Proceedings of the 5th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2026.36535

Clone-Free Two-Model Order-Flow Weighting: A 30-Stock A-Share Case Study

Yiming Sun

School of Mathematics, Tianjin University, Tianjin, China
13130609760@163.com

Abstract. This paper evaluates a two-model order-flow weighting diagnostic and its use in trend and reversal rules. The design removes a structural defect in an earlier three-candidate specification: splitting one signed coefficient into constrained half-axis models creates a candidate that can collapse to the persistence null. The proposed order-flow response weighting (OFRW) implementation instead compares exactly two predictive models—a persistence null and one signed-association model—using prequential Student-t log scores over a common rolling window. Conditional on the incremental-model weight, a normal-CDF transform of the signed coefficient and its Newey–West standard error divides exposure between trend and reversal sleeves. This transform is an allocation heuristic, not a posterior probability. The archived case sample contains 30 A-shares selected from December 2021 inputs; 2025 is a retrospectively designated evaluation year, not a prospective holdout. The first-stage small-order-flow forecast improves on a historical mean but not on a rolling AR(1), and the signed-association model has slightly lower mean log scores than the null in both stock groups. After 0.10% one-way costs, OFRW earns 1.68% with a 0.743 Sharpe ratio. The two reported HAC contrasts against response-equal and swapped mappings do not reject zero, but they are not exposure matched. The result is a transparent negative case study: clone-free weighting repairs model identity but does not establish incremental trading value.

Keywords: order-flow imbalance, predictive model weighting, technical trading, A-share market, negative result

1. Introduction

Technical rules compress price and volume histories into implementable decisions, but their apparent performance is sensitive to specification search, transaction costs, and the evaluation window [1-3]. This paper asks a deliberately narrow financial-technology question: what predictive and portfolio behavior results when a clone-free signed order-flow diagnostic is attached to pre-specified trend and reversal rules? Because the implementations are not exposure matched, the exercise does not identify a pure mapping effect.

The motivating idea is recursive response. Cognitive-hierarchy and level-k models organize strategic behavior by how agents respond to others [4-7]. Daily market aggregates, however, do not reveal individual beliefs, payoff functions, action sets, or reasoning depth. Order-size buckets are not

DOI: 10.54254/2754-1169/2026.36535

investor identities. The empirical object studied here is therefore not cognition or a temporally ordered behavioral response. It is a same-day predictive association: the component of small-order imbalance forecastable before day t is used to predict large-order imbalance on day t.

This distinction matters because labels can conceal non-identification. An earlier construction compared persistence, a non-negative same-direction term, and a non-negative opposite-direction term. With a single signed regressor, one constrained model necessarily lies at a zero boundary and can become a clone of persistence. Normalizing three scores then gives duplicated labels separate weight. The present OFRW design removes that defect by comparing exactly two models: persistence and one unrestricted signed association. The incremental-model weight is subsequently divided between trend and reversal sleeves according to the estimated sign.

The contribution is methodological and diagnostic rather than a profitability claim. First, the two-model comparison is invariant to artificial half-axis duplication. Second, the implementation is prequential: scores are generated one step at a time, evidence windows contain only common realized forecasts, and portfolio targets are shifted to the next open, conditional on next-open availability of the daily fields. Third, the benchmark includes alternative mappings that separate incremental-model weight from its sign-to-rule allocation. The trading exercise is secondary and descriptive because these mappings are not exposure matched. The 2025 evidence is negative: OFRW has slightly worse first-stage and density estimates than its parsimonious benchmarks and a lower annual return than response-equal and swapped mappings.

OFRW sits at the intersection of three literatures but is not a substitute for any of them. Order-book research links order-flow imbalance to short-horizon price impact using event-level data [8], whereas the present daily bucket aggregates support only predictive association. Forecast-pooling research studies log-score combinations [9]; OFRW uses simpler finite-window exponential scores and makes no Bayesian claim. Equal-predictive-accuracy tests motivate the loss-difference diagnostics [10], while the trading exercise remains a separate, non-exposure-matched implementation check.

