]

Continued on next page

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
Volume	10	CL	$+23.77$ $[+10.85, +36.70]$	$+12.78$ $[+4.49, +21.08]$	$+54.23$ $[+39.17, +69.30]$	$+38.79$ $[+31.68, +45.89]$
		LC	$+4.87$ $[-5.00, +14.73]$	$+3.11$ $[-7.95, +14.17]$	-3.58 $[-13.11, +5.95]$	-1.35 $[-8.54, +5.85]$
		LL	$+69.02$ $[+50.94, +87.10]$	$+74.03$ $[+55.68, +92.37]$	$+115.86$ $[+96.35, +135.36]$	$+124.20$ $[+107.04, +141.37]$
Range	1	CL	$+37.95$ $[+29.08, +46.83]$	$+12.40$ $[+9.01, +15.78]$	$+14.22$ $[+10.72, +17.72]$	$+8.86$ $[+6.31, +11.41]$
		LC	-0.59 $[-7.65, +6.47]$	$+0.41$ $[-3.90, +4.72]$	-1.43 $[-8.60, +5.73]$	-1.70 $[-4.93, +1.53]$
		LL	$+48.83$ $[+37.07, +60.59]$	$+43.45$ $[+33.55, +53.35]$	$+52.86$ $[+41.57, +64.15]$	$+40.75$ $[+33.98, +47.52]$
Range	2	CL	$+59.95$ $[+46.82, +73.08]$	$+19.85$ $[+14.16, +25.55]$	$+36.47$ $[+29.12, +43.81]$	$+21.98$ $[+17.21, +26.75]$
		LC	-0.17 $[-5.08, +4.73]$	$+0.73$ $[-4.50, +5.97]$	-1.20 $[-7.11, +4.71]$	-2.34 $[-5.89, +1.20]$
		LL	$+67.40$ $[+52.63, +82.16]$	$+57.55$ $[+44.96, +70.14]$	$+78.03$ $[+64.38, +91.68]$	$+61.42$ $[+50.99, +71.86]$
Range	5	CL	$+109.89$ $[+85.14, +134.65]$	$+38.37$ $[+25.61, +51.12]$	$+136.38$ $[+114.71, +158.06]$	$+52.13$ $[+40.21, +64.06]$
		LC	-0.95 $[-3.36, +1.46]$	$+1.76$ $[-6.37, +9.90]$	-1.05 $[-4.75, +2.64]$	-3.40 $[-9.78, +2.99]$
		LL	$+118.41$ $[+92.73, +144.09]$	$+110.68$ $[+86.49, +134.88]$	$+150.85$ $[+126.38, +175.31]$	$+134.65$ $[+112.38, +156.91]$
Range	10	CL	$+64.63$ $[+48.90, +80.36]$	$+26.42$ $[+17.13, +35.70]$	$+118.71$ $[+99.43, +138.00]$	$+38.90$ $[+29.88, +47.92]$
		LC	-0.47 $[-3.94, +2.99]$	-0.87 $[-9.44, +7.70]$	-0.71 $[-2.80, +1.39]$	-3.64 $[-11.63, +4.36]$
		LL	$+78.35$ $[+60.50, +96.20]$	$+70.33$ $[+50.56, +90.10]$	$+114.16$ $[+96.04, +132.28]$	$+106.52$ $[+87.54, +125.50]$

Table 23: **Future information does not uniformly improve predictive ranks.** The ΔRankIC grid separates gains from deterioration and unresolved changes.

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
All	1	CL	-0.0918 $[-0.0986, -0.0849]$	-0.0307 $[-0.0428, -0.0187]$	-0.0876 $[-0.0974, -0.0778]$	-0.0111 $[-0.0258, +0.0037]$
		LC	$+0.0125$ $[-0.0219, +0.0468]$	-0.0036 $[-0.0229, +0.0157]$	-0.0072 $[-0.0376, +0.0231]$	-0.0044 $[-0.0233, +0.0144]$
		LL	$+0.3968$ $[+0.3684, +0.4252]$	$+0.3622$ $[+0.3471, +0.3774]$	$+0.3868$ $[+0.3633, +0.4104]$	$+0.3737$ $[+0.3593, +0.3880]$
All	2	CL	-0.1193 $[-0.1316, -0.1071]$	-0.0375 $[-0.0566, -0.0183]$	-0.0924 $[-0.1067, -0.0781]$	$+0.0114$ $[-0.0110, +0.0338]$
		LC	$+0.0127$ $[-0.0293, +0.0547]$	-0.0061 $[-0.0324, +0.0202]$	-0.0103 $[-0.0462, +0.0256]$	-0.0085 $[-0.0326, +0.0156]$
		LL	$+0.5200$ $[+0.4864, +0.5536]$	$+0.4936$ $[+0.4742, +0.5129]$	$+0.4986$ $[+0.4709, +0.5263]$	$+0.4947$ $[+0.4762, +0.5132]$
All	5	CL	-0.3030 $[-0.3373, -0.2687]$	-0.1318 $[-0.1692, -0.0943]$	-0.1660 $[-0.2017, -0.1303]$	$+0.0232$ $[-0.0203, +0.0666]$
		LC	$+0.0164$ $[-0.0304, +0.0632]$	-0.0030 $[-0.0350, +0.0289]$	-0.0155 $[-0.0556, +0.0247]$	-0.0153 $[-0.0447, +0.0141]$
		LL	$+0.9765$ $[+0.9476, +1.0054]$	$+0.9576$ $[+0.9438, +0.9714]$	$+0.9552$ $[+0.9313, +0.9790]$	$+0.9566$ $[+0.9410, +0.9722]$

Continued on next page

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
All	10	CL	$+0.1107$ $[+0.0763, +0.1452]$	$+0.0351$ $[+0.0065, +0.0638]$	$+0.1624$ $[+0.1270, +0.1979]$	$+0.1374$ $[+0.0975, +0.1773]$
		LC	$+0.0041$ $[-0.0339, +0.0421]$	-0.0146 $[-0.0419, +0.0128]$	-0.0163 $[-0.0483, +0.0158]$	-0.0144 $[-0.0388, +0.0100]$
		LL	$+0.9722$ $[+0.9415, +1.0028]$	$+0.9529$ $[+0.9375, +0.9683]$	$+0.9509$ $[+0.9259, +0.9758]$	$+0.9517$ $[+0.9357, +0.9677]$
Returns	1	CL	-0.1416 $[-0.1491, -0.1341]$	-0.0116 $[-0.0231, -0.0002]$	-0.1628 $[-0.1715, -0.1541]$	$+0.0066$ $[-0.0087, +0.0218]$
		LC	-0.0011 $[-0.0428, +0.0406]$	-0.0011 $[-0.0231, +0.0209]$	-0.0125 $[-0.0507, +0.0257]$	-0.0007 $[-0.0228, +0.0214]$
		LL	$+0.3605$ $[+0.3247, +0.3963]$	$+0.3420$ $[+0.3229, +0.3611]$	$+0.3561$ $[+0.3251, +0.3871]$	$+0.3510$ $[+0.3327, +0.3693]$
Returns	2	CL	-0.2715 $[-0.2857, -0.2574]$	-0.0397 $[-0.0567, -0.0228]$	-0.2672 $[-0.2812, -0.2533]$	$+0.0019$ $[-0.0205, +0.0244]$
		LC	$+0.0013$ $[-0.0476, +0.0502]$	$+0.0009$ $[-0.0275, +0.0293]$	-0.0164 $[-0.0608, +0.0280]$	-0.0060 $[-0.0341, +0.0220]$
		LL	$+0.4373$ $[+0.3979, +0.4767]$	$+0.4223$ $[+0.4006, +0.4441]$	$+0.4222$ $[+0.3866, +0.4579]$	$+0.4199$ $[+0.3974, +0.4425]$
Returns	5	CL	-0.6774 $[-0.7111, -0.6438]$	-0.1376 $[-0.1759, -0.0993]$	-0.5830 $[-0.6174, -0.5487]$	-0.0154 $[-0.0623, +0.0315]$
		LC	$+0.0014$ $[-0.0532, +0.0560]$	-0.0011 $[-0.0362, +0.0340]$	-0.0217 $[-0.0705, +0.0272]$	-0.0107 $[-0.0449, +0.0234]$
		LL	$+0.9611$ $[+0.9276, +0.9945]$	$+0.9586$ $[+0.9432, +0.9740]$	$+0.9490$ $[+0.9188, +0.9793]$	$+0.9604$ $[+0.9413, +0.9795]$
Returns	10	CL	$+0.2417$ $[+0.1996, +0.2838]$	$+0.1058$ $[+0.0634, +0.1481]$	$+0.2206$ $[+0.1972, +0.2440]$	$+0.1816$ $[+0.1300, +0.2332]$
		LC	-0.0101 $[-0.0515, +0.0313]$	-0.0127 $[-0.0403, +0.0149]$	-0.0220 $[-0.0613, +0.0172]$	-0.0099 $[-0.0371, +0.0172]$
		LL	$+0.9564$ $[+0.9216, +0.9912]$	$+0.9543$ $[+0.9371, +0.9715]$	$+0.9438$ $[+0.9126, +0.9751]$	$+0.9555$ $[+0.9358, +0.9751]$
State	1	CL	$+0.0281$ $[+0.0216, +0.0347]$	$+0.0128$ $[+0.0078, +0.0178]$	$+0.0121$ $[+0.0076, +0.0166]$	$+0.0151$ $[+0.0097, +0.0205]$
		LC	$+0.0094$ $[-0.0023, +0.0211]$	$+0.0040$ $[-0.0016, +0.0096]$	-0.0015 $[-0.0158, +0.0127]$	$+0.0044$ $[-0.0031, +0.0119]$
		LL	$+0.0626$ $[+0.0461, +0.0791]$	$+0.0552$ $[+0.0448, +0.0657]$	$+0.0666$ $[+0.0493, +0.0838]$	$+0.0734$ $[+0.0625, +0.0842]$
State	2	CL	$+0.0585$ $[+0.0456, +0.0715]$	$+0.0226$ $[+0.0128, +0.0324]$	$+0.0311$ $[+0.0211, +0.0410]$	$+0.0384$ $[+0.0273, +0.0495]$
		LC	$+0.0089$ $[-0.0082, +0.0259]$	$+0.0093$ $[-0.0016, +0.0201]$	-0.0008 $[-0.0208, +0.0193]$	$+0.0125$ $[-0.0023, +0.0273]$
		LL	$+0.1136$ $[+0.0890, +0.1383]$	$+0.1194$ $[+0.0988, +0.1401]$	$+0.1409$ $[+0.1158, +0.1659]$	$+0.1687$ $[+0.1468, +0.1907]$
State	5	CL	$+0.1684$ $[+0.1290, +0.2079]$	$+0.0533$ $[+0.0309, +0.0757]$	$+0.1367$ $[+0.1065, +0.1670]$	$+0.1115$ $[+0.0854, +0.1377]$
		LC	$+0.0094$ $[-0.0120, +0.0308]$	$+0.0112$ $[-0.0055, +0.0278]$	$+0.0001$ $[-0.0218, +0.0221]$	$+0.0092$ $[-0.0091, +0.0275]$
		LL	$+0.1922$ $[+0.1555, +0.2290]$	$+0.2062$ $[+0.1722, +0.2402]$	$+0.2451$ $[+0.2096, +0.2807]$	$+0.2950$ $[+0.2631, +0.3270]$
State	10	CL	$+0.0879$ $[+0.0560, +0.1198]$	$+0.0348$ $[+0.0151, +0.0546]$	$+0.1171$ $[+0.0881, +0.1462]$	$+0.0842$ $[+0.0638, +0.1046]$
		LC	$+0.0132$ $[-0.0107, +0.0370]$	$+0.0045$ $[-0.0137, +0.0228]$	$+0.0038$ $[-0.0171, +0.0248]$	$+0.0149$ $[-0.0034, +0.0331]$
		LL	$+0.1764$ $[+0.1420, +0.2108]$	$+0.1637$ $[+0.1323, +0.1951]$	$+0.2335$ $[+0.2031, +0.2640]$	$+0.2556$ $[+0.2264, +0.2847]$
Volatility	1	CL	$+0.0242$ $[+0.0179, +0.0305]$	$+0.0082$ $[+0.0036, +0.0128]$	$+0.0199$ $[+0.0153, +0.0245]$	$+0.0103$ $[+0.0062, +0.0144]$
		LC	-0.0117 $[-0.0178, -0.0056]$	-0.0021 $[-0.0078, +0.0036]$	-0.0069 $[-0.0102, -0.0037]$	-0.0011 $[-0.0046, +0.0023]$
		LL	$+0.0154$ $[+0.0061, +0.0246]$	$+0.0133$ $[+0.0060, +0.0207]$	$+0.0156$ $[+0.0091, +0.0220]$	$+0.0120$ $[+0.0060, +0.0180]$

Continued on next page

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
Volatility	2	CL	$+0.0552$ $[+0.0438, +0.0667]$	$+0.0180$ $[+0.0092, +0.0268]$	$+0.0477$ $[+0.0392, +0.0561]$	$+0.0274$ $[+0.0199, +0.0350]$
		LC	-0.0171 $[-0.0258, -0.0085]$	-0.0017 $[-0.0110, +0.0077]$	-0.0103 $[-0.0150, -0.0056]$	$+0.0002$ $[-0.0050, +0.0053]$
		LL	$+0.0461$ $[+0.0310, +0.0612]$	$+0.0416$ $[+0.0283, +0.0549]$	$+0.0449$ $[+0.0341, +0.0557]$	$+0.0400$ $[+0.0296, +0.0503]$
Volatility	5	CL	$+0.1626$ $[+0.1291, +0.1962]$	$+0.0620$ $[+0.0401, +0.0838]$	$+0.1578$ $[+0.1362, +0.1793]$	$+0.1005$ $[+0.0816, +0.1193]$
		LC	-0.0236 $[-0.0355, -0.0117]$	-0.0091 $[-0.0255, +0.0073]$	-0.0136 $[-0.0215, -0.0056]$	$+0.0071$ $[-0.0041, +0.0183]$
		LL	$+0.1594$ $[+0.1228, +0.1961]$	$+0.1366$ $[+0.1053, +0.1679]$	$+0.1303$ $[+0.1043, +0.1562]$	$+0.1317$ $[+0.1077, +0.1557]$
Volatility	10	CL	$+0.0340$ $[+0.0007, +0.0674]$	$+0.0112$ $[-0.0082, +0.0306]$	$+0.0981$ $[+0.0742, +0.1220]$	$+0.0492$ $[+0.0321, +0.0664]$
		LC	-0.0326 $[-0.0610, -0.0041]$	-0.0165 $[-0.0338, +0.0008]$	-0.0283 $[-0.0432, -0.0134]$	$+0.0134$ $[-0.0040, +0.0308]$
		LL	$+0.0441$ $[+0.0080, +0.0803]$	$+0.0395$ $[+0.0138, +0.0652]$	$+0.0823$ $[+0.0567, +0.1080]$	$+0.0817$ $[+0.0610, +0.1024]$
Volume	1	CL	$+0.0198$ $[+0.0148, +0.0248]$	$+0.0144$ $[+0.0097, +0.0192]$	-0.0041 $[-0.0101, +0.0019]$	$+0.0123$ $[+0.0088, +0.0157]$
		LC	$+0.0182$ $[+0.0031, +0.0332]$	$+0.0009$ $[-0.0061, +0.0078]$	-0.0001 $[-0.0166, +0.0163]$	-0.0123 $[-0.0192, -0.0053]$
		LL	$+0.0730$ $[+0.0556, +0.0904]$	$+0.0513$ $[+0.0396, +0.0631]$	$+0.0749$ $[+0.0570, +0.0928]$	$+0.0541$ $[+0.0443, +0.0639]$
Volume	2	CL	$+0.0485$ $[+0.0378, +0.0591]$	$+0.0356$ $[+0.0267, +0.0445]$	$+0.0113$ $[-0.0008, +0.0235]$	$+0.0393$ $[+0.0320, +0.0466]$
		LC	$+0.0228$ $[+0.0043, +0.0412]$	$+0.0041$ $[-0.0046, +0.0127]$	-0.0007 $[-0.0188, +0.0173]$	-0.0142 $[-0.0239, -0.0046]$
		LL	$+0.1373$ $[+0.1134, +0.1613]$	$+0.1109$ $[+0.0912, +0.1306]$	$+0.1578$ $[+0.1347, +0.1808]$	$+0.1421$ $[+0.1235, +0.1607]$
Volume	5	CL	$+0.1461$ $[+0.1163, +0.1760]$	$+0.0766$ $[+0.0580, +0.0952]$	$+0.1481$ $[+0.1159, +0.1803]$	$+0.1311$ $[+0.1135, +0.1487]$
		LC	$+0.0260$ $[+0.0037, +0.0482]$	$+0.0065$ $[-0.0053, +0.0184]$	-0.0032 $[-0.0213, +0.0149]$	-0.0243 $[-0.0336, -0.0151]$
		LL	$+0.2423$ $[+0.2047, +0.2800]$	$+0.2202$ $[+0.1857, +0.2547]$	$+0.2952$ $[+0.2599, +0.3306]$	$+0.2806$ $[+0.2495, +0.3118]$
Volume	10	CL	$+0.0653$ $[+0.0398, +0.0909]$	$+0.0268$ $[+0.0113, +0.0422]$	$+0.1324$ $[+0.1055, +0.1593]$	$+0.0837$ $[+0.0686, +0.0988]$
		LC	$+0.0275$ $[-0.0007, +0.0556]$	$+0.0127$ $[-0.0044, +0.0299]$	-0.0033 $[-0.0250, +0.0183]$	-0.0165 $[-0.0272, -0.0059]$
		LL	$+0.1444$ $[+0.1034, +0.1854]$	$+0.1422$ $[+0.1099, +0.1744]$	$+0.1995$ $[+0.1698, +0.2292]$	$+0.2105$ $[+0.1874, +0.2335]$
Range	1	CL	$+0.0290$ $[+0.0190, +0.0390]$	$+0.0128$ $[+0.0072, +0.0185]$	$+0.0146$ $[+0.0107, +0.0185]$	$+0.0089$ $[+0.0047, +0.0132]$
		LC	$+0.0368$ $[+0.0215, +0.0522]$	$+0.0040$ $[-0.0015, +0.0094]$	$+0.0215$ $[+0.0096, +0.0333]$	$+0.0008$ $[-0.0040, +0.0055]$
		LL	$+0.0762$ $[+0.0561, +0.0962]$	$+0.0425$ $[+0.0326, +0.0525]$	$+0.0656$ $[+0.0498, +0.0814]$	$+0.0345$ $[+0.0282, +0.0408]$
Range	2	CL	$+0.0570$ $[+0.0417, +0.0722]$	$+0.0266$ $[+0.0169, +0.0364]$	$+0.0398$ $[+0.0314, +0.0482]$	$+0.0287$ $[+0.0211, +0.0364]$
		LC	$+0.0187$ $[+0.0108, +0.0265]$	-0.0013 $[-0.0078, +0.0053]$	$+0.0126$ $[+0.0038, +0.0214]$	$+0.0014$ $[-0.0043, +0.0071]$
		LL	$+0.0833$ $[+0.0638, +0.1028]$	$+0.0517$ $[+0.0386, +0.0649]$	$+0.0933$ $[+0.0759, +0.1107]$	$+0.0517$ $[+0.0427, +0.0608]$
Range	5	CL	$+0.1260$ $[+0.0934, +0.1586]$	$+0.0585$ $[+0.0367, +0.0804]$	$+0.1540$ $[+0.1263, +0.1816]$	$+0.0907$ $[+0.0690, +0.1123]$
		LC	-0.0042 $[-0.0072, -0.0012]$	-0.0026 $[-0.0149, +0.0096]$	$+0.0024$ $[-0.0014, +0.0063]$	$+0.0234$ $[+0.0070, +0.0397]$
		LL	$+0.1278$ $[+0.0953, +0.1604]$	$+0.1181$ $[+0.0917, +0.1445]$	$+0.1746$ $[+0.1437, +0.2054]$	$+0.1498$ $[+0.1292, +0.1703]$

Continued on next page

			CSI 300		CSI 500	
Features	L	Entry	Ridge	LightGBM	Ridge	LightGBM
Range	10	CL	$+0.0928$ $[+0.0667, +0.1188]$	$+0.0494$ $[+0.0318, +0.0670]$	$+0.1463$ $[+0.1212, +0.1715]$	$+0.0839$ $[+0.0620, +0.1059]$
		LC	-0.0068 $[-0.0106, -0.0030]$	-0.0082 $[-0.0191, +0.0027]$	-0.0004 $[-0.0016, +0.0008]$	$+0.0030$ $[-0.0080, +0.0140]$
		LL	$+0.0955$ $[+0.0672, +0.1237]$	$+0.0768$ $[+0.0550, +0.0985]$	$+0.1444$ $[+0.1200, +0.1688]$	$+0.1165$ $[+0.0975, +0.1354]$

Table 24: **Equal one-feature exposure produces very different return effects.** The complete ΔNet (pp) comparison shows why a feature count alone cannot rank economic severity.

	CSI 300		CSI 500	
Feature	Ridge	LightGBM	Ridge	LightGBM
ret_1	$+31.77$ $[+27.25, +36.30]$	$+12.78$ $[+9.83, +15.73]$	$+11.54$ $[+9.48, +13.59]$	-7.11 $[-12.74, -1.48]$
ret_5	-29.56 $[-34.50, -24.62]$	-6.99 $[-9.84, -4.13]$	-28.69 $[-33.02, -24.37]$	-2.01 $[-4.22, +0.20]$
ret_10	$+30.75$ $[+27.32, +34.19]$	$+7.97$ $[+5.21, +10.72]$	$+18.40$ $[+15.74, +21.07]$	$+14.20$ $[+9.69, +18.71]$
ret_20	$+0.22$ $[-2.83, +3.26]$	$+3.58$ $[+1.02, +6.14]$	-11.71 $[-16.06, -7.36]$	$+2.04$ $[-0.11, +4.19]$
ret_60	-6.97 $[-9.20, -4.75]$	-1.62 $[-3.14, -0.10]$	-8.70 $[-12.23, -5.18]$	-1.46 $[-2.88, -0.04]$
vol_5	-0.17 $[-1.49, +1.15]$	-0.76 $[-3.19, +1.68]$	$+2.38$ $[+1.36, +3.40]$	-0.22 $[-1.28, +0.84]$
vol_20	$+0.13$ $[-0.68, +0.93]$	$+0.07$ $[-1.51, +1.64]$	$+3.83$ $[+2.21, +5.45]$	$+1.44$ $[-0.15, +3.02]$
vol_60	-1.44 $[-2.26, -0.61]$	-0.15 $[-1.09, +0.80]$	-2.27 $[-3.28, -1.26]$	-0.42 $[-1.14, +0.29]$
hl_range_1	$+6.25$ $[+4.25, +8.25]$	$+2.95$ $[+1.07, +4.83]$	$+1.17$ $[+0.16, +2.18]$	$+3.59$ $[+1.54, +5.64]$
hl_range_5_mean	$+7.85$ $[+4.49, +11.21]$	$+1.06$ $[-0.87, +3.00]$	$+11.48$ $[+8.08, +14.87]$	$+2.35$ $[+0.31, +4.38]$
oc_gap_1	-55.23 $[-64.37, -46.10]$	-21.91 $[-25.75, -18.07]$	-47.16 $[-53.37, -40.95]$	-24.31 $[-28.51, -20.11]$
close_ma5_ratio	$+17.56$ $[+15.30, +19.82]$	$+1.89$ $[-0.80, +4.57]$	$+13.38$ $[+11.36, +15.40]$	$+9.28$ $[+6.27, +12.28]$
close_ma20_ratio	-39.02 $[-43.79, -34.26]$	-10.49 $[-12.93, -8.05]$	$+3.46$ $[-6.10, +13.02]$	-0.67 $[-2.83, +1.48]$
close_ma60_ratio	$+3.35$ $[-0.46, +7.16]$	-1.76 $[-3.91, +0.39]$	$+0.11$ $[-6.52, +6.74]$	$+1.84$ $[-0.58, +4.26]$
log_volume	-0.91 $[-1.95, +0.13]$	-0.87 $[-2.04, +0.29]$	-0.79 $[-1.62, +0.03]$	$+0.81$ $[-0.29, +1.92]$
vol_ratio_5	-3.06 $[-5.28, -0.83]$	$+0.16$ $[-1.60, +1.92]$	$+2.42$ $[+1.43, +3.40]$	$+0.62$ $[-0.80, +2.04]$
vol_ratio_20	$+17.33$ $[+12.74, +21.92]$	$+1.55$ $[-0.65, +3.74]$	$+1.11$ $[+0.43, +1.79]$	$+1.00$ $[-1.01, +3.00]$
log_dollar_volume	$+2.50$ $[+0.63, +4.37]$	$+1.51$ $[-0.17, +3.20]$	-3.59 $[-4.79, -2.39]$	-3.04 $[-4.49, -1.59]$
up_days_ratio_20	$+2.79$ $[+1.97, +3.61]$	-1.82 $[-3.05, -0.60]$	$+4.47$ $[+3.56, +5.37]$	-0.77 $[-1.73, +0.19]$

Table 25: **Predictive responses also depend on which feature is exposed.** All 19 ΔRankIC comparisons are retained, including near-zero and model-dependent effects.

	CSI 300		CSI 500	
Feature	Ridge	LightGBM	Ridge	LightGBM
ret_1	+0.0833 [+0.0743, +0.0922]	+0.0636 [+0.0549, +0.0723]	+0.0293 [+0.0254, +0.0333]	+0.0275 [+0.0147, +0.0403]
ret_5	−0.0753 [−0.0819, −0.0686]	−0.0244 [−0.0291, −0.0197]	−0.0720 [−0.0788, −0.0652]	−0.0143 [−0.0172, −0.0113]
ret_10	+0.0770 [+0.0730, +0.0809]	+0.0282 [+0.0247, +0.0317]	+0.0438 [+0.0389, +0.0486]	+0.0316 [+0.0250, +0.0381]
ret_20	+0.0040 [−0.0033, +0.0113]	+0.0094 [+0.0053, +0.0135]	−0.0255 [−0.0337, −0.0172]	+0.0076 [+0.0033, +0.0118]
ret_60	−0.0179 [−0.0221, −0.0136]	−0.0020 [−0.0041, 0.0000]	−0.0188 [−0.0253, −0.0122]	−0.0029 [−0.0057, −0.0001]
vol_5	+0.0002 [−0.0010, +0.0015]	−0.0039 [−0.0056, −0.0023]	+0.0024 [+0.0014, +0.0034]	−0.0012 [−0.0023, −0.0002]
vol_20	−0.0010 [−0.0020, 0.0000]	−0.0008 [−0.0026, +0.0009]	+0.0031 [+0.0018, +0.0044]	−0.0003 [−0.0016, +0.0010]
vol_60	−0.0019 [−0.0026, −0.0013]	−0.0010 [−0.0018, −0.0001]	−0.0032 [−0.0042, −0.0022]	−0.0007 [−0.0012, −0.0001]
hl_range_1	+0.0104 [+0.0074, +0.0133]	+0.0060 [+0.0028, +0.0092]	+0.0020 [+0.0006, +0.0033]	+0.0044 [+0.0022, +0.0066]
hl_range_5_mean	+0.0116 [+0.0086, +0.0147]	+0.0033 [+0.0018, +0.0048]	+0.0151 [+0.0110, +0.0193]	+0.0063 [+0.0041, +0.0086]
oc_gap_1	−0.1443 [−0.1566, −0.1321]	−0.0834 [−0.0914, −0.0753]	−0.1115 [−0.1208, −0.1023]	−0.0548 [−0.0609, −0.0487]
close_ma5_ratio	+0.0460 [+0.0429, +0.0492]	+0.0150 [+0.0106, +0.0194]	+0.0345 [+0.0307, +0.0383]	+0.0202 [+0.0156, +0.0248]
close_ma20_ratio	−0.1146 [−0.1260, −0.1031]	−0.0336 [−0.0371, −0.0301]	−0.0022 [−0.0221, +0.0178]	−0.0091 [−0.0120, −0.0062]
close_ma60_ratio	+0.0055 [−0.0036, +0.0145]	−0.0048 [−0.0081, −0.0015]	−0.0074 [−0.0212, +0.0065]	−0.0043 [−0.0086, −0.0001]
log_volume	−0.0059 [−0.0086, −0.0033]	−0.0036 [−0.0050, −0.0023]	−0.0021 [−0.0035, −0.0007]	+0.0002 [−0.0009, +0.0013]
vol_ratio_5	−0.0151 [−0.0200, −0.0101]	−0.0023 [−0.0042, −0.0003]	+0.0078 [+0.0063, +0.0093]	−0.0005 [−0.0020, +0.0009]
vol_ratio_20	+0.0302 [+0.0236, +0.0368]	+0.0096 [+0.0064, +0.0127]	+0.0036 [+0.0026, +0.0047]	+0.0058 [+0.0035, +0.0082]
log_dollar_volume	+0.0064 [+0.0029, +0.0099]	+0.0022 [+0.0005, +0.0039]	−0.0093 [−0.0113, −0.0074]	−0.0069 [−0.0082, −0.0057]
up_days_ratio_20	+0.0080 [+0.0069, +0.0092]	−0.0029 [−0.0037, −0.0021]	+0.0163 [+0.0140, +0.0186]	−0.0031 [−0.0045, −0.0017]

Equal-mask means and lag sensitivity. Variation across fixed masks measures sensitivity to exposure location; it is distinct from temporal uncertainty. For the latter, we form $\bar{d}_t = M^{-1} \sum_{m=1}^{M} d_{m,t}$, with $M = 20$ for date or stock–date masks and $M = 19$ for feature rotations, then apply Equation (15) to this daily sequence. Tables 26 and 27 report post-result descriptive intervals at every dose. Averaging within date before computing uncertainty retains dependence among mask responses; averaging individual-mask bounds would not. Zero-dose and full-exposure intervals match CC and CL.

Table 28 varies $b \in \{5, 10, 20, 40\}$ for four net-return comparisons selected after observing results: the All-feature $L = 1$ interaction, State CL at $L = 1$, ten-session-return minus intraday single-feature exposure, and equal-mask stock–date exposure at $p = 5\%$. Each contrast uses aligned daily differences before its standard error is calculated. These summaries assess lag sensitivity under the same pointwise, unadjusted interpretation.

The row labels identify these four contrasts as interaction (I), market state (S), feature difference (F), and sparse stock–date exposure (R). They label comparisons, not additional metrics.

The lag choice does not change the direction of these net-return comparisons. Across both models and universes, the interaction, State response and ten-session-return minus intraday contrast have intervals wholly above zero at every lag; stock–date exposure at $p = 5\%$ has intervals wholly below zero. This conclusion concerns the four portfolio comparisons, not all predictive endpoints in the grid.

Table 26: **Sparse contamination can produce a detectable economic response.** Full-exposure endpoints connect back to the dense experiment. ΔNet (pp); descriptive post-result intervals.

		CSI 300		CSI 500	
Exposure	Dose	Ridge	LightGBM	Ridge	LightGBM
Dates	0%	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]
	1%	−0.41 [−0.51, −0.30]	−0.09 [−0.21, +0.03]	−0.35 [−0.45, −0.25]	−0.12 [−0.22, −0.03]
	5%	−2.37 [−2.74, −2.00]	−0.64 [−1.01, −0.26]	−2.28 [−2.65, −1.91]	−1.21 [−1.59, −0.83]
	10%	−4.59 [−5.25, −3.93]	−1.25 [−1.88, −0.63]	−4.35 [−5.00, −3.70]	−2.34 [−3.02, −1.67]
	25%	−11.66 [−13.20, −10.13]	−3.48 [−4.89, −2.08]	−11.05 [−12.62, −9.48]	−5.95 [−7.66, −4.25]
	100%	−47.93 [−54.06, −41.80]	−14.64 [−20.03, −9.25]	−45.51 [−51.97, −39.05]	−23.78 [−30.55, −17.01]
Stock-dates	0%	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]
	1%	−0.43 [−0.55, −0.30]	−0.08 [−0.22, +0.07]	−0.47 [−0.60, −0.34]	−0.21 [−0.35, −0.06]
	5%	−2.11 [−2.63, −1.59]	−0.56 [−0.97, −0.15]	−2.13 [−2.59, −1.66]	−1.11 [−1.57, −0.65]
	10%	−4.45 [−5.31, −3.59]	−1.36 [−2.08, −0.65]	−4.37 [−5.20, −3.54]	−2.29 [−3.10, −1.48]
	25%	−11.52 [−13.26, −9.78]	−3.69 [−5.19, −2.18]	−11.21 [−13.09, −9.32]	−5.86 [−7.68, −4.04]
	100%	−47.93 [−54.06, −41.80]	−14.64 [−20.03, −9.25]	−45.51 [−51.97, −39.05]	−23.78 [−30.55, −17.01]
Features	0	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]
	1	−0.83 [−1.87, +0.20]	−0.68 [−1.59, +0.24]	−1.54 [−2.24, −0.83]	−0.15 [−1.03, +0.73]
	2	−3.78 [−5.34, −2.22]	−0.09 [−1.50, +1.32]	−3.62 [−4.73, −2.51]	+0.07 [−1.18, +1.31]
	4	−8.74 [−11.31, −6.17]	+0.65 [−1.23, +2.53]	−7.53 [−9.22, −5.83]	0.00 [−2.09, +2.09]
	8	−15.92 [−19.54, −12.29]	−0.44 [−3.27, +2.39]	−15.24 [−17.99, −12.48]	−2.67 [−6.07, +0.74]
	19	−47.93 [−54.06, −41.80]	−14.64 [−20.03, −9.25]	−45.51 [−51.97, −39.05]	−23.78 [−30.55, −17.01]

Table 27: **Sparse exposure can have small or unresolved predictive effects.** These ΔRankIC responses accompany the portfolio effects in Table 26; intervals are descriptive post-result summaries.

		CSI 300		CSI 500	
Exposure	Dose	Ridge	LightGBM	Ridge	LightGBM
Dates	0%	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]
	1%	−0.0008 [−0.0010, −0.0006]	−0.0001 [−0.0002, +0.0001]	−0.0007 [−0.0009, −0.0006]	0.0000 [−0.0002, +0.0002]
	5%	−0.0045 [−0.0050, −0.0039]	−0.0011 [−0.0018, −0.0005]	−0.0043 [−0.0049, −0.0037]	−0.0005 [−0.0012, +0.0003]
	10%	−0.0088 [−0.0097, −0.0079]	−0.0025 [−0.0036, −0.0013]	−0.0084 [−0.0094, −0.0074]	−0.0007 [−0.0021, +0.0007]
	25%	−0.0227 [−0.0246, −0.0208]	−0.0073 [−0.0102, −0.0044]	−0.0218 [−0.0242, −0.0194]	−0.0022 [−0.0058, +0.0014]
	100%	−0.0918 [−0.0986, −0.0849]	−0.0307 [−0.0428, −0.0187]	−0.0876 [−0.0974, −0.0778]	−0.0111 [−0.0258, +0.0037]
Stock-dates	0%	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]
	1%	−0.0009 [−0.0010, −0.0009]	−0.0003 [−0.0004, −0.0001]	−0.0008 [−0.0009, −0.0007]	−0.0002 [−0.0003, 0.0000]
	5%	−0.0045 [−0.0049, −0.0042]	−0.0015 [−0.0021, −0.0009]	−0.0042 [−0.0047, −0.0037]	−0.0006 [−0.0013, +0.0001]
	10%	−0.0090 [−0.0097, −0.0082]	−0.0029 [−0.0041, −0.0018]	−0.0085 [−0.0094, −0.0075]	−0.0012 [−0.0027, +0.0002]
	25%	−0.0225 [−0.0243, −0.0208]	−0.0076 [−0.0106, −0.0047]	−0.0213 [−0.0238, −0.0189]	−0.0028 [−0.0064, +0.0008]
	100%	−0.0918 [−0.0986, −0.0849]	−0.0307 [−0.0428, −0.0187]	−0.0876 [−0.0974, −0.0778]	−0.0111 [−0.0258, +0.0037]
Features	0	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]	0.0000 [0.0000, 0.0000]
	1	−0.0049 [−0.0057, −0.0041]	−0.0013 [−0.0023, −0.0004]	−0.0050 [−0.0055, −0.0044]	+0.0003 [−0.0007, +0.0013]
	2	−0.0106 [−0.0118, −0.0094]	−0.0011 [−0.0029, +0.0007]	−0.0100 [−0.0112, −0.0089]	+0.0014 [−0.0006, +0.0034]
	4	−0.0200 [−0.0222, −0.0178]	−0.0016 [−0.0049, +0.0016]	−0.0197 [−0.0217, −0.0176]	+0.0019 [−0.0018, +0.0056]
	8	−0.0346 [−0.0385, −0.0307]	−0.0041 [−0.0098, +0.0016]	−0.0377 [−0.0418, −0.0335]	+0.0032 [−0.0039, +0.0102]
	19	−0.0918 [−0.0986, −0.0849]	−0.0307 [−0.0428, −0.0187]	−0.0876 [−0.0974, −0.0778]	−0.0111 [−0.0258, +0.0037]

Table 28: **The four focused economic contrasts retain their directions across lag choices.** These post-result comparisons have positive intervals for interaction, state, and feature contrasts, and negative intervals for sparse stock–date exposure. ΔNet (pp).

Contrast	Universe/ model	Lag 5	Lag 10	Lag 20	Lag 40
I	300/Ridge	+214.57 [+201.62, +227.52]	+214.57 [+198.39, +230.75]	+214.57 [+194.20, +234.94]	+214.57 [+190.08, +239.07]
	300/LGBM	+175.88 [+163.94, +187.81]	+175.88 [+161.05, +190.70]	+175.88 [+157.30, +194.45]	+175.88 [+154.01, +197.74]
	500/Ridge	+220.00 [+206.70, +233.29]	+220.00 [+203.32, +236.68]	+220.00 [+199.03, +240.96]	+220.00 [+195.08, +244.92]
	500/LGBM	+194.11 [+181.05, +207.17]	+194.11 [+177.90, +210.31]	+194.11 [+173.89, +214.33]	+194.11 [+170.39, +217.83]

Continued on next page

Contrast	Universe/ model	Lag 5	Lag 10	Lag 20	Lag 40
S	300/Ridge	$+25.52$ $[+20.33, +30.72]$	$+25.52$ $[+19.97, +31.08]$	$+25.52$ $[+19.31, +31.73]$	$+25.52$ $[+18.98, +32.07]$
	300/LGBM	$+8.82$ $[+5.58, +12.06]$	$+8.82$ $[+5.52, +12.12]$	$+8.82$ $[+5.53, +12.12]$	$+8.82$ $[+5.49, +12.16]$
	500/Ridge	$+12.79$ $[+9.13, +16.44]$	$+12.79$ $[+9.21, +16.36]$	$+12.79$ $[+9.25, +16.32]$	$+12.79$ $[+9.41, +16.16]$
	500/LGBM	$+6.84$ $[+4.19, +9.48]$	$+6.84$ $[+4.20, +9.48]$	$+6.84$ $[+4.18, +9.50]$	$+6.84$ $[+4.17, +9.51]$
F	300/Ridge	$+85.99$ $[+78.93, +93.04]$	$+85.99$ $[+77.38, +94.59]$	$+85.99$ $[+75.25, +96.72]$	$+85.99$ $[+72.30, +99.67]$
	300/LGBM	$+29.88$ $[+25.57, +34.19]$	$+29.88$ $[+25.16, +34.60]$	$+29.88$ $[+24.60, +35.16]$	$+29.88$ $[+23.99, +35.77]$
	500/Ridge	$+65.57$ $[+60.92, +70.21]$	$+65.57$ $[+60.13, +71.00]$	$+65.57$ $[+59.09, +72.05]$	$+65.57$ $[+58.20, +72.93]$
	500/LGBM	$+38.51$ $[+32.60, +44.41]$	$+38.51$ $[+32.02, +44.99]$	$+38.51$ $[+31.47, +45.55]$	$+38.51$ $[+30.90, +46.12]$
R	300/Ridge	-2.11 $[-2.59, -1.63]$	-2.11 $[-2.59, -1.63]$	-2.11 $[-2.63, -1.59]$	-2.11 $[-2.68, -1.54]$
	300/LGBM	-0.56 $[-0.99, -0.14]$	-0.56 $[-0.97, -0.15]$	-0.56 $[-0.97, -0.15]$	-0.56 $[-0.95, -0.18]$
	500/Ridge	-2.13 $[-2.54, -1.72]$	-2.13 $[-2.55, -1.71]$	-2.13 $[-2.59, -1.66]$	-2.13 $[-2.66, -1.60]$
	500/LGBM	-1.11 $[-1.56, -0.67]$	-1.11 $[-1.57, -0.66]$	-1.11 $[-1.57, -0.65]$	-1.11 $[-1.60, -0.62]$

C.3 Complete training–scoring interactions

The four learning-entry conditions ask whether training changes the value of future scoring inputs. Their interaction ($I = LL - LC - CL + CC$) compares the scoring effect after exposed versus clean training. We report the interaction on the Net (pp) and RankIC scales, preserving covariance across the four matched daily paths. The one-session tables contrast All and market-state features; the latter describe prices and trading activity, not portfolio holdings. The clean $CC - CC$ reference is zero by construction, while Table 30 reports the actual Net levels. All clean paths are shared across lead labels.

The return comparison shows the consequences of learning from the exposure. Table 31 asks whether the same reversal is already present in the predictive ranks.

Training exposure reverses the All-feature RankIC response at $L = 1$: CL lowers RankIC by 0.011–0.092, while LL raises it by 0.362–0.397. Both compare with the group's clean prediction map, matching the return comparison in the main text.

Interaction and paired inference. The factorial estimand is

$$I_t = r_t^{LL} - r_t^{LC} - r_t^{CL} + r_t^{CC}, \qquad \mathcal{I} = 100 \times 252\,\overline{I}. \tag{16}$$

This compares scoring exposure under exposed training with scoring exposure under clean training. We form the daily contrast on the same 726 economic dates before estimating HAC uncertainty, retaining all cross-cell covariances. Adding separate standard errors or interval limits would not estimate this uncertainty. RankIC uses the same contrast without annualization.

The full grid shows where training and scoring reinforce each other. Tables 32 and 33 give all 96 interactions: six groups, four leads, two models and two universes. Net-return intervals are wholly positive in 92 cases, wholly negative in two, and include zero in two; the corresponding RankIC counts are 93, one and two. Positive interaction at $L = 1$ for All and State is therefore not an artifact of the contrast definition. The negative net interactions are CSI 500/Ridge with Range at $L = 10$ and Volatility at $L = 5$. For all eight one-session All/State interactions, net-return and RankIC intervals are positive in pooled and individual-year estimates at each bandwidth, $b \in \{5, 10, 20, 40\}$. These are pointwise retrospective comparisons; annual and model cells are not assumed independent.

Table 29: **Training exposure changes what future scoring inputs do to returns.** For All features, CL depresses returns while LL raises them; the market-state interaction is also positive. ΔNet (pp), one-session lead.

Contrast	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
All, $L = 1$				
$CC - CC$ (reference)	0 (ref.)	0 (ref.)	0 (ref.)	0 (ref.)
$CL - CC$	-47.9 [-54.1, -41.8]	-14.6 [-20.0, -9.2]	-45.5 [-52.0, -39.1]	-23.8 [-30.6, -17.0]
$LC - CC$	-6.4 [-21.0, 8.2]	-7.7 [-18.1, 2.7]	-13.5 [-28.3, 1.2]	-12.3 [-21.3, -3.3]
$LL - CC$	160.2 [143.6, 176.9]	153.5 [137.7, 169.4]	161.0 [143.1, 178.8]	158.0 [142.4, 173.6]
I	214.6 [194.2, 234.9]	175.9 [157.3, 194.4]	220.0 [199.0, 241.0]	194.1 [173.9, 214.3]
Market state, $L = 1$				
$CC - CC$ (reference)	0 (ref.)	0 (ref.)	0 (ref.)	0 (ref.)
$CL - CC$	25.5 [19.3, 31.7]	8.8 [5.5, 12.1]	12.8 [9.3, 16.3]	6.8 [4.2, 9.5]
$LC - CC$	4.4 [-1.9, 10.7]	-1.6 [-7.0, 3.9]	-3.5 [-13.0, 6.0]	-2.6 [-8.4, 3.1]
$LL - CC$	47.7 [36.5, 59.0]	39.4 [30.1, 48.7]	47.1 [35.7, 58.5]	51.5 [42.0, 61.0]
I	17.8 [13.1, 22.5]	32.1 [26.0, 38.2]	37.9 [31.5, 44.2]	47.3 [40.2, 54.4]

Table 30: **Joint exposure can turn negative reference returns into apparent profitability.** Clean Net (%) levels provide the reference for this comparison.