2. Data and method

2.1. Archived sample and information set

The panel contains forward-adjusted daily prices, volume, turnover, and Tushare money-flow buckets for 30 A-shares from 2021 through 2025, totaling 35,720 stock-days. Fifteen securities come from the CSI 300 and fifteen are high-turnover CSI 1000 securities. The list is reproduced exactly from three archived inputs dated no later than 31 December 2021: the two index constituent files and a daily-basic selection window. Small-stock turnover uses each candidate's latest available 20 rows through that date; all selected small stocks have 20 observations. File hashes, selection code, and the reproduced list are recorded in the run manifest. This makes the selection mechanically auditable, although a 30-stock case study cannot represent the full market or eliminate post-selection concerns.

Small-order imbalance (SOI) and combined large-plus-extra-large order imbalance (LIOI) are

$$SOI(i,t) = [Bsm(i,t) - Ssm(i,t)]/[Bsm(i,t) + Ssm(i,t)], \tag{1}$$

$$LIOI(i,t) = [BL(i,t) - SL(i,t)]/[BL(i,t) + SL(i,t)].$$

DOI: 10.54254/2754-1169/2026.36535

Here B and S denote buy and sell amounts; sm denotes small orders, while L combines large and extra-large orders. The daily market table is the left-join spine. Thirteen zero LIOI denominators are treated as missing; suspensions are not imputed. Data through 2022 warm up the rolling estimators, 2023–2024 are development periods, and 2025 is a designated historical evaluation period. The specification was finalized after 2025 data existed, so this is neither a prospective nor a preregistered holdout. The archived files record the extraction vintage, but not independent field-level publication timestamps. Next-open availability of all close-t inputs is consequently an explicit implementation assumption tested below with an additional one-session delay.

2.2. First-stage forecast

For each security, a rolling regression forecasts current SOI from information available by the preceding close. Predictors are lagged SOI, lagged return, lagged absolute return, and lagged abnormal turnover. Abnormal turnover is standardized on a rolling 20-valid-day window before being lagged. The estimation window contains at most 252 prior valid observations and requires at least 120:

$$SOI(i,t) = a + \rho SOI(i,t-1) + \gamma 1 r(i,t-1) + \gamma 2|r(i,t-1)| \tag{2}$$

$$+\gamma 3 ATurn(i,t-1) + u(i,t).$$

The fitted value ŜOI(i,t) is a generated regressor. Its estimation uncertainty is not propagated into the second stage. Accordingly, coefficient-sign diagnostics below are conditional, descriptive quantities rather than structural or fully calibrated inference [11].

2.3. Clone-free predictive weights

Using a rolling sample of at most 252 observations, OFRW compares two models for LIOI:

$$M0 : LIOI(i,t) = \alpha 0 + \phi 0 LIOI(i,t-1) + \varepsilon 0, t, \tag{3}$$

$$MR : LIOI(i,t) = \alpha R + \phi R LIOI(i,t-1) + \beta \widehat{SOI}(i,t) + \varepsilon R, t.$$

M0 is the persistence null. MR contains one unrestricted signed association coefficient β. Formally, projecting an unrestricted estimate b onto the two half-axes gives max(b,0) and −min(b,0); at least one equals zero, so its fitted conditional mean collapses to M0. The two-model construction removes this duplicated mean before score normalization. Both models use at most 252 valid prior observations, require 120, and produce one-step Student-t predictive densities with five degrees of freedom. For model k, residual variance is SSE/(n−pk), leverage is hk,t=xk,t′(Xk′Xk)+xk,t, and the Student-t scale is max{10−4, √[SSE/(n−pk)·(1+hk,t)]·√[(5−2)/5]}. Their log scores are accumulated over the latest 60 dates for which both forecasts and the realized outcome exist. With cumulative scores E0(i,t) and ER(i,t), the incremental-model weight is

$$p(k,i,t) = exp[Ek(i,t) - maxjEj(i,t)]/ \tag{4}$$

$$\Sigma jexp[Ej(i,t) - maxjEj(i,t)], k \in \{0, R\}.$$

DOI: 10.54254/2754-1169/2026.36535

Thus $q(i,t)=p(R,i,t)$ and $1-q(i,t)=p(0,i,t)$. These exponential weights summarize recent relative predictive scores; they are not Bayesian posterior probabilities.

To map the signed association into two trading directions, the rolling MR regression also yields β and a Newey–West standard error with five lags. Define $s(i,t)=\Phi[\beta/SEHAC(\beta)]$, where Φ is the standard normal cumulative distribution. The final weights are $wnull=1-q$, $wsame=q\cdot s$, and $wopposite=q\cdot(1-s)$. They are non-negative and sum exactly to one. At row t, q incorporates the score realized at close t, whereas the sign split uses a coefficient fitted through $t-1$; this is a conservative one-observation lag. The CDF transform is a bounded allocation heuristic. Because generated-regressor uncertainty is omitted and repeated rolling estimates are dependent, s must not be interpreted as a calibrated probability that β is positive.