Features	CSI300 Ridge ↑	CSI300 LightGBM ↑	CSI500 Ridge ↑	CSI500 LightGBM ↑
All	-8.57	-5.01	**-0.49**	**-1.55**
Returns	-10.16	**-2.34**	-8.63	-3.55
State	-11.42	-4.48	-6.31	-7.78
Volatility	**-4.82**	-2.99	-9.75	-8.07
Volume	-5.11	-5.35	-5.26	-7.07
Range	-10.05	-12.52	-10.82	-8.40

Table 31: **The predictive response also reverses between CL and LL for All features.** Market-state inputs retain a positive interaction without the signed-return reconstruction basis. ΔRankIC, one-session lead.

Contrast	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
All, $L = 1$				
$CC - CC$ (reference)	0 (ref.)	0 (ref.)	0 (ref.)	0 (ref.)
$CL - CC$	-0.092 [-0.099, -0.085]	-0.031 [-0.043, -0.019]	-0.088 [-0.097, -0.078]	-0.011 [-0.026, 0.004]
$LC - CC$	0.012 [-0.022, 0.047]	-0.004 [-0.023, 0.016]	-0.007 [-0.038, 0.023]	-0.004 [-0.023, 0.014]
$LL - CC$	0.397 [0.368, 0.425]	0.362 [0.347, 0.377]	0.387 [0.363, 0.410]	0.374 [0.359, 0.388]
I	0.476 [0.461, 0.492]	0.397 [0.378, 0.415]	0.482 [0.464, 0.499]	0.389 [0.368, 0.410]
State, $L = 1$				
$CC - CC$ (reference)	0 (ref.)	0 (ref.)	0 (ref.)	0 (ref.)
$CL - CC$	0.028 [0.022, 0.035]	0.013 [0.008, 0.018]	0.012 [0.008, 0.017]	0.015 [0.010, 0.020]
$LC - CC$	0.009 [-0.002, 0.021]	0.004 [-0.002, 0.010]	-0.002 [-0.016, 0.013]	0.004 [-0.003, 0.012]
$LL - CC$	0.063 [0.046, 0.079]	0.055 [0.045, 0.066]	0.067 [0.049, 0.084]	0.073 [0.063, 0.084]
I	0.025 [0.017, 0.033]	0.038 [0.032, 0.045]	0.056 [0.045, 0.067]	0.054 [0.046, 0.062]

Table 32: **Positive economic interaction occurs across most of the grid.** Of 96 intervals, 92 are positive, two negative, and two include zero. Net interaction I (pp).

Features	L	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
All	1	214.6 [194.2, 234.9]	175.9 [157.3, 194.4]	220.0 [199.0, 241.0]	194.1 [173.9, 214.3]
	2	287.8 [259.4, 316.1]	247.3 [220.1, 274.6]	285.6 [258.2, 313.0]	251.6 [224.2, 278.9]
	5	545.4 [488.2, 602.7]	460.7 [410.1, 511.3]	503.7 [451.3, 556.2]	437.8 [390.0, 485.6]
	10	347.2 [316.8, 377.5]	391.2 [354.3, 428.0]	314.7 [284.6, 344.8]	357.5 [322.6, 392.4]
Returns	1	209.2 [190.0, 228.5]	169.0 [151.8, 186.2]	222.3 [203.5, 241.1]	174.3 [157.1, 191.6]
	2	302.8 [273.0, 332.5]	234.9 [210.2, 259.7]	310.1 [281.7, 338.5]	232.9 [208.1, 257.7]
	5	643.0 [578.9, 707.1]	492.6 [445.2, 540.1]	630.9 [573.0, 688.7]	461.4 [412.9, 509.9]
	10	256.9 [231.6, 282.1]	390.2 [352.6, 427.8]	284.1 [258.6, 309.6]	361.5 [324.9, 398.1]
State	1	17.8 [13.1, 22.5]	32.1 [26.0, 38.2]	37.9 [31.5, 44.2]	47.3 [40.2, 54.4]
	2	20.1 [13.4, 26.8]	59.4 [49.0, 69.8]	65.2 [53.3, 77.1]	93.3 [79.5, 107.1]
	5	11.2 [-3.0, 25.4]	102.2 [82.7, 121.6]	67.0 [47.2, 86.9]	156.7 [132.7, 180.7]
	10	47.8 [35.8, 59.7]	81.6 [63.9, 99.3]	53.9 [40.1, 67.8]	117.1 [97.7, 136.6]

Features	L	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
Volatility	1	2.1 [0.3, 3.9]	4.0 [1.8, 6.2]	2.0 [0.8, 3.2]	3.0 [0.9, 5.1]
	2	6.5 [2.5, 10.6]	17.8 [13.7, 21.9]	4.7 [2.3, 7.2]	12.8 [8.4, 17.1]
	5	17.2 [8.8, 25.6]	54.6 [43.2, 66.1]	-17.4 [-28.0, -6.7]	34.7 [24.7, 44.6]
	10	18.3 [3.9, 32.6]	28.5 [14.9, 42.2]	6.0 [-0.9, 12.9]	35.2 [24.0, 46.5]
Volume	1	21.0 [15.4, 26.6]	25.3 [19.8, 30.9]	45.2 [36.1, 54.2]	37.6 [31.2, 44.0]
	2	32.9 [24.1, 41.7]	46.5 [37.6, 55.4]	75.8 [60.1, 91.5]	73.5 [62.8, 84.3]
	5	36.4 [22.8, 50.0]	70.4 [54.7, 86.2]	76.0 [49.8, 102.3]	111.2 [94.0, 128.4]
	10	40.4 [29.4, 51.4]	58.1 [44.0, 72.3]	65.2 [49.0, 81.4]	86.8 [72.4, 101.1]
Range	1	11.5 [7.6, 15.4]	30.6 [23.9, 37.4]	40.1 [33.0, 47.1]	33.6 [27.2, 40.0]
	2	7.6 [3.7, 11.5]	37.0 [28.1, 45.8]	42.8 [35.4, 50.1]	41.8 [32.6, 50.9]
	5	9.5 [4.9, 14.0]	70.6 [46.7, 94.4]	15.5 [8.2, 22.8]	85.9 [63.8, 108.1]
	10	14.2 [8.3, 20.0]	44.8 [26.6, 62.9]	-3.8 [-6.9, -0.8]	71.3 [51.4, 91.1]

Table 33: **The predictive interaction follows a similar pattern.** Of 96 intervals, 93 are positive, one negative, and two include zero. RankIC interaction I.

Features	L	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
All	1	0.4761 [0.4607, 0.4915]	0.3965 [0.3783, 0.4148]	0.4817 [0.4642, 0.4993]	0.3892 [0.3685, 0.4099]
	2	0.6266 [0.6053, 0.6480]	0.5372 [0.5097, 0.5646]	0.6013 [0.5794, 0.6231]	0.4918 [0.4638, 0.5199]
	5	1.2631 [1.2175, 1.3086]	1.0924 [1.0431, 1.1417]	1.1366 [1.0922, 1.1810]	0.9487 [0.8980, 0.9995]
	10	0.8573 [0.8140, 0.9005]	0.9324 [0.8979, 0.9669]	0.8047 [0.7668, 0.8427]	0.8287 [0.7957, 0.8618]
Returns	1	0.5032 [0.4882, 0.5181]	0.3548 [0.3393, 0.3702]	0.5314 [0.5165, 0.5463]	0.3451 [0.3254, 0.3649]
	2	0.7076 [0.6847, 0.7304]	0.4612 [0.4388, 0.4836]	0.7059 [0.6847, 0.7270]	0.4240 [0.3965, 0.4515]
	5	1.6371 [1.5887, 1.6855]	1.0973 [1.0500, 1.1446]	1.5537 [1.5070, 1.6004]	0.9865 [0.9315, 1.0415]
	10	0.7248 [0.6815, 0.7681]	0.8612 [0.8176, 0.9049]	0.7453 [0.7273, 0.7633]	0.7838 [0.7402, 0.8274]
State	1	0.0251 [0.0172, 0.0329]	0.0384 [0.0319, 0.0449]	0.0560 [0.0454, 0.0667]	0.0539 [0.0461, 0.0617]
	2	0.0463 [0.0343, 0.0582]	0.0876 [0.0741, 0.1010]	0.1106 [0.0919, 0.1292]	0.1178 [0.1007, 0.1350]
	5	0.0144 [-0.0094, 0.0383]	0.1417 [0.1222, 0.1613]	0.1082 [0.0778, 0.1387]	0.1743 [0.1467, 0.2019]
	10	0.0753 [0.0533, 0.0974]	0.1243 [0.1013, 0.1473]	0.1125 [0.0871, 0.1380]	0.1565 [0.1333, 0.1798]

Features	L	CSI300 Ridge	CSI300 LightGBM	CSI500 Ridge	CSI500 LightGBM
Volatility	1	0.0029 [0.0011, 0.0047]	0.0073 [0.0047, 0.0098]	0.0026 [0.0017, 0.0034]	0.0028 [0.0010, 0.0046]
	2	0.0080 [0.0033, 0.0127]	0.0253 [0.0198, 0.0307]	0.0075 [0.0053, 0.0097]	0.0123 [0.0086, 0.0161]
	5	0.0204 [0.0076, 0.0332]	0.0838 [0.0661, 0.1015]	-0.0140 [-0.0249, -0.0030]	0.0241 [0.0083, 0.0399]
	10	0.0427 [0.0156, 0.0698]	0.0448 [0.0217, 0.0679]	0.0125 [0.0015, 0.0236]	0.0191 [0.0004, 0.0377]
Volume	1	0.0350 [0.0251, 0.0449]	0.0360 [0.0280, 0.0440]	0.0792 [0.0639, 0.0944]	0.0541 [0.0453, 0.0629]
	2	0.0661 [0.0501, 0.0821]	0.0712 [0.0570, 0.0855]	0.1471 [0.1215, 0.1728]	0.1170 [0.0991, 0.1350]
	5	0.0702 [0.0494, 0.0910]	0.1371 [0.1151, 0.1591]	0.1504 [0.1113, 0.1894]	0.1739 [0.1489, 0.1988]
	10	0.0516 [0.0348, 0.0684]	0.1027 [0.0845, 0.1209]	0.0704 [0.0442, 0.0965]	0.1433 [0.1219, 0.1647]
Range	1	0.0103 [0.0046, 0.0161]	0.0257 [0.0186, 0.0329]	0.0295 [0.0202, 0.0388]	0.0248 [0.0192, 0.0305]
	2	0.0077 [0.0020, 0.0133]	0.0264 [0.0184, 0.0343]	0.0409 [0.0305, 0.0513]	0.0216 [0.0127, 0.0305]
	5	0.0061 [0.0023, 0.0099]	0.0622 [0.0425, 0.0818]	0.0182 [0.0118, 0.0246]	0.0357 [0.0107, 0.0608]
	10	0.0095 [0.0034, 0.0157]	0.0356 [0.0207, 0.0505]	-0.0015 [-0.0032, 0.0002]	0.0296 [0.0023, 0.0569]

C.4 How the information comparisons connect to protocol severity

Three controls connect the broad severity profile in Section 4.1 to the entry-point analysis. First, the nested-return identities can be summarized as

$$\mathrm{ret}_5(t+5) = \mathrm{ret}_{10}(t+10) - \mathrm{ret}_5(t+10) = \log(1+y_t) - g_{t+6} + g_{t+1}. \tag{17}$$

These identities explain why All-feature and Returns models approach target reconstruction at larger leads. Second, Market state and its Volume and Range subsets retain future information without that signed-return basis. Third, the frozen-model sparse curves change only prediction-time inputs, linking small departures to the exact CL endpoint. The controls distinguish near-target reconstruction from limited exposure and from exposure to other types of future information.

The fraction of selected rows exposed and the fraction of selected names replaced describe different effects. At requested 5% stock–date exposure, exposed records account for 5.22–5.48% of selected rows, compared with approximately 5% of the full scoring cross-section. The individual ten-session-return and intraday replacements both expose 5.26% of feature cells, yet replace 9.92–20.56% and 25.76–36.77% of clean selected names, respectively. Their positive and negative predictive and economic directions recur in every model–universe–year comparison. All-feature averages and balanced rotations retain the other feature identities, so these two channels do not stand in for all future information.

Appendix B reports the remaining protocol axes and unfitted Momentum control. Those comparisons establish which stages can change a backtest; the controls here explain why exposing the same amount of future information can have different effects.

D Matched exposure-location experiments

Figure 4 in the main text shows every condition; this table retains the numerical ranges across channels and doses.

Channels and fixed prediction maps. The location experiment fixes the clean annual fitted models and uses all completed 2022–2024 prediction panels. Three channels advance the same stock's

Table 34: **Smaller score changes produce larger economic distortions.** Near-cutoff exposure changes selections and returns more in all 24 future-input conditions.

Universe	Model	Near/far ratio		
		Score change	Rep.	Dev.
CSI 300	Ridge	0.93–0.94	2.44–11.92	1.60–6.69
	LightGBM	0.59–0.88	1.43–7.44	1.06–5.21
CSI 500	Ridge	0.91–0.94	4.03–32.99	1.98–18.05
	LightGBM	0.58–0.79	1.57–5.06	1.14–3.04
Conditions with near $>$ far		0/24	24/24	24/24

State vector, ten-session trailing return, or intraday open–close return by one session at prediction time, keeping the original row identity and target. State uses its eight-feature clean model; the two single-feature interventions use the All-feature clean model. Because preprocessing and prediction operate row-wise, choosing each row's clean or exposed endpoint score implements the feature intervention without refitting. Two further channels add symmetric score noise to the State and All clean score families.

Future availability is checked on the complete 19-feature endpoint vector before selecting the exposed feature group. A missing or non-finite endpoint feature causes the row to retain its full clean vector and score, keeping the sample unchanged. Advancing State removes the direct signed-return reconstruction basis but still supplies future information: next-session volatility and range describe part of the prediction interval.

Allocation at matched assigned counts. Order the clean scores from largest to smallest, with the fixed trading tie-break, and let $w_t = \lfloor K_t/2 \rfloor$. The near pools contain the last w_t selected and first w_t unselected stocks. The far pools contain the first w_t selected and last w_t unselected stocks. The stratified-random comparison samples from the complete selected and unselected groups. Near and far are therefore disjoint equal-width windows on both sides of the same cutoff.

For nominal dose $p \in \{0.01, 0.05\}$, each arm assigns $m_t = \lfloor pN_t \rfloor$ rows, split as evenly as possible between the two sides. An odd extra row is balanced exactly across the twenty fixed seeds on every date. Common hash-based sampling ranks couple the three arms and both doses. If a pool cannot supply its quota, all three arms retain clean scores on that date; no such shortage occurs in the completed experiment. Assignment depends on clean ranks and identities, not future availability. An unavailable assigned row retains its clean score without replenishment. Assigned, applied and numerically changed counts are recorded separately. Equal dose means equal assigned rows *within a channel*; eight exposed State features do not have the same feature-cell dose as one exposed return feature.

Development-scaled symmetric noise. Let IQR denote the clean-score interquartile range and write $\sigma_t = \mathrm{IQR}(s_t^0)/1.349$. For assignment $z_{ti}^{a,u}$, common future-availability indicator e_{ti} and symmetric hash-generated sign $\xi_{ti}^u \in \{-1, +1\}$, the noise score is

$$s_{ti}^{a,u} = s_{ti}^0 + z_{ti}^{a,u} e_{ti} \xi_{ti}^u g_m \sigma_t. \tag{18}$$

Signs and assignment use separate hash domains. A zero IQR produces zero noise. The calibration uses 1,420 clean CSI 300 development dates, January 4, 2016–November 4, 2021. Standardize each date's scores by its median and σ_t, take the median absolute same-stock change between adjacent dates within each year, then the median over dates. Pairs crossing an absent stock date, year boundary or zero-spread date are excluded. The resulting g_m values are approximately 0.45648 (Ridge) and 0.48104 (LightGBM), transferred unchanged to CSI 500 and the State score family. Noise does not use future feature values, but it shares the information channels' availability gate. This differs from the ungated, 2016–2018 calibrated stress used for audit prioritization.

Diagnostics, dates, and uncertainty. For each date and seed, score magnitude is mean absolute clean-to-exposed score change among assigned rows divided by σ_t. Stock replacement is $|S_t^0 \setminus S_t^{a,u}|/K_t$; economic deviation is $10^4 |r_t^{a,u} - r_t^0|$ in bp/day. The score diagnostic averages all 726 prediction dates; replacement averages 685 sleeve-opening dates; economic deviation averages all

726 portfolio dates. Absolute values are taken before seed averaging. Each location contrast first averages the twenty paired near-minus-far differences within date, then averages dates.

The design contains 5 channels $\times$ 2 models $\times$ 2 universes $\times$ 2 doses: forty pooled conditions, including twenty-four with future inputs. Three arms and twenty seeds give 2,400 portfolio paths. Pointwise Bartlett–Newey–West intervals use lag 20; the machine-readable supplement also reports lags 5, 10 and 40. These intervals describe temporal variation conditional on this design. Seeds, universes and forecasting models are not independent market replications, and the intervals do not adjust for multiple comparisons or selection of examples from earlier experiments.

The endpoint tables report near-minus-far estimates and 95% CIs: Rep. in pp, Dev. in bp/day, ΔNet in pp, and ΔRankIC in units of 10^{-3}. Intraday denotes the open–close feature; State/All noise controls use the corresponding clean feature sets. Dose always means the assigned fraction, before availability checks.

Complete endpoint comparisons. Replacement and absolute economic deviation follow $\mathrm{far} < \mathrm{random} < \mathrm{near}$ in all forty pooled conditions. Tables 35–38 report both doses and every channel–model–universe comparison, including signed net-return change and RankIC. Both near-minus-far intervals have positive lower bounds at all four bandwidths. These intervals test near minus far directly; the ordering through the random arm compares point estimates.

Figures 16 and 17 show all location contrasts at nominal 1% and 5% exposure, using shared endpoint-specific scales across doses. Their near-minus-far differences complement the near/far diagnostic ratios in Table 34 and Figure 4 and summarize the same completed comparisons.

Near-cutoff future inputs produce larger selection and economic effects despite smaller score changes and similar applied exposure. Across all twenty-four future-input conditions, the near/far ratio of mean standardized score magnitude is 0.576–0.945. Applied fractions average 0.761–0.808% at nominal 1% and 4.784–4.844% at nominal 5%, reflecting integer quotas and eligibility; the largest between-arm difference is below 0.00081 percentage points. The noise controls match normalized amplitude apart from small differences in realized gate reach.

The exposed feature determines whether near-cutoff exposure raises or lowers returns relative to far exposure. At 5%, the ten-session return's near-minus-far effect is $+1.03$–$+7.29$ pp/year and the intraday channel's is -6.69–-2.43; their HAC20 intervals retain these signs. At 1%, all four ten-session-return estimates are positive with three positive intervals, and all four intraday intervals are negative. All sixteen noise signed-return intervals span zero.

Annual replacement contrasts and their HAC20 lower bounds are positive in all 120 conditions. Annual absolute-deviation contrasts are positive in 119, with positive lower bounds in 118. Both intervals spanning zero are State/CSI 300/LightGBM at 5%: -0.153 $[-0.712, 0.407]$ bp/day in 2022 and 0.549 $[-0.035, 1.133]$ in 2023. Increasing dose raises the near-minus-far replacement contrast in all twenty channel–stratum pairs and the absolute-deviation contrast in nineteen; the latter contrast decreases from 0.649 to 0.391 bp/day in the same State stratum. The supplement retains all three arms, signed outcomes, RankIC, annual subdivisions and bandwidths, including these exceptions.

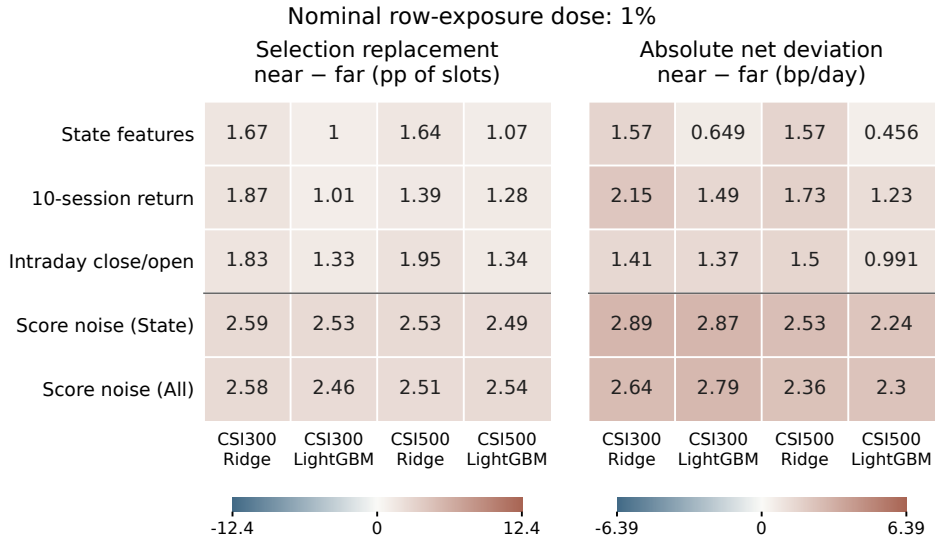


Figure 16: Near-minus-far exposure-location sensitivity at a nominal 1% row-exposure dose. Each arm balances selected/unselected quotas 50/50 across twenty fixed seeds; endpoints are averaged over seeds within date. Left: clean selected slots replaced (percentage points; 685 decision dates). Right: absolute net-return deviation from the matched clean portfolio, computed before seed averaging (basis points/day; 726 dates). Future features use a one-session lead. Development-scaled symmetric score noise shares the future-eligibility gate. Tables report pointwise 95% HAC20 intervals for every cell; panel-specific scales are shared across doses.

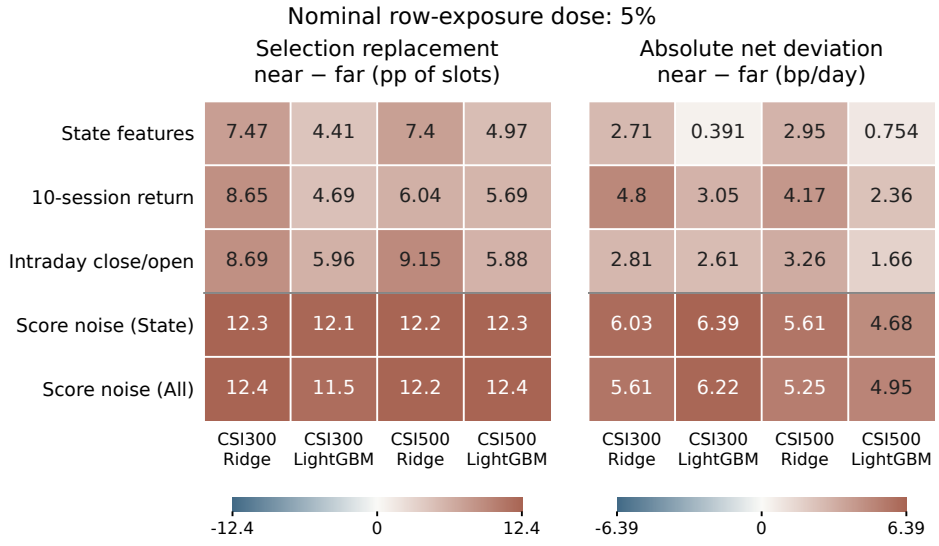


Figure 17: Equal exposure, different decision and economic consequences. Near-minus-far contrasts at a nominal 5% row dose: the future-information arms have smaller mean absolute standardized score changes among assigned rows near the boundary, yet change selections and portfolio paths more. Left: clean selected slots replaced (percentage points; 685 decision dates). Right: absolute net-return deviation from the matched clean portfolio (basis points/day; 726 dates). Assigned counts and selected/unselected quotas match across arms. Absolute deviations precede within-date averaging over twenty seeds. Future features use a one-session lead; development-scaled symmetric noise shares their availability gate. Complete pointwise HAC20 intervals appear in Appendix D; panel-specific scales are shared across doses.

Table 35: **Near-cutoff exposure replaces more selected stocks in every pooled condition.** The pattern holds for future inputs and matched-noise controls. Near-minus-far Rep. (pp).

Input	Dose	CSI300		CSI500	
		Ridge	LightGBM	Ridge	LightGBM
State	1%	1.67 [1.54, 1.81]	1 [0.9, 1.11]	1.64 [1.58, 1.7]	1.07 [0.998, 1.15]
	5%	7.47 [7.06, 7.88]	4.41 [4.03, 4.79]	7.4 [7.08, 7.72]	4.97 [4.64, 5.31]
ret_{10}	1%	1.87 [1.76, 1.99]	1.01 [0.939, 1.09]	1.39 [1.27, 1.51]	1.28 [1.24, 1.32]
	5%	8.65 [8.29, 9]	4.69 [4.44, 4.94]	6.04 [5.47, 6.62]	5.69 [5.5, 5.88]
Intraday	1%	1.83 [1.67, 2]	1.33 [1.23, 1.42]	1.95 [1.9, 2.01]	1.34 [1.29, 1.39]
	5%	8.69 [8.18, 9.21]	5.96 [5.72, 6.2]	9.15 [8.88, 9.43]	5.88 [5.63, 6.13]
Noise (State)	1%	2.59 [2.42, 2.75]	2.53 [2.38, 2.69]	2.53 [2.49, 2.56]	2.49 [2.45, 2.52]
	5%	12.3 [12.1, 12.5]	12.1 [11.8, 12.4]	12.2 [12, 12.4]	12.3 [12.1, 12.5]
Noise (All)	1%	2.58 [2.41, 2.75]	2.46 [2.31, 2.62]	2.51 [2.48, 2.55]	2.54 [2.51, 2.58]
	5%	12.4 [12.2, 12.6]	11.5 [11.2, 11.8]	12.2 [12.1, 12.4]	12.4 [12.2, 12.6]

Table 36: **Near-cutoff exposure also creates greater economic deviation.** Every pooled comparison is positive, connecting selection changes to portfolio consequences. Near-minus-far Dev. (bp/day).

Input	Dose	CSI300		CSI500	
		Ridge	LightGBM	Ridge	LightGBM
State	1%	1.57 [1.39, 1.75]	0.649 [0.539, 0.759]	1.57 [1.37, 1.77]	0.456 [0.361, 0.551]
	5%	2.71 [2.29, 3.14]	0.391 [0.0399, 0.742]	2.95 [2.45, 3.45]	0.754 [0.461, 1.05]
ret_{10}	1%	2.15 [1.95, 2.34]	1.49 [1.32, 1.66]	1.73 [1.56, 1.9]	1.23 [1.11, 1.36]
	5%	4.8 [4.35, 5.24]	3.05 [2.68, 3.41]	4.17 [3.72, 4.62]	2.36 [2.09, 2.63]
Intraday	1%	1.41 [1.22, 1.61]	1.37 [1.22, 1.52]	1.5 [1.32, 1.68]	0.991 [0.865, 1.12]
	5%	2.81 [2.27, 3.34]	2.61 [2.2, 3.02]	3.26 [2.72, 3.8]	1.66 [1.37, 1.95]
Noise (State)	1%	2.89 [2.61, 3.16]	2.87 [2.61, 3.13]	2.53 [2.3, 2.77]	2.24 [2.05, 2.43]
	5%	6.03 [5.52, 6.54]	6.39 [5.81, 6.98]	5.61 [5.12, 6.1]	4.68 [4.28, 5.08]
Noise (All)	1%	2.64 [2.41, 2.87]	2.79 [2.54, 3.03]	2.36 [2.16, 2.56]	2.3 [2.11, 2.49]
	5%	5.61 [5.14, 6.08]	6.22 [5.65, 6.79]	5.25 [4.77, 5.73]	4.95 [4.52, 5.37]

Table 37: **Greater distortion does not imply greater return inflation.** Return and intraday channels have opposite signed effects, while all noise intervals include zero. Near-minus-far ΔNet (pp).

Input	Dose	CSI300		CSI500	
		Ridge	LightGBM	Ridge	LightGBM
State	1%	0.318 [−0.0951, 0.731]	0.231 [−0.0705, 0.533]	0.202 [−0.00906, 0.414]	−0.0177 [−0.263, 0.227]
	5%	2.71 [1.11, 4.31]	0.406 [−1.12, 1.93]	1.49 [0.613, 2.38]	1.06 [−0.231, 2.36]
ret_{10}	1%	0.908 [0.625, 1.19]	0.183 [−0.0662, 0.432]	1.02 [0.771, 1.28]	0.411 [0.204, 0.618]
	5%	7.29 [6.2, 8.37]	1.03 [0.232, 1.82]	6.1 [5.09, 7.12]	2 [1.05, 2.95]
Intraday	1%	−0.644 [−0.974, −0.315]	−0.43 [−0.683, −0.177]	−0.825 [−1.12, −0.527]	−0.273 [−0.522, −0.0242]
	5%	−4.33 [−6.09, −2.57]	−3.33 [−4.46, −2.19]	−6.69 [−8.11, −5.27]	−2.43 [−3.44, −1.42]
Noise (State)	1%	0.135 [−0.206, 0.476]	0.0083 [−0.375, 0.392]	0.151 [−0.0772, 0.379]	0.16 [−0.0924, 0.413]
	5%	0.639 [−0.506, 1.78]	−0.303 [−1.41, 0.802]	0.0756 [−0.631, 0.782]	−0.0752 [−0.952, 0.802]
Noise (All)	1%	0.106 [−0.157, 0.368]	−0.0423 [−0.297, 0.212]	−0.00957 [−0.253, 0.234]	−0.00676 [−0.263, 0.249]
	5%	0.267 [−0.76, 1.29]	−0.548 [−1.46, 0.369]	−0.532 [−1.44, 0.372]	−0.4 [−1.09, 0.294]

Table 38: **Predictive effects vary with the exposed information.** Read alongside the consistently larger replacement and deviation near the cutoff, these contrasts distinguish predictive direction from economic severity. Near-minus-far ΔRankIC (10^{-3}).

Input	Dose	CSI300		CSI500	
		Ridge	LightGBM	Ridge	LightGBM
State	1%	0.037 [−0.0664, 0.14]	0.0555 [−0.174, 0.285]	−0.0221 [−0.084, 0.0398]	0.122 [−0.0417, 0.286]
	5%	0.311 [−0.267, 0.889]	0.428 [−0.996, 1.85]	0.00853 [−0.322, 0.339]	1.34 [0.259, 2.42]
ret_{10}	1%	0.229 [0.169, 0.289]	0.0663 [−0.00642, 0.139]	0.15 [0.121, 0.178]	0.0459 [−0.0138, 0.105]
	5%	1.52 [1.23, 1.81]	0.41 [0.0361, 0.784]	0.875 [0.72, 1.03]	0.272 [−0.0391, 0.584]
Intraday	1%	−0.226 [−0.351, −0.1]	−0.295 [−0.46, −0.13]	−0.331 [−0.405, −0.257]	−0.0379 [−0.146, 0.07]
	5%	−1.43 [−2.19, −0.673]	−2.22 [−3.09, −1.34]	−1.99 [−2.38, −1.61]	−0.609 [−1.16, −0.0575]
Noise (State)	1%	0.00978 [−0.015, 0.0345]	−0.00587 [−0.0331, 0.0213]	0.00319 [−0.0168, 0.0231]	−0.00366 [−0.0249, 0.0176]
	5%	0.0858 [−0.0232, 0.195]	0.0943 [−0.00344, 0.192]	0.0105 [−0.0709, 0.0919]	0.00784 [−0.0819, 0.0975]
Noise (All)	1%	0.0231 [−0.00512, 0.0514]	−0.00477 [−0.0296, 0.0201]	0.00306 [−0.0233, 0.0294]	0.016 [−0.0102, 0.0422]
	5%	0.114 [−0.00717, 0.235]	0.036 [−0.059, 0.131]	0.0678 [−0.0411, 0.177]	0.0939 [0.0051, 0.183]

E HETEROGENEOUS FUTURE-INFORMATION AUDIT

Diagnostic hierarchy. We construct three nested diagnostic vectors. *Score-only* (S17) uses 17 descriptors of exposure, score change, clean score dispersion and past portfolio volatility. These in-

clude raw and normalized mean absolute score change, its maximum and RMS, and signed-change fractions. Our *Decision* priority (D28) adds eleven selection and cutoff descriptors: replacement, ordered-selection change, cutoff gap and ties, the fraction of scores within the maximum disturbance of the cutoff, selected and unselected margins, affected-stock cutoff distance, and replacement scaled by past volatility. The added descriptors capture whether a disturbance reaches a decision and the geometry of the stocks it affects.

E.1 A SHARED FORECASTING MODEL AND VARIED CANDIDATE INTERVENTIONS

This task tests audit priorities when future-information interventions produce different score changes. We use All19 Ridge and LightGBM forecasts on CSI 300 and CSI 500, keeping the clean fitted model and portfolio fixed within each universe–model pair. Each intervention substitutes the next-session value of one feature: ten-session return (`ret_10`), intraday open–close return (`oc_gap_1`), or log trading volume (`log_volume`). These channels cover return history, intraday prices and trading activity. The location study's Market-state (State8) forecaster uses a different clean feature set and is excluded from this common-model comparison.

For each decision date we cross the three channels with assigned stock fractions $p \in \{0.01, 0.05, 0.10\}$ and ten fixed seeds. A keyed uniform ordering selects the first $\lfloor pN_t \rfloor$ stocks. Its key depends on universe, date, seed and stock identity, not on scores, selection margins, channel or future availability. The resulting masks are nested across doses and shared across channels and comparable models. No boundary enrichment or selected/unselected quota is imposed.

Training and preprocessing remain fixed. For an assigned stock with an available future endpoint, we use its saved score under that single-feature substitution; every other stock keeps its saved clean score. For these row-wise models, this choice exactly implements the input intervention without score interpolation. We retain the original full-vector availability gate and clean fallback, and distinguish assigned, applied and numerically changed rows. Score changes therefore depend on the feature value, stock and dose rather than a fixed stress amplitude.

Score magnitude varies within every date and every fixed channel–dose group. All 2,740 universe–model–date cross-sections and all 24,660 channel–dose cross-sections contain different mean absolute score changes across candidates, verified at relative tolerance 10^{-12}. Within a channel–dose group, only assigned stocks vary across seeds, so score diagnostics can rank candidates even when their information channel and dose are fixed.

There are 90 candidate interventions per universe–model–date, or 246,600 over the four pairs and 685 decision dates in 2022–2024. The economic calendar has 726 sessions through December 31, 2024; candidate decisions end on November 4, matching the existing clean policy's last selection date.

E.2 ECONOMIC TARGETS, AUDIT CASES AND FORWARD EVALUATION

Each universe–model clean account starts with unit cash on January 4, 2022; its daily path is reproduced against the existing clean reference. Each check starts two continuations from the same post-open-t cash and lot snapshot and changes only the first selection. The subsequent policy and market path are shared, with no further clean entries after the original decision endpoint. Each account then evolves under the unchanged five-sleeve rule, five-session holding horizon, next-open execution and 10 bp costs on each side. Equation (2) averages the absolute paired net-return difference over six sessions, without forcing liquidation of outstanding positions at the endpoint.

At a given universe–model–date, the same complete ordered selection from the shared snapshot produces the same economic continuation. We therefore merge those candidate interventions into one audit case, retain their membership, and use the lexicographically first (channel, numeric dose, seed) candidate's whole diagnostic vector. Every audit case has unit weight, regardless of source multiplicity, and unchanged selections remain eligible. Each selected audit case requires one check.

Audit regressions are fitted separately for each universe and forecasting model. The 2023 fit uses 2022 cases whose complete six-session response ends before January 1, 2023. The 2024 fit expands

to eligible 2022–2023 cases ending before January 1, 2024. Each fit uses training-only feature standardization, Ridge regression with penalty 10 and an intercept, and nonnegative truncation of predicted impact. The numerical solver is SVD. Dates and their candidates are never randomly split. These are forward audit fits on a previously examined market period.

Each retrospective universe–model–month queue receives $\lfloor bn_q \rfloor$ checks at $b \in \{0, 0.05, 0.10, 0.20, 0.50, 1\}$. The primary budget is 10%. All priorities use the same candidates and quotas, with a common outcome-independent tie order. Simple priorities use random ordering, assigned exposure, raw mean absolute score change, its normalized version, root mean square score change, or selected-name replacement. The oracle orders realized impact and provides an upper reference.

The root mean square (RMS) priority gives larger score disturbances more weight than the raw mean-absolute-change priority. This tests whether emphasizing score magnitude can substitute for observing the changed decisions.

E.3 Nested score, decision and state representations

The Score-only priority uses 17 descriptors. To describe the clean cross-section, it includes N_t, K_t, score scale $\sigma_t = \mathrm{IQR}(s_t^0)/1.349$ and a zero-scale indicator. To describe the intervention, it includes assigned, applied and changed fractions; mean absolute, maximum absolute, root mean square, normalized mean absolute and signed mean score change; and positive- and negative-change fractions. Means and fractions use all scored names; normalization divides by σ_t when positive, and one otherwise.

The remaining descriptors give the score changes an economic scale: past clean gross-return volatility ν_t and two interactions. Volatility is the population standard deviation of the latest twenty observed gross returns, in bp, using the available history at the beginning of 2022. The interactions multiply ν_t by normalized mean absolute change and by σ_t, respectively.

Decision retains those 17 descriptors and adds eleven describing which stocks change and where they lie relative to the cutoff: replacement fraction, an ordered-selection-change indicator, normalized clean cutoff gap, cutoff-tie fraction, the fraction of clean scores within the candidate's maximum absolute score change of the cutoff, four mean normalized margins over selected, unselected and adjacent inside/outside half-K stocks, mean normalized cutoff distance over added or removed names, and replacement fraction times ν_t. Empty affected sets have distance zero. The fitting procedure stays fixed, so the comparison measures the added value of these decision and boundary descriptors.

Decision+State adds the 27 holding, maturity, renewal and current-mark projection descriptors specified in Appendix G, giving 55 total. Projections use only current quotes and known maturities; future execution prices and realized future execution eligibility are not diagnostic inputs. Channel labels, seeds, case identifiers and response outcomes are not regression features. All candidate descriptors are fixed before their economic responses are joined.

E.4 Primary comparison, uncertainty and restricted pools

The primary development comparison is Decision minus Score-only impact capture at the 10% budget in the mixed candidate pool. Before computation, we specified a positive paired 95% HAC20 interval as the criterion. Secondary comparisons include replacement minus raw score change, Decision minus raw score change, and Decision+State minus Decision.

In the stratum and pool tables, the paired Capture difference (Gain) is Decision minus Score-only, in pp. The method table also shows Recall and Found to connect recovered impact to cases: 8,861 checks are available among 89,000 cases, of which 28,549 are high-impact. Capture CIs and paired Gain CIs use the same date-level inference, with each comparison retaining its own denominator. The original equal-amplitude task's separate State-minus-Decision comparison remains unchanged.

Inference uses the date-level ratio influence in Equation (23). Within each shared date, pooled numerators and denominators average the four universe–model totals; their ratio preserves the complete distinct-case sums. Paired influences retain the common denominator and covariance. Bartlett HAC bandwidths 5, 10, 20 and 40 are reported, with 20 primary. Recall uses the stipulated tolerance $A_i > 1$ bp/day; zero denominators produce undefined ratios.

In addition to the mixed pool, three channel-specific, three dose-specific and nine channel–dose pools retain the same fitting and checking rules. Each restricted pool independently merges its economically identical candidates and selects its own whole-row representatives. These comparisons separate prioritization within similar interventions from prioritization across different contents and doses. They have overlapping cases and pointwise intervals, not independent-replication or simultaneous-coverage interpretations. Complete methods, budgets, periods and bandwidths are retained in the numerical supplement.

Decision finds more high-impact cases at the primary budget. It corrects 3,926 misses made by Score-only and introduces 2,567, a net increase of 1,359 cases.

Decision also improves capture within every fixed channel–dose pool, with a positive HAC20 interval in all nine pools. The gain therefore does not arise solely from distinguishing channels or doses. Each pool uses its own denominator and forward fits, so comparisons between pools do not isolate causal dose effects.

Table 39: **Decision information improves audit recovery in development.** Decision finds more impact and more high-impact cases than Score-only at the same 10% budget.

Priority	Capture (%)↑	Recall (%)↑	Found↑
Random	9.92 [9.65, 10.18]	9.95	2,841
Exposure	13.17 [12.20, 14.13]	13.16	3,757
Raw score change	22.11 [19.73, 24.48]	21.47	6,130
Normalized score change	22.10 [20.03, 24.17]	21.47	6,129
RMS score change	21.81 [19.51, 24.11]	21.13	6,031
Replacement	26.49 [24.89, 28.09]	26.45	7,552
Score-only	22.64 [20.42, 24.86]	21.94	6,264
Decision	**26.82** [25.35, 28.30]	**26.70**	**7,623**
Decision+State	26.78 [25.43, 28.13]	26.59	7,590
Oracle	34.79 [32.59, 36.99]	31.04	8,861

Decision finds 1,359 additional high-impact cases with the same checks; the text decomposes corrected and introduced misses.

Table 40: **The development gain holds within both models and both stock universes.** Every stratum has a positive paired interval. Capture (%), 10% budget.

Stratum	Score-only↑	Decision↑	Gain (pp)
CSI300 / Ridge	23.64	**26.90**	+3.26 [2.00, 4.52]
CSI300 / LightGBM	20.33	**26.23**	+5.90 [4.55, 7.25]
CSI500 / Ridge	25.25	**29.02**	+3.76 [2.33, 5.19]
CSI500 / LightGBM	20.78	**25.06**	+4.28 [2.97, 5.59]
Pooled / 2023	22.25	**26.35**	+4.09 [3.22, 4.96]
Pooled / 2024	23.10	**27.40**	+4.30 [2.42, 6.19]

Table 41: **The gain is not only a difference between channels or doses.** Decision also improves recovery within every fixed channel–dose pool. Capture (%), 10% budget.

Pool	Cases	Score-only ↑	Decision ↑	Gain (pp)
Mixed (primary)	89,000	22.64	**26.82**	+4.19 [3.23, 5.14]
Return / all doses	33,621	21.17	**26.55**	+5.38 [3.81, 6.96]
Intraday / all doses	37,367	17.88	**21.61**	+3.73 [2.72, 4.74]
Volume / all doses	22,728	18.71	**38.48**	+19.77 [17.38, 22.15]
1% / all channels	13,759	22.83	**31.13**	+8.30 [6.44, 10.16]
5% / all channels	40,403	21.14	**27.90**	+6.76 [5.55, 7.97]
10% / all channels	49,098	21.06	**25.42**	+4.36 [3.19, 5.53]
Return @ 1%	6,306	23.44	**38.32**	+14.88 [11.73, 18.03]
Return @ 5%	15,404	19.25	**27.87**	+8.62 [6.54, 10.71]
Return @ 10%	17,254	18.05	**23.04**	+4.98 [3.64, 6.32]
Intraday @ 1%	7,323	18.12	**27.51**	+9.39 [7.21, 11.58]
Intraday @ 5%	16,427	15.96	**21.69**	+5.73 [4.66, 6.80]
Intraday @ 10%	17,602	15.01	**18.38**	+3.37 [2.23, 4.51]
Volume @ 1%	4,270	20.80	**66.90**	+46.10 [39.32, 52.88]
Volume @ 5%	11,547	17.28	**44.42**	+27.14 [24.07, 30.21]
Volume @ 10%	15,106	16.37	**36.12**	+19.74 [17.28, 22.21]

F TEMPORAL CONFIRMATION OF DECISION-AWARE AUDIT VALUE

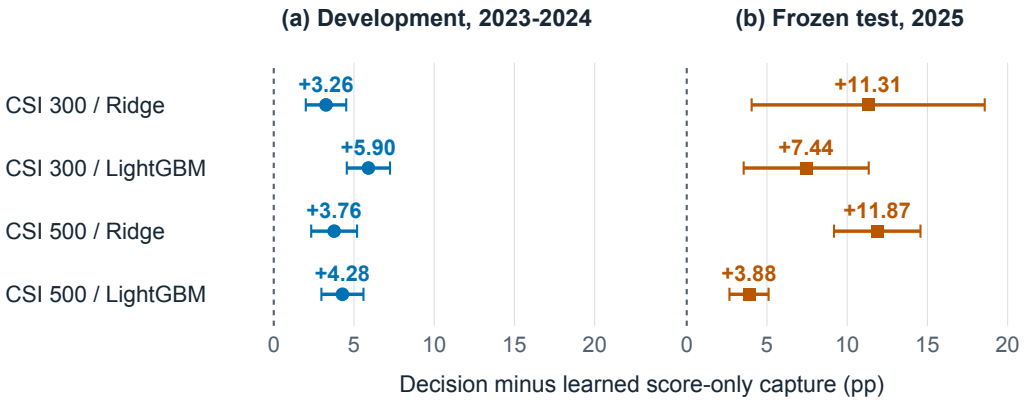


Figure 18: **The audit gain holds across both forecasters and both stock universes.** Decision-minus-Score-only impact capture at the same 10% budget. Points and bars are estimates and paired 95% intervals: HAC20 in development, the fixed four-block rule in 2025. These are secondary pointwise comparisons; Table 1 reports the pooled primary test. Both panels use the same scale.