2.4. Trading rules and falsification controls

Null weight remains in zero-return cash. The trend sleeve enters on an EMA(12)-minus-EMA(26) MACD crossing above its EMA(9) signal with 14-day ADX above 25, and exits on a downward MACD crossing. The reversal sleeve enters when 14-day RSI is below 30 and the close is below the 20-day lower Bollinger band (two population standard deviations), and exits when RSI reaches 50 or the close reaches the band midpoint. Signals observed at close t combine with weights updated using outcomes through close t. The resulting target is shifted one row and first becomes eligible at the next valid open. Nonpositive or missing opens are ineligible. If an order is blocked, the engine does not carry a stale instruction indefinitely: a later close may supersede it with a newer target.

The local engine assigns 1/30 of initial capital to each security without daily cross-stock rebalancing and uses fractional exposure with a standard weight-turnover cost approximation. Risky exposure drifts with open-to-open price movement; rebalancing back to the target creates turnover, and $c|q-p|$ is charged to current wealth for one-way rate c, pre-trade exposure p, and target q. This is not an exact share-and-cash brokerage ledger. Portfolio exposure, turnover, and cost are wealth weighted across securities. The baseline cost is 0.10% per traded side. Annual return is a 252-session CAGR; volatility and the zero-risk-free-rate Sharpe ratio annualize daily sample moments by $\sqrt{252}$.

Three controls interrogate different claims. "Response-equal" keeps response weight q, hence null weight $1-q$, but splits active exposure equally between trend and reversal. "Swapped" reverses the sign-to-rule mapping. "Equal-active" gives the two active sleeves equal fixed weights and holds the remainder in cash. Fixed trend and fixed reversal portfolios provide estimation-free rule references. These portfolios are not exposure matched; raw return differences therefore combine mapping, sleeve activation, exposure, and turnover. Newey–West inference with 20 lags describes daily return differences but cannot isolate a pure mapping effect [12].

The panel lacks official daily price-limit fields and order-queue or fill data. When explicit permissions are unavailable, the engine blocks only observable one-price locked sessions. This is weaker than exchange-level execution reconstruction and is retained as a limitation.

3. Results

3.1. Predictive diagnostics

Table 1 compares the first-stage OFRW forecast with a causally fitted rolling AR(1); the OOS R2 column separately uses a historical-mean benchmark. The enriched first stage improves on that mean but is slightly worse than AR(1). In 2025 their OOS R2 values are 0.2373 and 0.2397, while

DOI: 10.54254/2754-1169/2026.36535

mean absolute errors are 0.08671 and 0.08667. The extra price and turnover variables therefore provide no incremental first-stage advantage in this sample.

Table 1. First-stage SOI forecast diagnostics

Period	Model	Observations	MAE	Correlation	OOS R2 vs. mean
2023 development	OFRW first stage	7,259	0.09524	0.5022	0.3691
2023 development	Rolling AR(1)	7,259	0.09476	0.5077	0.3762
2024 development	OFRW first stage	7,249	0.08200	0.3060	0.3451
2024 development	Rolling AR(1)	7,249	0.08147	0.3195	0.3571
2025 test	OFRW first stage	7,287	0.08671	0.3301	0.2373
2025 test	Rolling AR(1)	7,287	0.08667	0.3338	0.2397

The second stage is equally cautious. In 2025 the mean MR-minus-M0 log-score difference is −0.00181 for large stocks and −0.00043 for small high-turnover stocks. Hence the signed-association density is slightly worse than persistence in both groups. Aggregating losses by date and applying a 20-lag HAC estimator, the pooled 2025 MR-minus-M0 log-score difference is −0.00112 (standard error 0.00118, t=−0.946, p=0.344; 95% interval [−0.00344, 0.00120]). The analogous OFRW-minus-AR(1) first-stage squared-error difference is 0.000048 (standard error 0.000062, p=0.437). These standard equal-predictive-accuracy summaries [10] do not reject zero; because the rolling models are nested, they are descriptive rather than definitive tests. The point estimates show no improvement, but non-rejection is not evidence of equivalence.