F.1 A FIXED AUDIT IN A NEW EVALUATION PERIOD

The 2025 test asks whether adding decision information to a learned exposure-and-score priority helps the same number of checks find more economic impact. After the 2022–2024 development study, we specified the comparison, candidates, fitting rules, budget and success criterion before inspecting this study's 2025 predictions or economic responses. This was an internal prospective plan, not an external preregistration. The primary February–December account window matches the prior non-inspection declaration. January observations had already supported earlier analyses, so January–December is a descriptive supplement.

The stock forecasters and audit regressions use different training histories. The All19 stock forecasters use 2019–2023 training observations and the 2024 validation fold, with the previously fixed Ridge and LightGBM configurations and no new tuning. Labels whose exit dates cross their train-

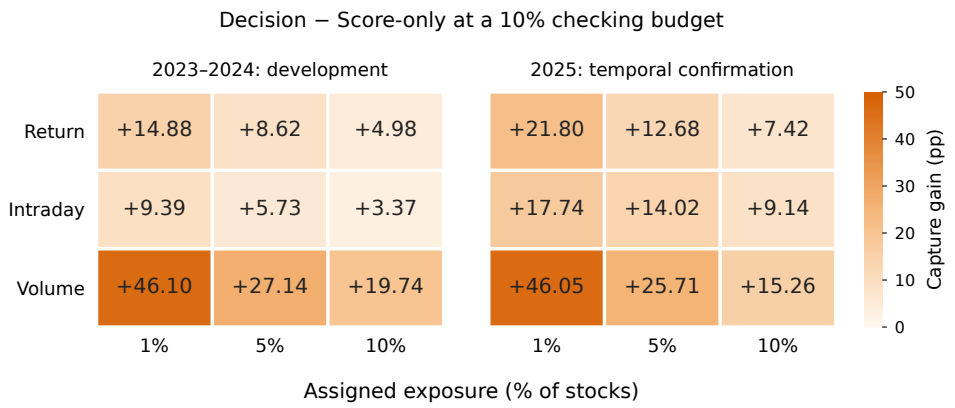


Figure 19: **Decision diagnostics help within the same channel and dose.** Paired Decision-minus-Score-only impact capture at the same 10% monthly checking budget, in percentage points. Both panels show the full three-channel, three-dose grid. Horizontal exposure is the nominal fraction assigned future information, not the checking budget. Each cell uses its own audit cases and historical fits. These are secondary point estimates; overlapping pools have different denominators, so cross-cell differences are not causal dose effects. Pointwise intervals and case counts appear in Appendices E and F.

ing or validation boundaries are removed. Training-only preprocessing, date-effective CSI 300 and CSI 500 membership, and score eligibility are unchanged. Each of the four universe–model pairs has one clean fitted forecaster, shared by all interventions and both account windows.

The *audit* regressions learn economic impact from 2022–2024 intervention cases whose entire six-session response ends before January 1, 2025. We reuse the 17-descriptor Score-only representation (S17), its 28-descriptor decision-aware extension (D28), and the 55-descriptor Decision+State extension specified in Appendix E. Each uses training-only standardization and Ridge regression with penalty 10, an intercept and nonnegative prediction truncation. We fit separate models for each universe–model pair and candidate pool. The procedure, scalers and coefficients were fixed before 2025 response access; no audit model is updated during the test year.

F.2 Interventions and independently funded accounts

We retain the candidates from Appendix E: next-session substitution of ten-session return, intraday open–close return or log volume, crossed with assigned stock fractions of 1%, 5% and 10% and ten fixed seeds. Uniform stock masks are nested across doses and shared across feature channels. The existing full-vector future-availability rule determines which assigned rows receive the single-feature substitution; all other rows retain their clean scores. Thus each universe–model–date has 90 candidate interventions, without conditioning assignment on scores or selection margins.

The primary account spans February 5–December 31, 2025 (225 market sessions), with 184 candidate decision dates from February 5 to November 4. The supplementary account spans January 2–December 31 (243 sessions), with 202 decision dates from January 2 to November 4. These endpoints retain the existing terminal selection rule: the final new entry is on November 5, with its scheduled five-session exit on November 12. Every candidate has all six subsequent response sessions within 2025; no 2026 observations are introduced. After the last selection, accounts continue marking holdings and retrying failed exits through December 31, without forced liquidation. The December account endpoint therefore does not imply candidate decisions or six-session audit responses extending through December.

Each window starts independently with unit cash and no holdings. The February account therefore inherits neither January positions nor returns; its past-return volatility uses only its own history. The windows share fitted forecasters and scores, but their accounts evolve separately. At each candidate date, a check starts the clean and intervened accounts from the same post-open cash and lot snapshot. Only the first selection differs; subsequent clean decisions and market observations are shared. Both

accounts retain self-financing, five-sleeve allocation, next-open execution, five-session scheduled holdings, 10 bp costs per filled side, failed-sale retention and cash-constrained purchases. Impact is the six-session mean absolute paired net-return difference A_i, in bp/day, from Equation (2).

Within each window, universe, model and date, identical complete ordered selections form one audit case. Each case has unit weight, including unchanged selections, and its representative is chosen without outcomes. Every priority faces the same audit cases and quota $\lfloor bn_q \rfloor$ in each retrospective universe–model–month queue, with primary budget $b = 0.10$.

F.3 One primary comparison and a paired temporal interval

The sole primary contrast is D28 minus S17 impact capture in the mixed candidate pool at the 10% budget. Let $\mathcal{C}_{st}$ denote audit cases in stratum s on date t, and I_{mi} indicate that method m checks case i under its monthly quota. Date-level selected and total impact average the four stratum totals:

$$N_{mt} = \frac{1}{4}\sum_{s=1}^{4}\sum_{i\in\mathcal{C}_{st}} I_{mi}A_i, \qquad D_t = \frac{1}{4}\sum_{s=1}^{4}\sum_{i\in\mathcal{C}_{st}} A_i. \tag{19}$$

The pooled capture and its ratio influence are

$$\widehat{c}_m = \frac{\sum_t N_{mt}}{\sum_t D_t}, \qquad \psi_{mt} = \frac{N_{mt} - \widehat{c}_m D_t}{\overline{D}}, \qquad \overline{D} = \frac{1}{184}\sum_t D_t. \tag{20}$$

Thus $\widehat{\theta} = \widehat{c}_{\mathrm{D28}} - \widehat{c}_{\mathrm{S17}}$ compares pooled case-impact ratios, not averages of daily or stratum-specific capture rates. The paired influence is $\psi_t = \psi_{\mathrm{D28},t} - \psi_{\mathrm{S17},t}$.

We partition the 184 chronological decision dates into four consecutive blocks B_j of 46 dates, retaining all four strata together on each date. With $z_j = 46^{-1}\sum_{t\in B_j}\psi_t$ and $\overline{z} = 4^{-1}\sum_j z_j$, the interval is

$$\widehat{\theta} \pm t_{3,0.975}\sqrt{\frac{\sum_{j=1}^{4}(z_j - \overline{z})^2}{4\cdot 3}}. \tag{21}$$

The point estimate remains the full-episode pooled ratio difference; block-specific ratios are not averaged. Pairing precedes uncertainty estimation, preserving the covariance of methods evaluated on the same cases. The three-degree-of-freedom critical value reflects four temporal groups, not four independent markets. A strictly positive lower bound meets the prespecified primary criterion; an interval containing zero is unresolved, and a wholly negative interval is contrary evidence. Zero-denominator ratios are reported as undefined. No primary outcome changes the designated contrast, window or criterion.

F.4 Pre-outcome calibration and secondary reporting

Pre-outcome calibration selected the four-block interval rule. The initial normal-critical-value HAC20 rule attained 87.34%–90.00% coverage in the prespecified null simulations, below the required level. We recorded these results and fixed the four-block rule before a separate calibration, keeping the scientific contrast and point estimator unchanged. Inference from a few long temporal groups is motivated by Ibragimov & Müller (2010); Equation (21) uses a pooled influence approximation rather than invoking an exact finite-sample theorem.

The revised rule met the coverage criterion under the tested scenarios. Across 5,000 replications per setting, it attained 94.02%–94.88% coverage on the fixed 16-setting grid, which combines shared cross-stratum shocks, six-session overlap, 20-date regimes, Gaussian or Student-t_5 innovations, and null or positive paired effects. These simulations assess the finite-episode effect under those dependence scenarios, not arbitrary market dependence. All inference conditions on the fixed fitted forecasting and audit models.

We retain all methods, the four strata, monthly summaries, six budgets and all sixteen candidate pools. The latter comprise the mixed pool, three channel pools, three dose pools and nine channel–dose pools, each with its own audit cases and historical audit fits. Secondary primary-window intervals use the same temporal rule and are pointwise; only the pooled mixed-pool D28–S17 comparison at 10% has the primary confirmation criterion.

In the method table, Primary means the February-start account and Full year means the independently funded January-start account. Primary offers 3,482 checks among 34,984 cases, including 7,894 high-impact cases. Both columns show Capture; Found refers to Primary. Gain in the stratum and pool tables again means Decision minus Score-only Capture, in pp. Undefined zero-denominator quantities are reported as N/A. Restricted pools overlap and are not independent replications or isolated causal dose effects. High-impact recall and corrected or introduced misses use the fixed threshold $A_i > 1$ bp/day. January–December results retain their separate accounts and quotas and are reported as descriptive estimates and counts, without confidence intervals or pooling with the primary window. The study tests temporal transfer within one equity market, with the two universes providing stratified rather than independent-market evidence.

F.5 Complete comparisons and computational checks

Tables 42–44 report methods, strata and restricted pools. Decision+State captures 30.49% versus Decision's 30.24%; the 0.25 pp increment has interval $[-0.40, 0.90]$. The frozen primary test establishes the value of decision information beyond score diagnostics; the added state increment remains uncertain.

The separately funded full-year account gives a similar descriptive comparison: Score-only and Decision capture 22.46% and 30.23% across 38,480 audit cases, with a common budget of 3,830.

The score baseline retains within-date information in the confirmation task. Score-change magnitude varies across candidates on all 736 primary-window universe–model–dates and within all 6,624 channel–dose–date groups. These checks use candidate interventions before ordered-selection deduplication or economic-target access. The baseline is therefore not restricted to ranking dates, as it is in the equal-amplitude control.

Independent calculations reconstruct all 28 score/decision descriptors for 432 sampled candidates, 48 complete counterfactual paths and all eight clean account paths. The economic checks also reconcile the paired daily responses and account projections. Across all 440,784 six-session response rows from both windows, minimum cash is zero and maximum gross exposure is one, with no negative-cash or leverage violation. Separate statistical reconstruction refits all twelve historical audit regressions for the mixed pool and reproduces both windows' rankings, integer quotas, pooled metrics and primary-window intervals. Independent reconstruction covers the mixed-pool Score-only and Decision fits; restricted-pool fits and the full Decision+State representation lie outside that reconstruction. The numerical supplement records this coverage alongside the complete result grid.

Table 42: **The frozen test preserves the advantage of decision-aware priorities.** Replacement already improves recovery; learning adds to it, while extra state features add little. Capture (%), 10% budget.

Priority	Primary ↑	Found ↑	Full year ↑
Random	9.92 [9.21, 10.62]	784	9.78
Exposure	14.29 [10.14, 18.45]	1,144	13.62
Raw score change	21.83 [18.36, 25.30]	1,740	21.63
Normalized score change	21.85 [19.82, 23.89]	1,747	21.71
RMS score change	21.44 [18.66, 24.21]	1,726	21.32
Replacement	29.04 [28.19, 29.90]	2,365	29.00
Score-only	22.56 [19.51, 25.61]	1,792	22.46
Decision	30.24 [28.80, 31.67]	2,508	30.23
Decision+State	**30.49** [28.74, 32.23]	**2,527**	**30.41**
Oracle	41.09 [35.18, 47.01]	3,412	40.91

The February-start account provides the frozen primary test. The separately funded full-year account is descriptive and shows a similar ordering.

Table 43: **The 2025 gain is positive in every model–universe stratum.** The pooled result is the primary test; the remaining comparisons show where the gain occurs. Capture (%).

Stratum	Score-only ↑	Decision ↑	Gain (pp)
Pooled	22.56	**30.24**	+7.68 [4.77, 10.58]
CSI300 / Ridge	25.12	**36.42**	+11.31 [4.04, 18.57]
CSI300 / LightGBM	21.55	**28.99**	+7.44 [3.55, 11.34]
CSI500 / Ridge	23.18	**35.05**	+11.87 [9.18, 14.57]
CSI500 / LightGBM	21.69	**25.58**	+3.88 [2.66, 5.10]

Table 44: **Decision information remains useful within each 2025 channel–dose pool.** The gain persists when the intervention type and assigned exposure are held fixed. Capture (%), 10% budget.

Pool	Cases	Score-only ↑	Decision ↑	Gain (pp)
Mixed (primary)	34,984	22.56	**30.24**	+7.68 [4.77, 10.58]
Return / all doses	12,971	20.00	**27.78**	+7.78 [3.24, 12.31]
Intraday / all doses	12,981	19.72	**29.21**	+9.49 [7.80, 11.18]
Volume / all doses	11,251	18.09	**34.19**	+16.10 [12.89, 19.32]
1% / all channels	4,879	27.52	**44.81**	+17.28 [13.67, 20.89]
5% / all channels	16,077	22.38	**33.20**	+10.82 [7.59, 14.05]
10% / all channels	20,316	20.93	**28.29**	+7.36 [3.64, 11.09]
Return @ 1%	2,348	26.03	**47.83**	+21.80 [13.08, 30.52]
Return @ 5%	6,096	18.61	**31.28**	+12.68 [10.54, 14.81]
Return @ 10%	7,052	17.10	**24.52**	+7.42 [4.36, 10.47]
Intraday @ 1%	2,347	24.43	**42.16**	+17.74 [10.77, 24.70]
Intraday @ 5%	6,083	17.71	**31.74**	+14.02 [8.77, 19.28]
Intraday @ 10%	7,059	17.30	**26.44**	+9.14 [6.17, 12.11]
Volume @ 1%	1,958	20.54	**66.59**	+46.05 [26.81, 65.29]
Volume @ 5%	5,504	16.49	**42.20**	+25.71 [22.67, 28.76]
Volume @ 10%	6,872	15.45	**30.71**	+15.26 [10.09, 20.43]

G EQUAL-AMPLITUDE AUDIT CONTROL AND COMPLETE COMPARISONS

This original equal-amplitude audit uses 22-descriptor Decision and 49-descriptor State variants. These labels are local to this appendix; the heterogeneous audit in the main text uses Decision (D28) and Decision+State (55 descriptors).

G.1 CANDIDATE CONSTRUCTION AND ECONOMIC TARGETS

This control tests whether a perturbation's location helps prioritize economic checks when its magnitude is fixed. We use All-feature CSI 300 clean forecasts and a nominal 5% one-date score stress. A keyed uniform ordering assigns $\lfloor 0.05N_t \rfloor$ stocks, with independent symmetric signs adding $\pm g_m \,\mathrm{IQR}(s_t^0)/1.349$. Calibration follows the adjacent-date procedure in Appendix D, but uses only 2016–2018 forecasts: $g_m = 0.43442031$ for Ridge and 0.43653475 for LightGBM. There is no future-feature value or future-availability gate. Twenty fixed seeds generate the candidates for each model–date. The five-percent intervention dose is distinct from the fraction of candidates that an auditor is allowed to check.

For a fixed model and date, let $q = \lfloor 0.05N_t \rfloor$ and $a_t = g_m \,\mathrm{IQR}(s_t^0)/1.349$. Then, apart from floating-point roundoff, every seed has the same mean absolute score change qa_t/N_t, maximum change a_t, and root mean square change $a_t\sqrt{q/N_t}$. Score-magnitude priority can therefore distinguish dates within a monthly queue, but not the differently located candidates on the same date. The comparison isolates location at matched perturbation size.

Each check starts the clean and stressed accounts from the same post-open-t cash-and-lot snapshot. Only the close-t selection changes. Subsequent clean selections, prices and execution restrictions are shared, while each account's cash, holdings, transactions and fees evolve separately. Equation (2)

measures the mean absolute paired net-return difference over the next six sessions, in bp/day. The horizon covers next-open entry and the scheduled five-session exit, without liquidating unresolved positions at its endpoint.

One economic check, one observation. At each model–date, we merge seeds that produce the same complete ordered selection and verify that their full six-session paired return vectors and targets agree exactly. The lowest seed supplies its whole feature vector, without using outcomes to choose a coordinate or representative. Each distinct check receives unit weight regardless of seed multiplicity; unchanged selections remain eligible. The results retain all source seeds and group membership.

The 55,840 seed responses in 2016–2021 yield 50,074 audit cases. Forward evaluation uses 24,610: 12,896 Ridge and 11,714 LightGBM cases on 684 shared dates, January 2, 2019–October 28, 2021. The yearly date counts are 244, 243 and 197. There are 9,200 high-impact cases with $A_i > 1$ bp/day.

Each model–calendar-month batch permits $\lfloor bn_q \rfloor$ checks at $b \in \{0, 0.05, 0.10, 0.20, 0.50, 1\}$. At the primary 10% budget, the quotas total 2,429 checks: 1,275 Ridge and 1,154 LightGBM, or 9.87% after rounding. All methods share the candidates, quotas and outcome-independent tie order. These are retrospective monthly queues.

Alongside Capture, this control reports the fraction of high-impact cases left unchecked (Missed): 100 minus Recall, not 100 minus Capture. Lower Missed therefore means fewer consequential cases overlooked. The final row of the main control table gives paired State-minus-Decision changes in pp. Budget endpoints are fixed: Capture is zero with no checks and 100% with exhaustive checks.

G.2 All 22 decision and 49 state descriptors

Simple priorities use a fixed random order, assigned-row fraction, mean absolute raw score change over *all* scored stocks, or selected-name replacement fraction. Exposure is nearly constant at this fixed dose; its priority provides a reference rather than testing how to rank different leakage doses. The learned Decision and State priorities both use standardized Ridge regression with penalty 10, an intercept and nonnegative prediction truncation.

For the descriptors below, rank the clean scores in descending order and write $b_t = s^0_{(K_t)}$, $\sigma_t = \mathrm{IQR}(s^0_t)/1.349$, $\bar{\sigma}_t = \sigma_t$ if positive and one otherwise, $d_i = |s^0_i - b_t|/\bar{\sigma}_t$, and $\delta_i = s^a_i - s^0_i$. Let I and O contain the closest $\max\{1, \lfloor K_t/2 \rfloor\}$ selected and unselected names. Let ν_t be the population standard deviation of the twenty most recent observed clean *gross* returns through t, multiplied by 10^4. The complete Decision vector is:

1. **Clean geometry (12).** N_t; K_t; σ_t; $\mathbf{1}\{\sigma_t = 0\}$; $(s^0_{(K_t)} - s^0_{(K_t+1)})/\bar{\sigma}_t$; the fraction with $s^0_i = b_t$; the fraction with $|s^0_i - b_t| \le g_m\sigma_t$; the mean d_i over selected names, unselected names, I, and O (four descriptors); and ν_t.
2. **Observed intervention (8).** Assigned fraction; mean $|\delta_i|$; maximum $|\delta_i|$; root mean square δ_i; mean $|\delta_i|/\bar{\sigma}_t$; replacement fraction R_t; an indicator that the ordered selection changes; and mean d_i over the symmetric difference of clean and stressed selections, set to zero when empty. Score means in this group use all scored names.
3. **Economic-scale interactions (2).** $R_t\nu_t$ and $(\mathrm{mean}_i\, |\delta_i|/\bar{\sigma}_t)\nu_t$.

State contains these same 22 descriptors plus the following 27. Let V_t be the common current wealth. All holding, due and residual weights sum lot notionals at their current or carried marks and divide by V_t. Due weight counts non-residual lots scheduled for the next execution date. For a set of names U, renewal capacity is

$$\mathrm{Roll}(U) = \frac{1}{V_t} \sum_{i \in U} \min\left\{ \mathrm{DueNotional}_i, \frac{V_t}{5K_t} \right\}, \tag{22}$$

excluding names with pre-existing residual holdings or unavailable current marks. Define added and removed sets as $S^a \setminus S^0$ and $S^0 \setminus S^a$, respectively.

1. **Current portfolio (10).** Cash weight; total held weight; number of lots divided by $5K_t$; next-date due weight; residual weight; held weight in the clean selection; renewal capacity in the clean

selection, I, and O (three descriptors); and the fraction of clean selected names with a residual holding.

2. **Affected holdings (8).** Added and removed held weight; added and removed next-date due weight; added and removed residual weight; and added and removed renewal capacity.
3. **Current-mark projections (9).** Projected asset-notional L_1 difference L_t; absolute cash difference C_t; $L_t + C_t$; signed fee difference F_t and $|F_t|$; signed renewal difference U_t and $|U_t|$; $L_t\nu_t$; and $L_t\nu_t + |F_t|$. Asset, cash and renewal differences divide by common V_t; fee differences additionally multiply by 10^4 and are in bp.

What the projection assumes. One hypothetical clean execution and one stressed execution take place at the next scheduled open from the same observed post-open-t snapshot, holding observed open-t quotes fixed. The projection uses the same allocation, costs, renewal and financing rules as the economic account. Hypothetical eligibility requires a finite positive current quote; an absent quote remains absent, with held lots marked under the carried-mark rule. The projected pre-trade NAV must equal current NAV. Future prices and realized future fill flags are not inputs. If $\widehat{w}_i^a$ and $\widehat{c}^a$ are projected asset notionals and cash divided by V_t, then $L_t = \sum_i |\widehat{w}_i^a - \widehat{w}_i^0|$ and $C_t = |\widehat{c}^a - \widehat{c}^0|$. F_t and U_t similarly compare the two projections' fees and internal renewal. These descriptors measure differences under stated current-mark planning assumptions, not predicted actual fills.

G.3 Forward fitting and paired uncertainty

We fit a separate audit regression for each base forecasting model before each prediction year. The first fit uses mature 2016–2018 cases; the window expands before 2020 and 2021. A training case is eligible only when its full six-session response ends before January 1 of the prediction year. Dates and their seed groups are never randomly split. Standardization uses training means and population standard deviations, replacing zero standard deviations by one. Feature lists, penalty, tolerance, budgets and the primary comparison remain fixed without prediction-year tuning. These forward fits use previously selected forecasts and examined development responses; they are not an untouched temporal confirmation.

For each model–date, sum impact across audit cases to obtain denominator d_t and checked impact n_t^g. Pooled analysis averages the two models' numerators and denominators within the shared date. Thus $C_g = \overline{n}^g/\overline{d}$ equals the ratio over audit cases, not an average of daily ratios or of model-specific capture percentages. Its date influence is

$$z_t^g = \frac{n_t^g - C_g d_t}{\overline{d}}. \tag{23}$$

For the paired State-minus-Decision contrast, use $z_t = z_t^{\text{State}} - z_t^{\text{Decision}}$, retaining their common denominator and covariance. With centered date influences $\widetilde{z}_t$ and T dates, the Bartlett HAC estimate is

$$\widehat{\text{Var}}(\overline{z}) = \frac{1}{T}\left[\widehat{\gamma}_0 + 2\sum_{\ell=1}^{\min(L,T-1)}\left(1 - \frac{\ell}{L+1}\right)\widehat{\gamma}_\ell\right], \quad \widehat{\gamma}_\ell = \frac{1}{T}\sum_{t=\ell+1}^{T}\widetilde{z}_t\widetilde{z}_{t-\ell}. \tag{24}$$

Pointwise 95% intervals use the estimate plus/minus $\Phi^{-1}(0.975) \simeq 1.96$ standard errors, with $L \in \{5, 10, 20, 40\}$ and lag 20 as the main report. They are not clipped to a feasible range. Recall replaces impact by $\mathbf{1}\{A_i > 1\}$; missed fraction is one minus recall. Zero impact or high-impact denominators give an undefined ratio, not zero error. Intervals are not simultaneous across methods, budgets and years.

G.4 ALL BUDGETS, MODELS, AND YEARS

Table 45: **Decision changes remain informative when score magnitudes are matched.** Replacement and learned decision priorities recover more impact than score magnitude alone at the same checking budget.

	Capture (%)			Missed (%)
Priority	Ridge ↑	LightGBM ↑	Pooled ↑	Pooled ↓
Random	9.94	10.29	10.08	90.00
Exposure	10.53	10.39	10.47	89.55
Score change	9.71	11.11	10.28	89.78
Replacement	18.36	24.00	20.67	78.97
Decision	18.32	23.92	20.62	78.79
State	**18.62**	**24.03**	**20.84**	**78.67**
Oracle (upper bound)	28.17	34.84	30.91	73.60
State − Decision (pp)	+0.30	+0.10	+0.22	-0.12

State corrects 305 Decision misses and introduces 294, for 11 additional high-impact cases.

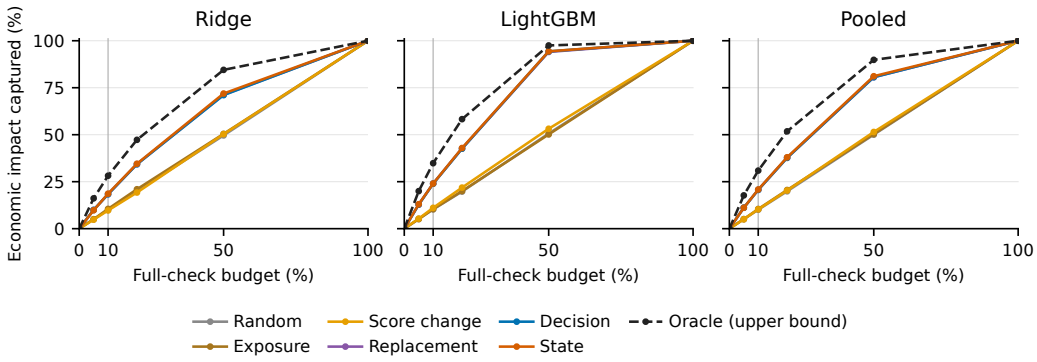


Figure 20: **Matched-magnitude score-stress control.** Forward 2019–2021 impact capture under equal-amplitude, fixed-dose candidates. Every priority faces the same audit cases and rounded monthly quota. Score magnitude is equal across same-date candidates; the priority can distinguish dates, not the locations perturbed within one date. The original primary State-minus-Decision comparison is retained at 10%. The outcome-informed oracle is an upper reference.

At the primary 10% budget, the pooled State-minus-Decision gain is $+0.22$ pp with HAC20 interval $[-0.19, 0.64]$. Each model's interval and the pooled interval include zero at every reported bandwidth (Table 48). State finds 1,962 of 9,200 high-impact cases versus Decision's 1,951, a recall increment of $+0.12$ pp $[-0.36, 0.60]$. At 5%, pooled State capture is lower than Decision capture. The original primary comparison therefore remains statistically unresolved.

Tables 46 and 47 report every method, both base models and every evaluation year. The machine-readable supplement retains their pointwise intervals, missed-case fractions, counts and all bandwidths.

Table 46: **Decision-aware priorities retain their advantage across checking budgets.** Complete method comparisons, Capture (%), 2019–2021.

Model	Priority	5% ↑	10% ↑	20% ↑	50% ↑
Ridge	Random	5.02	9.94	19.99	49.62
	Exposure	5.02	10.53	20.98	50.46
	Score change	4.80	9.71	19.23	50.28
	Replacement	9.96	18.36	34.35	71.24
	Decision	**9.98**	18.32	34.26	71.07
	State	9.81	**18.62**	**34.55**	**71.90**
	Oracle	16.14	28.17	47.27	84.50
LightGBM	Random	5.21	10.29	19.75	50.61
	Exposure	5.07	10.39	20.01	50.14
	Score change	5.34	11.11	21.86	53.17
	Replacement	12.80	24.00	42.63	94.04
	Decision	12.83	23.92	42.56	94.31
	State	**13.01**	**24.03**	**42.91**	**94.36**
	Oracle	19.97	34.84	58.28	97.50
Pooled	Random	5.10	10.08	19.89	50.03
	Exposure	5.04	10.47	20.58	50.33
	Score change	5.02	10.28	20.31	51.46
	Replacement	11.13	20.67	37.75	80.60
	Decision	**11.15**	20.62	37.66	80.60
	State	11.12	**20.84**	**37.98**	**81.12**
	Oracle	17.71	30.91	51.79	89.83

Table 47: **Decision-aware recovery persists across years, while the best priority varies.** Annual Capture (%) at the 10% budget.

Model	Year	Random ↑	Exposure ↑	Score ↑	Replacement ↑	Decision ↑	State ↑	Oracle ↑
Ridge	2019	10.20	9.89	9.22	17.99	17.94	**18.36**	28.99
	2020	9.93	11.61	10.20	**18.36**	18.31	18.27	28.45
	2021	9.72	9.97	9.62	18.68	18.66	**19.22**	27.16
LightGBM	2019	10.48	10.08	11.20	**23.19**	22.64	22.69	34.88
	2020	9.70	10.66	11.81	23.28	23.50	**23.93**	33.83
	2021	10.79	10.40	10.18	25.67	**25.71**	25.49	35.99
Pooled	2019	10.32	9.97	10.06	**20.20**	19.93	20.19	31.49
	2020	9.83	11.21	10.88	20.41	20.48	**20.63**	30.70
	2021	10.14	10.14	9.84	21.40	21.40	**21.66**	30.60

Table 48: **The State-minus-Decision comparison is stable across bandwidths.** Every paired interval includes zero, for both Capture and Recall (pp).

Model	Period	Budget (%)	Lag	Capture gain	Recall gain
Ridge	All	10	5	+0.30 [-0.48, 1.08]	+0.25 [-0.64, 1.13]
			10	+0.30 [-0.43, 1.04]	+0.25 [-0.60, 1.09]
			20	+0.30 [-0.35, 0.95]	+0.25 [-0.51, 1.00]
			40	+0.30 [-0.21, 0.81]	+0.25 [-0.38, 0.88]
LightGBM	All	10	5	+0.10 [-0.48, 0.69]	-0.05 [-0.74, 0.63]
			10	+0.10 [-0.39, 0.60]	-0.05 [-0.64, 0.54]
			20	+0.10 [-0.30, 0.51]	-0.05 [-0.56, 0.46]
			40	+0.10 [-0.22, 0.43]	-0.05 [-0.44, 0.33]

Model	Period	Budget (%)	Lag	Capture gain	Recall gain
Pooled	All	10	5	+0.22 [-0.32, 0.76]	+0.12 [-0.48, 0.72]
			10	+0.22 [-0.27, 0.71]	+0.12 [-0.44, 0.68]
			20	+0.22 [-0.19, 0.64]	+0.12 [-0.36, 0.60]
			40	+0.22 [-0.08, 0.53]	+0.12 [-0.25, 0.49]

Planning cost and scope of the budget. State descriptors require planning work beyond the complete six-session checks counted by the budget. Constructing current-mark descriptors requires 58,632 logical clean/stressed planning requests; reuse of identical ordered selections reduces these to 50,316 one-step projections. The twelve model–year feature sets take 74.1–93.1 seconds each with six concurrent workers; their elapsed times sum to 1,042.6 seconds, not the overall parallel wall time. Reconstructing the clean initial histories takes a further 44.2 seconds. The full response targets already existed from the earlier study. These measurements cover state-feature construction. Appendix N separates preparation, ranking and continuation costs for the main audit.

H THE SELF-FINANCING ECONOMIC ACCOUNT

This account converts selected stocks into positions paid for with available cash. The information and location comparisons share its portfolio rule and market path; each audit pair also starts with the same cash and lots. Execution-clock experiments change their specified timing assumption. We describe the order of renewal, sales and purchases below, then report accounting checks for each study.

Adjusted units and marks. An ordinary full path starts with cash one and no positions. Lot ℓ holds $u_{\ell,t-1} \geq 0$ adjusted units, a scheduled exit date and a residual indicator. At the execution bar on date t, its mark is the stock's HFQ price $\widetilde{P}_{\ell,t}$. The next-open experiments use opening prices; the same-close comparison uses closing prices for both transactions and valuation. When a quote is unavailable, the lot retains its last valid mark for valuation but cannot execute or renew internally at that mark. Cash earns no interest. Pre-trade wealth is

$$V_t^- = c_{t-1} + \sum_{\ell} u_{\ell,t-1} \widetilde{P}_{\ell,t}. \tag{25}$$

Raw quotes and status fields determine directional eligibility, including suspensions, special-treatment status, volume and one-price limit bars. An eligible external purchase of notional b creates $b/\widetilde{P}_{i,t}$ adjusted units; an external sale converts adjusted units into cash using the same denomination.

HFQ incorporates the provider's backward/post-adjustment for price changes[3]. Corporate actions enter through those adjusted prices, without an additional dividend payment to the account. This consistent unit convention supports self-financing, but does not reconstruct physical shares, subscriptions or shareholder distributions, and therefore does not establish equivalence to all historical corporate-action entitlements.

Targets and maturity-matched renewal. For horizon $h = 5$, the new sleeve targets $V_t^-/5$, before current transaction costs. Each selected stock has intended notional $w_{i,t} = V_t^-/(5K_t)$. Its new lot is due five market sessions after entry. Before external sales, a non-residual lot due on the current date can supply the new sleeve when it selects the same stock. The transferred notional is the smaller of the lot's marked value and the stock's remaining target. The corresponding adjusted units move to the new lot with a new exit date, without changing cash or aggregate units and without a fee. Other sleeves retain their exit schedules. Stocks with pre-existing residuals or stale marks cannot transfer. Renewal thus matches a maturing lot to a new target; it does not net all holdings of the same stock.

Failed sales and affordable purchases. After transfers, remaining due lots and existing residual lots are submitted for sale. A successful sale removes its units and credits notional less the selling fee.

[3] BaoStock adjustment-factor documentation (in Chinese): An Introduction to Adjustment Factors.

An unsuccessful attempt retains all units, releases no cash, pays no fee, and leaves a residual for retry. If filled external sales total S_t and per-side cost is $\kappa = 0.001$, cash after sales is $c_t^s = c_{t-1} + (1-\kappa)S_t$.

Let $d_{i,t}$ be a selected stock's unfilled target after internal renewal, and $D_t = \sum_i d_{i,t}$. Freeze the common affordability factor

$$\alpha_t = \begin{cases} \min\{1, c_t^s/[(1+\kappa)D_t]\}, & D_t > 0, \\ 0, & D_t = 0. \end{cases} \tag{26}$$

A positive intention $\alpha_t d_{i,t}$ fills only when buying that stock is permitted and it has no pre-existing residual holding. Otherwise its allocation remains cash rather than being redistributed. Residual status is recorded before current sale attempts: an old residual blocks a new same-stock purchase even if its sale retry succeeds that bar; a first failure today joins the blocking set on the next bar. Zero allocations are not submitted as orders. At the floating-point cash boundary, only a bounded settlement residue may adjust the final affordable purchase; a substantive budget overrun is rejected.

For filled purchases B_t, the resulting account satisfies

$$\begin{aligned} c_t &= c_t^s - (1+\kappa)B_t \geq 0, \\ V_t^+ &= c_t + \sum_\ell u_{\ell,t}\widetilde{P}_{\ell,t} = V_t^- - \kappa(S_t + B_t), \end{aligned} \tag{27}$$

$$r_t^{\text{net}} = \frac{V_t^+}{V_{t-1}^+} - 1 = \underbrace{\frac{V_t^-}{V_{t-1}^+} - 1}_{r_t^{\text{gross}}} - \underbrace{\frac{\kappa(S_t + B_t)}{V_{t-1}^+}}_{r_t^{\text{cost}}}. \tag{28}$$

Insufficient cash reduces new buys; it does not create borrowing to maintain the one-fifth target. Non-negative cash and units imply post-trade gross stock exposure $(V_t^+ - c_t)/V_t^+ \in [0,1]$. Unresolved terminal positions remain valued holdings, without an assumed unfillable sale.

Registered severity and observation paths. The original severity and expanded future-information paths obey the cash and exposure constraints. Table 49 reports all four families for 2022–2024: 2,268 registered paths and 1,646,568 path–date observations. TRAD uses individual mask paths rather than their averaged daily series; the dense and sparse observation families retain registered baselines and endpoint aliases. These counts specify the accounting coverage.

The daily return identity also reconciles to numerical precision. In every family, minimum saved cash is zero, no cash observation is negative, and maximum gross and net stock exposure is one. There are no stale-mark observations, and the maximum absolute residual in the daily return identity, Equation (28), is below 3×10^{-15}. These results document cash-funded execution across the protocol and information comparisons, separately from the location and one-decision audit studies.

For reading the account table, cash is in initial-wealth units ($V_0 = 1$) and stock exposure is a fraction of current wealth. Long-only holdings make gross and signed net exposure equal. Event columns give minimum–maximum counts per 726-bar path, including terminal exits and retries. Residual blocks are purchases prevented by pre-existing failed exits; they differ from rejected buy orders. OBS path counts include baseline and endpoint aliases.

Table 49: **The economic paths remain cash-funded despite failed trades.** No family has negative cash or stock exposure above one; blocked execution remains active.

Path family	Paths	Min. cash	Max. exposure	Failed buys	Failed sells	Residual blocks
Severity: single paths	68	0	1	0–71	28–309	0–5
Severity: TRAD masks	400	0	1	0–23	46–231	0–3
OBS: dense	384	0	1	2–80	10–488	0–5
OBS: sparse	1,416	0	1	1–34	6–291	0–7

Execution restrictions affect trades while preserving cash conservation and adjusted-unit accounting.

Exposure-location paths. The location paths also obey the cash and exposure constraints despite failed orders. Their 2,400 paths contain 1,742,400 portfolio–date observations, excluding the eight

reused clean references. Minimum cash is zero, gross stock exposure remains in $[0, 1]$, and all paths finish fully liquidated, with terminal wealth in cash and no stale marks. The maximum absolute residual in the daily gross-minus-cost identity is 2.925×10^{-15}. Execution restrictions remain active: per-path failed buys range from 2 to 28, failed liquidation attempts from 10 to 254, and residual holdings block at most four buys, affecting 850 paths. Positive-notional internal transfers occur on 676–680 dates per path. Sale counts include repeated retries, and transfer counts count dates, not individual orders. These diagnostics summarize the recorded execution paths.

Both sets of full-path records distinguish rejected positive buys from residual-blocked buys and count failed sales again at each retry. They do not separately count proportional scaling or zero-cash orders, so the event columns do not cover every reduced or skipped purchase intention. Internal renewals are not counted as external fills.

State comparisons and 2025 audit continuations. The fixed-selection historical-state experiment has its own paired six-session reconciliation in Appendix I.2. The 2025 audit instead retains two independently funded account windows. Across its 440,784 saved six-session response rows, minimum cash is zero and maximum gross exposure is one, with no negative-cash or leverage violation. Appendix F specifies its independent reconstruction scope. These shorter candidate responses and their terminal holdings are not pooled with the 2,268 registered full paths or the 2,400 location paths; the location study's complete terminal liquidation is not assumed for the other studies.

Controlled accounting checks. The same portfolio function passes 48 targeted tests. Constant positive HFQ rescalings by $1/8$, 8 and 13 preserve wealth paths and inversely rescale units to numerical precision. Split and ex-distribution proxies combine a raw-price drop with a stipulated adjustment factor to hold adjusted prices constant; wealth is preserved without a separate cash payment. These tests verify the adjusted-unit convention. A flat-price round trip charges 10 bp per filled side, or 4 bp of initial portfolio wealth for a one-fifth sleeve. Further tests cover pro-rata scaling as five sleeves exhaust cash, retained rejected-buy allocations, failed-exit capital constraints, maturity-specific transfers, and the timing of residual-based purchase blocks.

I PORTFOLIO-STATE CONTROLS AND PRE-STRESS SENSITIVITY PREDICTION

We give the construction and historical comparisons supporting Section 4.4, then examine whether sensitivity can be predicted before a stress is drawn. This prediction task uses average response and MAE; the later audit ranks observed candidates by economic impact. Each retains its original outcome and evaluation criterion.

I.1 A FIXED SUBSTITUTION NEED NOT HAVE A FIXED ECONOMIC EFFECT

Proposition 1 (Same substitution, opposite economic effects)**.** *Under the five-sleeve rule with per-side cost $0 < \kappa < 1$, fix equal initial wealth V, flat valid prices, unrestricted execution and a clean selection of A versus an intervened selection of B ($K = 1$). There exist three admissible initial states for which the first-execution net-return difference, intervened minus clean, is respectively*

$$-\frac{2\kappa}{5(1+\kappa)}, \qquad +\frac{2\kappa}{5(1+\kappa)}, \qquad 0. \tag{29}$$

Proof of Proposition 1. Write $q = V/5$. All marks equal one. In H_A, cash is zero, an A lot worth q is due at the next open, and $V-q$ is held in an unselected stock C due one session later. Clean renews A without a fee. Stress sells A, obtaining $(1-\kappa)q$, and buys B with notional $B = (1-\kappa)q/(1+\kappa)$. Its total fee is $F = \kappa(q+B) = 2\kappa q/(1+\kappa)$; flat prices give return difference $-F/V$. In H_B, replace the due A lot by B: stress renews internally, clean pays the same external fees, and the difference is $+F/V$. In H_0, all wealth is in C due one session later. Neither branch has cash or a due sale, so both intended buys scale to zero. The difference is zero. All states obey nonnegative cash and the five-session schedule. □

At the study's cost rate, these are -3.996, $+3.996$ and 0 bp; all three vanish at zero cost. The substitution and prices stay fixed while the executed trades change. These specified endowments

establish a possibility, without asserting their historical frequency. The construction extends to $K > 1$ and $h \geq 2$ by using $q = V/(hK)$ with $K - 1$ common selected names lacking due holdings. The external arm divides its available cash across slots; its total purchase and fees retain the same expression.

Multi-session accounting. With common valid marks P_j, aggregate carried units u^a_{j-1}, fees F^a_j, and branch difference Δ, self-financing implies

$$\Delta V_j = \Delta V_{j-1} + \Delta u_{j-1}^\top (P_j - P_{j-1}) - \Delta F_j,$$
$$\Delta r_j = \frac{\Delta u_{j-1}^\top (P_j - P_{j-1}) - \Delta F_j - r_j^0 \Delta V_{j-1}}{V^1_{j-1}}. \tag{30}$$

Subtract the two wealth recursions and substitute $r^a_j = V^a_j / V^a_{j-1} - 1$ to obtain these identities. With equal initial wealth, telescoping and the triangle inequality yield

$$|\Delta V_J| \leq \sum_{j,i} |\Delta u_{j-1,i}(P_{j,i} - P_{j-1,i})| + \left| \sum_j \Delta F_j \right|. \tag{31}$$

For carried marks, replace the unit–price term by the difference in actual lotwise marked profit, ΔG_j. Then $\Delta V_J = \sum_j \Delta G_j - \sum_j \Delta F_j$ from a common initial state, with bound $\sum_j |\Delta G_j| + |\sum_j \Delta F_j|$. These identities describe realized paths using their positions and marks; score changes alone do not supply the quantities needed to evaluate them. Eighteen fixed constructions ($K \in \{1, 2, 30\}$, three initial states, two cost rates) and five additional price/failure paths reconcile fifty paired transitions, including cancellation and changing wealth denominators.

I.2 Historical versus equal-wealth all-cash state

Holding selections fixed isolates what the starting account adds. Historical holdings and equal-wealth cash generate two six-session impacts (A_{hist} and A_{cash}); their paired difference ($\Delta A = A_{\text{hist}} - A_{\text{cash}}$) shows which account is more sensitive. A negative value means lower absolute sensitivity under historical holdings, not a negative investment return. Table 50 averages twenty seeds and 684 dates, pooling models within date; its CI uses paired HAC20 inference.

Table 50: **The same selection change has a different impact from a different starting account.** Historical holdings slightly reduce average sensitivity relative to equal-wealth cash. Impact A (bp/day).

Model	Historical	Cash	ΔA	CI
Ridge	0.9039	0.9339	-0.0300	[-0.0330, -0.0269]
LightGBM	0.6337	0.6494	-0.0156	[-0.0180, -0.0132]
Pooled	0.7688	0.7916	-0.0228	[-0.0250, -0.0206]

The 2019–2021 CSI 300 experiment fixes the complete clean/stressed ordered selection pair, subsequent decisions, prices and execution restrictions. Both branches start from either the historical cash-and-lot snapshot or an all-cash endowment of identical wealth. This changes holdings, maturities and cash together, rather than liquidating the historical account without fees or changing only its cash ratio. The one-date stress is the same 2016–2018 calibrated, ungated stress used in Appendix G. The all-cash account initially funds only the next one-fifth sleeve. Table 50 reports the comparison over twenty seeds, two models and 684 shared dates.

Paired estimand and complete comparisons. For initial state H, decision date t and seed s, write

$$A(H, t, s) = \frac{10^4}{6} \sum_{j=1}^{6} \left| r^{\text{stress}}_{t+j}(H, s) - r^{\text{clean}}_{t+j}(H) \right|. \tag{32}$$

The primary contrast averages $A(H_{\text{hist}}, t, s) - A(H_{\text{cash}}, t, s)$ over the twenty seeds within date and then over dates; pooling averages the two model contributions on their shared date. Its sign

indicates which starting state produces the larger absolute economic response; it is not the sign of the intervention's return effect.

To retain differences that might cancel across dates or seeds, we also report mean $|\Delta A|$, taking the absolute value before pooling. Both summaries are in bp/day; Dates counts the temporal observations used for the paired intervals. These comparisons condition on the existing base forecasts and development market path.

Historical state reduces the pooled six-session absolute response by 0.0228 bp/day, or 2.88% of the all-cash response. The difference is negative for both models in all three years and at every reported bandwidth. This comparison isolates account state's effect on the impact of fixed selections; the audit experiments separately assess ranking performance.

The complete comparisons retain variation across models and years. Table 51 reports every model–year comparison and the mean absolute state difference, which is 0.0308 bp/day when pooled. The 684 evaluation dates comprise 244 in 2019, 243 in 2020 and 197 in 2021. Table 52 uses the same paired observations and point estimates at Bartlett bandwidths 5, 10, 20 and 40. Dates, rather than seeds or model labels, are the temporal sampling units for these pointwise descriptive intervals.

Economic reconciliation. The new cash-state continuations reconcile with the account equations. The 27,360 seed continuations contain 164,160 paired daily observations, with a separately continued cash-clean reference for each model–date. The complete ordered selection pair, future clean decisions, six market dates and execution flags agree across the two initial-state conditions. Neither continuation resets its positions after entry or forces liquidation at the six-session endpoint. Every first-execution response satisfies the paired cost identity; subsequent steps satisfy the lotwise marked-profit wealth recursion and its conservative bound, including carried marks, with each branch's own return denominator. Across the new continuations, minimum cash is zero, maximum gross holdings divided by NAV is one, and the largest absolute wealth-recursion residual is below 3.55×10^{-15} account units. These checks concern the historical state intervention, separately from the 2025 audit confirmation.

Table 51: **The state effect recurs in both models and all three years.** Historical-minus-cash sensitivity is negative throughout; its magnitude varies across the paired accounts. Impact A (bp/day).

Model	Year	Dates	Historical	Cash	ΔA	Mean $\lvert\Delta A\rvert$
Ridge	2019	244	0.7597	0.7838	-0.0241 [-0.0281, -0.0200]	0.0312
	2020	243	0.9086	0.9402	-0.0316 [-0.0365, -0.0267]	0.0421
	2021	197	1.0767	1.1119	-0.0352 [-0.0398, -0.0306]	0.0457
	All	684	0.9039	0.9339	-0.0300 [-0.0330, -0.0269]	0.0393
LightGBM	2019	244	0.5649	0.5825	-0.0176 [-0.0219, -0.0133]	0.0224
	2020	243	0.6580	0.6759	-0.0179 [-0.0219, -0.0140]	0.0254
	2021	197	0.6891	0.6994	-0.0103 [-0.0126, -0.0081]	0.0182
	All	684	0.6337	0.6494	-0.0156 [-0.0180, -0.0132]	0.0222
Pooled	2019	244	0.6623	0.6831	-0.0208 [-0.0247, -0.0170]	0.0268
	2020	243	0.7833	0.8081	-0.0248 [-0.0285, -0.0210]	0.0338
	2021	197	0.8829	0.9057	-0.0228 [-0.0256, -0.0199]	0.0319
	All	684	0.7688	0.7916	-0.0228 [-0.0250, -0.0206]	0.0308

Table 52: **The historical-minus-cash effect remains negative across bandwidths.** Every interval lies below zero. Paired ΔA (bp/day).

Year	Lag 5	Lag 10	Lag 20	Lag 40
Ridge				
2019	[-0.0280, -0.0202]	[-0.0281, -0.0201]	[-0.0281, -0.0200]	[-0.0283, -0.0199]
2020	[-0.0374, -0.0258]	[-0.0370, -0.0262]	[-0.0365, -0.0267]	[-0.0359, -0.0273]
2021	[-0.0409, -0.0295]	[-0.0400, -0.0304]	[-0.0398, -0.0306]	[-0.0397, -0.0307]
All	[-0.0331, -0.0269]	[-0.0330, -0.0270]	[-0.0330, -0.0269]	[-0.0333, -0.0267]
LightGBM				
2019	[-0.0213, -0.0140]	[-0.0216, -0.0136]	[-0.0219, -0.0133]	[-0.0223, -0.0130]
2020	[-0.0216, -0.0142]	[-0.0218, -0.0141]	[-0.0219, -0.0140]	[-0.0223, -0.0136]
2021	[-0.0132, -0.0074]	[-0.0131, -0.0075]	[-0.0126, -0.0081]	[-0.0123, -0.0084]
All	[-0.0177, -0.0135]	[-0.0179, -0.0134]	[-0.0180, -0.0132]	[-0.0184, -0.0128]
Pooled				
2019	[-0.0241, -0.0176]	[-0.0244, -0.0173]	[-0.0247, -0.0170]	[-0.0251, -0.0166]
2020	[-0.0290, -0.0205]	[-0.0288, -0.0207]	[-0.0285, -0.0210]	[-0.0286, -0.0210]
2021	[-0.0264, -0.0192]	[-0.0259, -0.0196]	[-0.0256, -0.0199]	[-0.0254, -0.0201]
All	[-0.0250, -0.0206]	[-0.0250, -0.0206]	[-0.0250, -0.0206]	[-0.0251, -0.0205]

I.3 Predicting sensitivity before a stress is drawn

The earlier prediction task asks how sensitive an account will be before a particular stress mask or score change is observed. Its target is the *twenty-seed mean* six-session response for each date. It uses the first twelve clean-geometry/volatility descriptors listed in Appendix G, with or without the ten current- portfolio descriptors. Neither predictor sees realized stress or selection changes. Both use standardized Ridge regression with penalty 10 and nonnegative truncation. The training-mean mature response is a constant baseline. The expanding 2016–2018 onward fits require every target to mature before January 1 of the prediction year, giving 706, 950 and 1,193 training dates before 2019, 2020 and 2021. Six preceding-year dates are excluded at each boundary. Evaluation uses the same 684 dates.

The tables call the constant predictor Mean, the geometry/volatility predictor Boundary, and its state extension State. Prediction quality is measured by mean absolute error (MAE), in bp/day, so lower is better. The paired MAE reduction (Gain) is Boundary minus State; a positive value favors State. Predictive R^2 measures squared-error improvement over the fold-specific training-mean predictor; a negative value means worse performance than that baseline. Pooled errors average the two signals within each shared date. Gain CIs use HAC20, with the same observations at lags 5, 10, and 40 for sensitivity.

State descriptors slightly improve pooled prediction error. MAE falls from 0.27004 bp/day for the constant to 0.24296 for geometry/volatility (10.03%), then to 0.24013 with state (another 1.16%). The paired state improvement is 0.00283 bp/day with HAC20 interval $[0.00045, 0.00521]$; all four bandwidth intervals are positive. The pooled gain masks a difference between models.

The state benefit depends on the forecasting model. Ridge MAE improves 2.94%, or 0.00789 bp/day $[0.00323, 0.01255]$; LightGBM MAE increases 1.03%, with gain -0.00223 $[-0.00470, 0.00024]$. These directions persist in every evaluation year. The original criterion required MAE improvement for both models, whereas the observed improvement is confined to Ridge. Pooled R^2 rises from 0.1785 to 0.2101; LightGBM's 2021 values remain negative relative to the training-mean predictor. This pre-stress MAE task is separate from candidate-level impact capture.

Tables 53–55 retain every model–year MAE and squared-error comparison, together with the pooled paired MAE intervals at all four bandwidths. The labels Boundary and State in these tables refer to the pre-stress predictors described above, not to the later Score-only (S17), Decision (D28) and Decision+State (55-descriptor) audit priorities.

Table 53: **The state contribution to sensitivity prediction depends on the model.** Complete model–year comparisons of MAE (bp/day); paired Gain is Boundary minus State.

Model	Period	Dates	Mean ↓	Boundary ↓	State ↓	Gain
Ridge	2019	244	0.2191	0.2046	**0.1998**	+0.00481 [−0.00138, +0.01100]
	2020	243	0.3042	0.2686	**0.2632**	+0.00540 [−0.00218, +0.01299]
	2021	197	0.3922	0.3485	**0.3337**	+0.01477 [+0.00553, +0.02401]
	Pooled	684	0.2992	0.2688	**0.2609**	+0.00789 [+0.00323, +0.01255]
LightGBM	2019	244	0.2135	**0.1881**	0.1903	−0.00213 [−0.00599, +0.00174]
	2020	243	0.2555	**0.2182**	0.2187	−0.00048 [−0.00538, +0.00442]
	2021	197	0.2568	**0.2519**	0.2564	−0.00452 [−0.00841, −0.00063]
	Pooled	684	0.2409	**0.2172**	0.2194	−0.00223 [−0.00470, +0.00024]
Both signals	2019	244	0.2163	0.1964	**0.1950**	+0.00134 [−0.00257, +0.00526]
	2020	243	0.2799	0.2434	**0.2409**	+0.00246 [−0.00067, +0.00559]
	2021	197	0.3245	0.3002	**0.2950**	+0.00513 [−0.00011, +0.01036]
	Pooled	684	0.2700	0.2430	**0.2401**	+0.00283 [+0.00045, +0.00521]

Table 54: **The state benefit is model-dependent under squared error too.** LightGBM retains negative out-of-sample R^2 in 2021.

Model	Period	Boundary ↑	State ↑
Ridge	2019	0.1415	**0.1867**
	2020	0.2910	**0.3364**
	2021	0.2073	**0.2680**
	Pooled	0.2213	**0.2732**
LightGBM	2019	**0.0414**	0.0184
	2020	0.2353	**0.2468**
	2021	**-0.0276**	-0.0406
	Pooled	**0.1007**	0.0954
Both signals	2019	0.1029	**0.1218**
	2020	0.2687	**0.3006**
	2021	0.1388	**0.1780**
	Pooled	0.1785	**0.2101**

Table 55: **Bandwidth changes leave the prediction conclusion unchanged.** Ridge and pooled MAE gains have positive intervals; LightGBM remains unresolved. Gain (bp/day).

Model	Lag 5	Lag 10	Lag 20	Lag 40
Ridge	[+0.00365, +0.01213]	[+0.00347, +0.01231]	[+0.00323, +0.01255]	[+0.00274, +0.01304]
LightGBM	[−0.00513, +0.00067]	[−0.00499, +0.00053]	[−0.00470, +0.00024]	[−0.00447, +0.00001]
Both signals	[+0.00043, +0.00523]	[+0.00042, +0.00524]	[+0.00045, +0.00521]	[+0.00026, +0.00540]

J PUBLIC IMPLEMENTATIONS: FULL BEFORE-AND-AFTER COMPARISONS

Table 56: **Public implementations expose different links in the decision chain.** FinRL changes portfolio returns through execution timing; LumiBot changes requested positions through a future-price lookup.

System	Change	Consequence	Interpretation
FinRL execution timing	Same close versus next-session open; a separate PPO fit for each convention at seeds 11, 23, and 37	$\Delta R = R_{\text{same}} - R_{\text{next}}$ is -1.88, $+2.28$, and $+1.74$ percentage points; mixed sign	Released modelling convention, not a reported defect. Chronological test returns are independently reproduced and show no uniform inflation direction.
LumiBot future-price incident	Corrected release versus the preceding revision on the public incident example and sizing path	Price $97.43 \rightarrow 90.98$; requested shares $102 \rightarrow 109$; nominal exposure $+636.86$ units	The maintainers disclosed and corrected the defect. The decision effect is independently reproduced; the approximately 2% production-return effect is maintainer-reported only.

The public cases in Table 56 show how information and execution timing affect two different outcomes: a trained policy's portfolio return in FinRL and the number of shares requested in LumiBot. We fixed each design before its local measurements. These implementation examples trace specific errors beyond our common simulator; the CSI experiments evaluate audit prioritization.

Independently reproduced quantities are those obtained in these experiments; maintainer-reported quantities are identified separately. The two cases are illustrative, not a representative software sample.

J.1 FINRL: EXECUTION CONVENTION AND A NEWLY TRAINED POLICY

The released FinRL v0.3.8 stock-trading environment supplies the agent with the completed current close and executes its action at that close before advancing the day.[4] We compare that convention with fills at the next session's open. Both conditions retain the same close-t observation, action constraints, 10-basis-point fee per filled side, and close-$(t+1)$ valuation. The same-close condition uses the original public environment. For next-open execution, we change only the price passed to its inherited buy and sell operations; the policy still cannot observe the next opening price when choosing its action.

Each convention receives a separate PPO fit for seeds 11, 23 and 37 (Schulman et al., 2017). Initial parameters are identical within each same-seed pair, while fitted parameters are allowed to adapt to the different execution environment. The comparison therefore includes how the policy adapts to each execution convention, as well as the prices at which it trades. Table 57 reports all three paired fits.

Every arm starts with 100,000 cash and no shares. Here Return is the terminal gain $100(V_T/V_0 - 1)$, not annualized Net. Its paired change (ΔReturn) is same-close minus next-open in pp; Wealth is the corresponding terminal account value. This comparison asks which evidence the execution convention produces, without treating the larger reported return as the more valid protocol.

[4] FinRL v0.3.8 source: https://github.com/AI4Finance-Foundation/FinRL/tree/v0.3.8.

Table 57: **Earlier execution does not uniformly inflate test returns.** The same-close advantage changes sign across seeds in the FinRL comparison.

	Same close		Next open		
Seed	Wealth	Return (%)	Wealth	Return (%)	ΔReturn (pp)
11	133,348.98	33.35	135,225.52	35.23	−1.88
23	133,348.98	33.35	131,068.59	31.07	+2.28
37	133,362.37	33.36	131,624.55	31.62	+1.74
Mean	133,353.44	33.35	132,639.55	32.64	+0.71

Changing the execution clock changes returns in different directions across the trained policies. Same-close minus next-open terminal return is -1.8765, $+2.2804$ and $+1.7378$ points, averaging $+0.7139$. The earlier clock therefore does not uniformly inflate test performance.

The accounting records retain every policy, including repeated portfolio paths. Table 58 reports fees, nonzero executions and terminal positions. All six ledgers use 542 decision/valuation transitions and mark final holdings at the March 1, 2013 close of 806.19, without terminal liquidation. Seeds 11 and 23 produce the same observed same-close path despite their distinct fits; retaining all six runs preserves the paired comparison.

In the account breakdown, Fills counts transitions with a nonzero filled share change, excluding unfilled requests. Fees and End cash use initial-cash units; End shares records the unliquidated position.

Table 58: **Execution conventions produce different trading activity and terminal positions.** Distinct policy fits can nevertheless produce identical observed paths.

Seed	Execution	Fills	Fees	End cash	End shares
11	Same close	2	99.57	327.63	165
11	Next open	70	705.71	591.79	167
23	Same close	2	99.57	327.63	165
23	Next open	26	294.20	465.81	162
37	Same close	2	99.56	341.02	165
37	Next open	295	453.42	1,021.77	162

Data and chronological split. We use the public GOOG daily data distributed with `backtesting.py` v0.6.6. After the twenty-observation warm-up, the prepared panel contains 2,128 rows from September 17, 2004 through March 1, 2013. Training ends on December 31, 2010 (1,585 rows); testing starts on January 3, 2011 (543 rows). The six market inputs are the one-session simple return, close-to-five- and twenty-session moving-average ratios, twenty-session return volatility, same-session high–low range divided by close, and volume relative to its twenty-session mean. The training mean and population standard deviation standardize each feature and are applied unchanged to the test period. Cash, holdings, the current close and a constant trade-availability indicator complete the observation.

Learning and portfolio constraints. PPO uses two 64-unit tanh hidden layers, learning rate 2.5×10^{-4}, 256-step rollouts, batch size 64, ten optimization epochs per update, discount factor 0.99, GAE parameter 0.95, clipping parameter 0.2, entropy coefficient zero, value-loss coefficient 0.5, and gradient clipping at 0.5. Advantage normalization is enabled. The nominal 20,000-step training budget completes 20,224 transitions in whole rollouts in every arm. Test actions are deterministic. The single-stock account starts with 100,000 cash, permits at most 100 requested shares per action, is long-only, uses integer shares and does not borrow. Reward scaling is 10^{-4}; both conventions retain the released portfolio constraints. This is a controlled study of the public environment's execution convention, not a reconstruction of the multi-asset experiment in Yang et al. (2020).

J.2 LUMIBOT: FROM FUTURE-PRICE LOOKUP TO REQUESTED EXPOSURE

LumiBot's official April 2026 correction restricts a daily-price lookup to cached bars available by simulation time.[5] We compare the corrected v4.5.0 revision with its immediate predecessor using the project's seven-case regression suite and documented COP example. The correction passes all seven cases, compared with two for the predecessor. Here we reproduce a disclosed software defect; the FinRL case instead changes a released modeling convention.

At 10:30 a.m. New York time on July 1, 2022, the last admissible completed bar in the example closes at 90.98 on June 30. The same cached frame also contains a later row, dated April 17, 2026, at 97.43. The predecessor returns the later price, whereas the corrected lookup returns 90.98. Passing each price through the unchanged public portfolio rebalancing method gives the complete calculation in Table 59.

The target allocation is 10,000 units. Nominal exposure values both requested quantities at the common admissible price of 90.98, so the comparison isolates order sizing rather than a simulated fill or subsequent investment return.

Table 59: **Correcting the future-price lookup increases the requested position.** A lower admissible price raises the request from 102 to 109 shares; this is an order-sizing effect, not a realized return.

Quantity	Before	After
Returned price	97.43	90.98
Requested shares, $\lfloor 10{,}000/P \rfloor$	102	109
Nominal exposure at 90.98	9,279.96	9,916.82
Unused target allocation	720.04	83.18
Allocation covered (%)	92.7996	99.1682
Passed public regression cases	2/7	7/7

Correcting the price lookup increases the requested position. The price falls by 6.45 units and the request rises by seven shares; at the common admissible price, nominal exposure rises by 636.86 units, or 6.3686 percentage points of the target allocation. This calculation measures requested exposure at the admissible price; executed trades and subsequent returns are outside this comparison.

The maintainers separately reported approximately 2% return understatement for a private strategy. Its strategy and trade records are unavailable, so we identify that number as maintainer-reported and do not treat it as a locally reproduced portfolio result.

Public sources and redistribution. The FinRL release is MIT-licensed. The repository distributing the GOOG data is AGPL-3.0-or-later, while the data file identifies neither its original market-data vendor nor a data-specific license; the data are not redistributed. The LumiBot repository's root license and package metadata report GPLv3 and MIT, respectively. We retain the cited source revisions and local numerical results without resolving that licensing discrepancy. The reproduced return and sizing comparisons use no undisclosed private strategy results.

K STUDY CHRONOLOGY AND ANALYSIS ROLES

Internal dated records identify which choices preceded each computation. Table 60 gives the sequence; the records below distinguish earlier research plans, mechanism studies and the completed 2025 audit test. All dates refer to 2026.

[5] LumiBot pull request 993: https://github.com/Lumiwealth/lumibot/pull/993.

Table 60: **The study separates mechanism development from temporal confirmation.** The timeline records what was already known when each design was fixed.

Date	Design and prior knowledge
September 4–6	Severity consolidation and mechanism grid; historical exploration and some later-period anchor results had already been examined.
September 8	Location grid fixed after inspecting 2022–2024 mechanisms; the market period was already examined.
September 9–10	Sensitivity, state controls and equal-amplitude audit; prior development responses informed their designs.
September 13	Heterogeneous future-information audit fixed after earlier stress results; forward evaluation uses an examined market period.
September 13	Separate 2025 audit specification and interval calibration precede study access to 2025 predictions and outcomes at 19:59 UTC.

Development and mechanism evidence. Feature construction, forecasting configurations and economic rules used clean CSI 300 evidence from 2016–2021 and historical exploration. Severity profiles were consolidated on September 4–5 after later-period anchors had been examined. On September 6, the feature/entry study fixed six groups, four leads, four training–scoring cells and all sparse doses before computing new estimates. The one-session State and ten-session/intraday examples were selected for presentation after inspecting results from September 7.

The September 8 location plan fixed three future-input channels, two noise controls, two doses, three arms, twenty seeds and two principal endpoints. Noise calibration used clean 2016–2021 forecasts. Appendix D reports all forty pooled conditions and annual results. The pointwise HAC intervals describe within-market temporal variation. Earlier proposed simultaneous procedures did not meet calibration criteria and were excluded.

Account-state studies and original criteria. The pre-stress sensitivity design fixed the stress distribution, target, features, penalty, mature-response splits and comparisons before computing targets. Its 2019–2021 evaluation improved MAE for Ridge but not LightGBM, so the joint criterion requiring both models to improve was not met. The September 10 All/State factorial analysis reused daily paths; the historical/ all-cash control fixed selection pairs before computing cash-state responses.

The equal-amplitude audit was designed after examining development responses. Audit cases, representatives, monthly queues, budgets, the 1 bp/day tolerance, D22/S49 inputs, regression and uncertainty rules were fixed before features were joined to targets. Its original pooled State-minus-Decision interval includes zero. Appendix G retains the complete comparison.

Heterogeneous audit specification. On September 13, the study fixed next-session substitutions of ten-session return, intraday return and log volume under common All19 forecasters, three exposure fractions and ten shared masks. The six-session target, ordered-selection audit cases, whole-row representatives and monthly quotas were fixed for the mixed pool and fifteen restricted pools. S17, D28, State55, regression and uncertainty rules preceded response computation. The primary comparison was pooled D28-minus-S17 capture at 10%, requiring a positive paired 95% HAC20 interval; State-minus-Decision was secondary. Completed 2022 responses train 2023 fits, and eligible 2022–2023 responses train 2024 fits.

Earlier mechanism plan and the completed 2025 audit. The September 8 record documented no prior inspection of February–December 2025 prediction or portfolio performance and proposed four mechanism comparisons for that window. Those analyses were not conducted; the 2025 evidence reported here comes from the separate September 13 audit specification. January observations had already supported late-2024 labels and future features, and membership interventions had queried 2025 source data. The access boundary therefore concerns prediction and portfolio performance. A

separately funded January-start descriptive account was added before either account's new results; February-start remained primary.

The September 13 audit specification fixed the D28-minus-S17 contrast, candidates, forecasting and audit fits, secondary grid, 10% budget and positive-interval criterion before inspecting 2025 performance. Compatibility checks reproduced 2024 forecasts exactly. Pre-outcome calibration replaced the initial normal-critical-value HAC20 interval with four consecutive 46-date blocks and a t_3 critical value. A separate simulation seed met the unchanged coverage criteria; both calibration records were retained.

First study access was September 13 at 19:59 UTC. Audit coefficients were fixed before 2025 responses, and priorities were sealed before targets were generated or joined. The primary gain, +7.68 points with paired 95% interval [4.77, 10.58], meets the fixed criterion. Appendix F reports all secondary comparisons, the descriptive full-year account and verification coverage.

Additional comparisons after the primary test. The direct D28–Replacement comparison and rank-weighted evaluation were added after the original 2025 results. The new policy, candidate support and frozen-fit transfer rule were recorded before computing new-policy responses. Allocation was introduced after inspecting initial transfer results and is exploratory. A later cache optimization reproduced all original candidate orders and representative descriptors. Appendix N reports these additional analyses.

L Relationship to backtest evaluation and selective inspection

The contribution sits at the intersection of economic evaluation, data integrity and selective inspection. Tables 61 and 62 distinguish the experimental objects and the decisions each line of work supports.

Economic evidence and strategy selection. Financial prediction studies evaluate both forecast accuracy and the investment outcomes obtained from those forecasts (Gu et al., 2020). Data-snooping inference accounts for searching across candidate models or strategies (White, 2000), and backtest-overfitting analysis assesses the reliability of selection based on historical performance. Our unit is a specified intervention in an evaluated pipeline. The paired account continuation measures how that intervention changes net returns; the audit priority determines which cases to check first. These procedures support different decisions: selection safeguards assess the search process, while economic auditing locates consequential violations within the evidence produced by a backtest.

From leakage mechanisms to economic consequences. Unsupervised preprocessing can bias cross-validation estimates, with effects depending on the preprocessing and prediction pipeline (Moscovich & Rosset, 2022). In recommendation, failure to respect the global timeline can expose future interactions and change models' relative performance (Ji et al., 2023). These findings motivate treating the evaluation procedure as part of the scientific object. Our financial experiments then trace specified interventions through selection, allocation and a self-financing account. At matched exposure, their economic effects differ with ranking location and account state. This within-mechanism variation creates the opportunity for decision-aware prioritization.

Table 61: **The audit question follows the path from information access to decisions and returns.** Related studies address complementary links; this study asks which economic checks to perform first.

Study	Setting	Comparison	Outcome	Question
Roth (2026)	2,047 tabular benchmarks; 129-dataset boundary analysis	28 core paired experiments; temporal boundary and graded comparisons	ROC-area change; standardized paired effect	Leakage mechanism and dose linked to predictive response
Bhand–Joshi (2026)	30 synthetic stocks; French 49 industries	Eight mechanisms; rolling-normalization validation on industry portfolios	Hit rate, Sharpe, and alpha t-statistics	Financial-panel leakage linked to apparent strategy performance
Fonseca (2026)	Synthetic 100-stock universes; NASDAQ-100, 2008–2025	Three timing-bias classes; 14 average-true-range configurations with point-in-time membership	Compound annual growth, risk, and investor experience	Point-in-time conventions linked to strategy and investor outcomes
Yin et al. (2026)	180 S&P 500 stocks in 30 buckets; 15 strategy instances	Five matched engines; four benchmark-specific cost regimes	Cross-engine return/risk disagreement	Engine and cost assumptions linked to return and risk disagreement
This study	CSI 300/500 stocks; distinct selection interventions	Five protocol axes; learning entry, location, initial state; temporal audit	RankIC, selection, net-return effects; impact capture	Exposure and induced decisions linked to economic effects and audit value

Inspection objectives determine useful priorities. ActiveClean couples selective data cleaning with model updates (Krishnan et al., 2016). Active testing selects test labels for accurate risk estimation and corrects selection-induced bias (Kossen et al., 2021). Guided exploration uses oracle feedback to discover high-confidence prediction errors (Lakkaraju et al., 2017). Influence functions attribute predictions to training observations (Koh & Liang, 2017). Our audit instead measures finite interventions through paired portfolio continuations. Its objective rewards recovering large absolute economic effects from supplied cases, which makes induced investment decisions directly relevant to prioritization. The budget counts these counterfactual checks.

Decision information for evaluation. Decision-focused learning aligns prediction with downstream optimization, including discrete decisions (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022); Mandi et al. synthesize the gradient-based and gradient-free methods for this training objective (Mandi et al., 2024). Our forecasting models and portfolio rules define the system being audited. Replacement and Allocation summarize how an intervention changes that system's intended decisions; Decision learns audit impact from mature responses. Thus the optimization target is the auditor's allocation of checks. The frozen comparison tests the value of decision information for this target, while the rank-weighted extension explains why the diagnostic must track the actual allocation rule.

Table 62: **What an inspection is meant to achieve determines what it should prioritize.** Economic auditing allocates checks to recover consequential distortions.

Approach	Object	Target	Decision
ActiveClean (Krishnan et al., 2016)	Dirty training records	Model quality as records are cleaned	Which records to clean and use for model updates
Active testing (Kossen et al., 2021)	Test points to label	Accurate model-risk estimation	Which labels to acquire for evaluation
Guided exploration (Lakkaraju et al., 2017)	Confident predictions for oracle queries	Discovery of unknown errors	Where to search for model failures
Influence functions (Koh & Liang, 2017)	Training observations	Influence on a model prediction	Which training points explain a prediction
Decision-focused learning (Wilder et al., 2019; Mandi et al., 2024)	Predictor parameters	Downstream decision utility	How to train a predictor for its decision task
This study	Distinct supplied audit cases	Economic impact captured at a fixed checking budget	Which paired portfolio continuations to evaluate

M Scope of the evidence and operational use

The demonstrated result is that observing decisions improves the ordering of supplied economic checks. Replacement provides a model-free implementation for equal-weight selection; Allocation extends it to unequal intended weights. Decision learns useful priorities from historical checks and has a positive direct increment over Replacement in 2025. The original D28–S17 comparison remains the sole frozen primary test. Direct D28–Replacement inference and policy transfer are additional analyses, with the allocation comparator introduced after the initial policy-transfer results.

The evidence covers two constituent sets in one Chinese equity market, two forecasters, and equal- or rank-weighted top-decile portfolios with five-session holdings. Candidates are controlled interventions with known clean counterparts. The task prioritizes their economic checks; it does not discover unknown pipeline violations.

Checking-stage measurements in Appendix N quantify descriptor and continuation costs when score pairs and historical auditors are available. They exclude forecaster construction, historical label acquisition and human investigation. The initial uncached timing and the verified caching optimization are both reported. Prioritization provides an earlier partial impact inventory; exhaustive replay supplies the complete inventory. Its operational value depends on the available checking resources and the cost of producing candidate information.

N Direct baseline comparison and rank-weighted evaluation

N.1 Direct paired comparison with Replacement

This additional analysis uses the original audit cases, monthly quotas, priorities and economic responses. It was specified after the 2025 results had been examined and does not replace the original Decision–Score-only primary comparison. For each date, subtract Replacement's selected-impact numerator from Decision's and retain their common total-impact denominator. Applying Equation (23) to this signed numerator is algebraically identical to differencing the paired ratio influences. Development uses HAC20; 2025 uses the same four consecutive 46-date blocks and t_3 critical value as Equation (21). Intervals are pointwise and conditional on the fixed fitted models. Gain reports Decision minus Replacement Capture at the 10% budget, in pp.

Table 63: **Decision recovers additional impact beyond Replacement in 2025.** Paired post hoc Capture gains (pp) in development and the 2025 test.

Period	Stratum	Gain (pp)	CI
2023–2024	Pooled	+0.34	[-0.17, +0.84]
	CSI 300 / LightGBM	+0.13	[-0.63, +0.89]
	CSI 300 / Ridge	+0.21	[-0.57, +0.99]
	CSI 500 / LightGBM	+0.24	[-0.39, +0.88]
	CSI 500 / Ridge	+0.76	[+0.05, +1.47]
2025	Pooled	+1.19	[+0.54, +1.85]
	CSI 300 / LightGBM	+1.12	[+0.09, +2.15]
	CSI 300 / Ridge	+2.67	[+0.53, +4.81]
	CSI 500 / LightGBM	+0.31	[-0.62, +1.24]
	CSI 500 / Ridge	+1.59	[-0.80, +3.99]

Independent reconstruction reproduces the original Decision–Score-only point estimates and both interval endpoints to numerical precision. The new 2025 Decision–Replacement estimate is +1.19 points [0.54, 1.85]. The four within-block differences are +1.28, +1.28, +1.67 and +0.75 points in chronological order. These describe temporal consistency; the reported pooled estimate and interval use the pooled ratio and its influence, not an average of block-specific ratios.

In 2025, both priorities check 3,482 cases and share 2,567. Decision's 915 exclusive checks recover 6.48% of total impact and find 549 high-impact cases; Replacement's 915 exclusive checks recover 5.29% and find 406. The net improvement is 143 additional high-impact cases. Mean replacement is lower in Decision's exclusive set (3.07% versus 3.33%), while its mean economic impact is higher (1.39 versus 1.14 bp/day). This descriptive decomposition shows that the additional capture is not obtained simply by selecting more stock replacements. Development has 7,227 shared checks and 1,634 exclusive checks per priority; the exclusive sets find 1,289 and 1,218 high-impact cases, respectively. Its +0.34-point capture increment has interval [-0.17, 0.84].

N.2 A DIFFERENT MAPPING FROM RANKS TO PORTFOLIO ALLOCATIONS

The rank-weighted evaluation changes portfolio construction from equal weights to linearly decreasing rank weights. For rank r among K selected stocks, the intended sleeve weight is $w_r = 2(K - r + 1)/[K(K + 1)]$. The weights sum to one; the same five-session ledger allocates one fifth of pretrade NAV to each new sleeve. Within-sleeve allocations follow the ordered selected names. Next-open execution, 10 bp costs per filled side, internal renewal, blocked-trade retention and proportional cash scaling remain unchanged. Unlike equal weights, this policy can change intended asset allocations when selected names stay fixed but their order changes.

We evaluate all four original universe–forecaster strata, all 184 primary-window dates and all 90 candidate interventions per stratum–date. Clean forecasting models and intervention scores are reused. The new policy starts an independent account from cash on February 5, 2025; clean accounts continue through December 31. The candidate design and ordered-selection equivalence classes therefore remain the same, giving 34,984 audit cases and 3,482 checks at the 10% budget. Each economic response is recomputed under the new policy from a common new-policy account state, retaining the six-session absolute-return target.

Score-only and Decision reuse their 2022–2024 equal-weight coefficients and training scalers. Only four inputs require updating: past-return volatility and its products with score change, score scale and replacement. These use the new clean account's already observed gross returns. The other 24 D28 inputs are unchanged because they depend on the same score pairs and ordered selections. The five initial priorities and monthly ranks were saved before the new counterfactual targets were computed. This secondary rank-weighted evaluation uses the already examined 2025 period. Pointwise intervals use the four-block approximation and calibration from the original audit.

To connect overall recovery to the cases found, each transfer cell lists Capture (%), its CI, and Found, in that order. The paired contrast table reports first-named minus second-named Capture in pp. All use the same 10% quota. Allocation was added after the initial transfer results and remains exploratory; its contrasts share the same post hoc, pointwise interpretation.

Table 64: **Allocation changes reveal impact that membership alone misses.** The exploratory Allocation priority captures the most impact and finds the most high-impact cases under rank weights.

Capture (%) / [CI] / Found Stratum	Random ↑	Raw score ↑	Replacement ↑	Score-only ↑	Decision ↑	Allocation ↑
Pooled	9.86 [9.46, 10.26] 478	23.09 [18.18, 28.01] 1,415	23.32 [20.43, 26.20] 1,410	24.23 [19.97, 28.48] 1,498	26.36 [24.07, 28.65] 1,660	**31.99** [29.44, 34.54] **2,140**
CSI 300 / Ridge	9.83 [7.60, 12.07] 65	23.74 [21.09, 26.40] 257	23.58 [17.95, 29.22] 248	25.70 [23.41, 27.99] 280	26.50 [21.41, 31.59] 286	**34.31** [29.90, 38.72] **438**
CSI 300 / LGBM	9.63 [7.93, 11.32] 134	22.01 [16.54, 27.47] 336	22.16 [19.18, 25.14] 340	22.68 [18.39, 26.97] 342	26.04 [22.35, 29.73] 407	**33.44** [28.83, 38.04] **551**
CSI 500 / Ridge	10.15 [8.79, 11.50] 52	24.43 [21.62, 27.24] 223	22.89 [18.09, 27.69] 196	24.82 [22.77, 26.87] 235	25.37 [20.47, 30.27] 237	**32.71** [29.54, 35.89] **335**
CSI 500 / LGBM	9.92 [9.03, 10.80] 227	22.96 [14.62, 31.31] 599	24.15 [17.10, 31.20] 626	24.36 [16.64, 32.09] 641	26.93 [19.34, 34.52] 730	**29.74** [23.58, 35.89] **816**

Table 65: **The allocation-based advantage holds across all four strata.** Allocation outperforms Score-only, Replacement, and Decision in every stratum. Post hoc Capture contrasts (pp).

Stratum	Decision − Score-only	Decision − Replacement	Replacement − Score-only	Allocation − Score-only	Allocation − Replacement	Decision − Allocation
Pooled	+2.13 [0.15, 4.12]	+3.04 [1.23, 4.85]	-0.91 [-3.89, 2.07]	+7.77 [5.62, 9.92]	+8.68 [5.74, 11.61]	-5.64 [-6.84, -4.43]
CSI 300 / Ridge	+0.80 [-4.93, 6.52]	+2.91 [1.74, 4.09]	-2.12 [-8.18, 3.95]	+8.61 [3.95, 13.27]	+10.73 [4.81, 16.64]	-7.81 [-12.72, -2.90]
CSI 300 / LGBM	+3.36 [-0.86, 7.57]	+3.88 [2.16, 5.60]	-0.52 [-3.31, 2.27]	+10.76 [7.73, 13.79]	+11.28 [8.78, 13.78]	-7.40 [-9.84, -4.97]
CSI 500 / Ridge	+0.55 [-2.38, 3.48]	+2.48 [0.03, 4.93]	-1.93 [-5.35, 1.49]	+7.89 [5.39, 10.40]	+9.82 [6.89, 12.76]	-7.34 [-11.24, -3.45]
CSI 500 / LGBM	+2.57 [1.33, 3.80]	+2.78 [-0.09, 5.65]	-0.21 [-3.76, 3.33]	+5.38 [3.30, 7.45]	+5.59 [2.63, 8.55]	-2.81 [-4.31, -1.31]

N.3 Verification and computational measurements

The rank-weighted engine changes only the intended within-sleeve allocation in the original ledger. Hand-calculated three-stock weights verify the 3:2:1 allocation; a uniform-weight control reproduces every original daily field and portfolio snapshot over 30 synthetic sessions, including blocked trades, costs and renewal. On market data, all four original clean account paths are reproduced, as are 48 original counterfactual targets sampled across four time blocks and four strata. Every new continuation satisfies nonnegative cash, gross exposure at most one and the net-return accounting identity; the largest identity residual is below 1.2×10^{-15}. Unchanged ordered selections have exactly zero response.

Allocation ranks the total variation of the intended within-sleeve asset weights, $V_i = \frac{1}{2}\sum_j |w_j^i - w_j^0|$, with the same outcome-independent tie rule. Under equal weights this

equals Replacement. Under linear rank weights it is computed with an exact integer rank-mass numerator divided by $K(K+1)$, avoiding floating summation-order effects in tied priorities. This natural policy-specific comparator was added after the initial Decision rank-weighted results. It has no fitted parameters and no outcome-based tuning. Its 31.99% capture exceeds Decision's 26.36%: the paired Decision-minus-Allocation difference is -5.64 points [-6.84, -4.43]. Under this portfolio rule, observing intended weights improves audit priorities beyond selected-stock membership.

Membership-preserving candidates account for 18,460 checks and 26.08% of total rank-weighted impact, including 618 high-impact cases. Of these, 17,724 change the ordering of retained stocks; the other 736 have unchanged ordered selections and zero impact. At the 10% budget, neither frozen Decision nor Replacement selects a membership-preserving case. Allocation selects 360 membership-preserving checks and finds 130 high-impact cases, recovering 1.97% of total impact from this group. This decomposition identifies within-set weight changes as a source of Allocation's additional capture.

Timing uses the same CPU node with BLAS threads limited to one. The initial instrumented descriptor loop totals 212.17 s across four workers; it repeats clean-score processing and includes validation. Caching each date's invariant clean geometry and tie ordering reproduces all 66,240 source selected orders and every representative D28 feature exactly. Three serial passes, including cache construction and the allocation comparator, take 12.63, 12.71 and 12.72 s. These measurements start from available score-pair mappings and exclude input loading, forecasting, historical label acquisition and refitting.

For the rank-weighted accounts, summing per-case continuation timings gives 231.18 s for exhaustive checking, with unchanged selections reusing the clean continuation for every method. The selected Decision checks use 23.49 s; Allocation's selected checks use 23.52 s. Allocation's queue creation, sorting and quota selection takes 0.15 s (median of three serial passes). Computing all cached diagnostics, this ranking and the selected continuations gives 36.38 s of aggregate stage work, recovering 31.99% of total impact. This is a partial-inventory cost comparison, not the runtime of obtaining the complete inventory. Continuation times sum work measured within the four workers; diagnostic times are three serial passes on the same node, not a single end-to-end latency benchmark.

A practical cost model separates shared preparation from candidate checking. With available score pairs, let C_{diag} be all-candidate diagnostics, C_{fit} be incremental historical fitting costs, and c_i be each continuation cost. Prioritization costs $C_{\mathrm{diag}} + C_{\mathrm{fit}} + \sum_{i \in Q_g(b)} c_i$ beyond shared preparation; exhaustive checking costs $\sum_i c_i$. The crossover is whether diagnostic and fitting work is smaller than the omitted continuation work. Allocation has no historical fitting term. First-use learned auditing must additionally account for any historical responses that would not otherwise exist.

N.4 Structural interpretation of allocation changes

This explanatory calculation motivates the diagnostic through a simplified return model. It was added after the reported rank-weighted results; the experimental roles and rankings remain unchanged. Consider two fully invested sleeve allocations w^0, w^i with nonnegative entries summing to one, fixed at decision time t. Unselected assets have zero weight. Write $\Delta w = w^i - w^0$ and $V_i = \frac{1}{2}\|\Delta w\|_1$. Both branches allocate the same fraction α of common initial wealth to the sleeve, with identical remaining exposures. For one holding interval with fixed exposures, asset simple returns u and costs c^0, c^i expressed as fractions of that initial wealth, the paired net-return difference is

$$\Delta\rho = \alpha \Delta w^\top u - \Delta c, \qquad \Delta c = c^i - c^0. \tag{33}$$

The triangle inequality gives

$$|\Delta\rho| \le \alpha \sum_j |\Delta w_j|\,|u_j| + |\Delta c|. \tag{34}$$

Since α and Δw are fixed conditional on decision-time information $\mathcal{F}_t$, conditional expectation yields Equation (5). No independence of asset returns is required. If m_t bounds every conditional absolute return $\mathbb{E}[|u_j| \mid \mathcal{F}_t]$, then

$$\mathbb{E}_t|\Delta\rho| \leq 2\alpha m_t V_i + \mathbb{E}_t|\Delta c|. \tag{35}$$

This bound explains the role of allocation distance as a measure of capital exposed to different asset outcomes. It supplies structural motivation, rather than an ordering guarantee for realized economic impact: the directions and joint behavior of returns, execution and costs determine how much of the potential change materializes. The future-return terms in the calculation are not inputs to the implemented Allocation priority.

Relation to Replacement and trading turnover. With a common equal-weight selection size K, replacing m stocks removes mass m/K and adds the same mass elsewhere. Thus $V_i = m/K = R_i$ exactly. For rank-weighted portfolios, both membership changes and redistributions among retained stocks contribute to V_i. This is a distance between two counterfactual target allocations at the same decision time. Executed turnover instead depends on trades from pre-existing holdings, including internal renewal and blocked trades; V_i is neither that turnover nor its difference between branches.

Connection to the six-session target and learning. The paper's target is $A_i = (10^4/6)\sum_{h=1}^{6} |r^i_{t+h} - r^0_{t+h}|$ in bp/day. It averages absolute daily net-return differences and is not the absolute value of a single signed holding-period return difference. A fixed-exposure calculation can be applied interval by interval only with the corresponding exposures and cost normalization. In the actual account, each new sleeve targets one fifth of pretrade wealth; fills, retained positions, cash scaling and subsequent wealth can diverge across branches. Paired replay follows these states explicitly and supplies the reported A_i, without replacing it by the simplified bound.

Learning can use mature economic responses to rank which observable changes warrant checking. Return scales, joint exposures and execution conditions explain why allocation distance alone need not determine that ranking. The transfer experiment keeps Decision's equal-weight coefficients and existing descriptor definitions fixed. It measures how those priorities transfer to a new allocation rule; the separate Allocation priority measures the value of observing that rule's intended weights.