Across all score-available rows from 15 April 2022 through 31 December 2025, mean null weights are 0.5159 for large stocks and 0.5050 for small high-turnover stocks. The small-stock group has mean weights of 0.5050 null, 0.1100 same-direction, and 0.3850 opposite-direction. These full-sample model-allocation summaries are not 2025 estimates and do not identify investor types.

3.2. Portfolio test

Table 2 reports the three dynamic mapping strategies in 2025. At 0.10% one-way cost, OFRW earns 1.6775% with 2.2749% annualized volatility, a 0.7426 Sharpe ratio, and a −1.4416% maximum drawdown. Response-equal earns 1.9616% with a 0.8477 Sharpe ratio, while the swapped mapping earns 1.9580% with a 0.6311 Sharpe ratio. Outside the table, fixed reversal is the strongest technical rule, returning 4.7938% with a 0.8915 Sharpe ratio. The CSI 300 returns 17.1945%, but its 100% exposure makes it a market reference rather than a risk-matched comparator.

Table 2. Dynamic mapping results in the 2025 test period

Strategy	Annual return	Volatility	Sharpe	Max. drawdown	Annual turnover	Mean exposure
OFRW	1.6775%	2.2749%	0.7426	−1.4416%	4.3364	8.0315%
Response-equal	1.9616%	2.3233%	0.8477	−1.4605%	5.2978	11.7007%
Swapped mapping	1.9580%	3.1510%	0.6311	−1.9504%	7.3623	15.3198%

DOI: 10.54254/2754-1169/2026.36535

The annualized arithmetic daily-return difference between OFRW and response-equal is −0.2802 percentage points (HAC t=−0.2381, two-sided p=0.8118). The approximate 95% daily interval is [−0.01027%, 0.00804%]. Against the swapped mapping, the annualized difference is −0.2993 points (t=−0.1264, p=0.8994). Neither comparison rejects equality. More importantly, neither identifies a pure benefit from signed allocation because mean exposures and turnovers differ. An exposure-matched contrast was not implemented, so no causal mapping claim is made.

3.3. Sensitivity

The negative incremental conclusion survives the reported perturbations. Evidence windows of 40, 60, and 80 common events give OFRW returns of 1.6548%, 1.6775%, and 1.3540%; response-equal and swapped controls are higher in every case. Student-t degrees of freedom of 3, 5, and 10 produce OFRW returns of 1.6317%, 1.6775%, and 1.7071%. Sign-HAC lags of 0, 5, and 10 produce 1.6857%, 1.6775%, and 1.6705%. Raising round-trip cost from 0.10% to 0.20% and 0.30% reduces OFRW return from 1.8976% to 1.6775% and 1.4581%. Shifting the complete close-t target by one additional session gives a 1.6899% return, 0.754 Sharpe ratio, and 8.02% mean exposure. This delay check reduces, but cannot eliminate, concern about unarchived publication timestamps. These are specification checks, not a multiple-testing claim of robustness.

4. Discussion

OFRW solves one real design problem: an unrestricted signed-association model cannot be multiplied into redundant half-axis candidates that each receive a normalized score weight. Exact weight normalization, temporal invariance, synthetic null behavior, signal timing, fractional accounting, cost handling, and the sample-selection cutoff are covered by 120 automated tests. Archived raw inputs reproduce the selected universe exactly, and a run manifest hashes code, data-derived artifacts, configuration, and software.

That repair does not create predictive or economic value. The first stage fails to beat a rolling AR(1); MR loses slightly to persistence on mean test-year log score; and the dynamic portfolio trails simpler controls in annual return. The appropriate interpretation is therefore asymmetric: the experiment rejects an easy positive narrative, but it does not prove that same-day order-flow association is universally useless. With 30 selected stocks and one designated year, statistical power and external validity are limited.

Four qualifications are decisive. First, $\hat{S}OI$ is generated, and its uncertainty is not propagated to the MR sign statistic; the CDF split is only a conditional heuristic [11]. Second, raw portfolio comparisons are not exposure matched and cannot isolate the sign mapping from activity and turnover. Third, field publication times and official limit, queue, and fill fields are unavailable, so the trading exercise remains an implementation illustration even after the conservative extra-session delay. Fourth, daily order-size aggregates reveal neither account identity nor beliefs, payoffs, or strategic best responses. References to cognitive hierarchy motivate response ordering only; they provide no license to call the weights cognitive levels.

Future work should use repeated prospective windows, broader pre-specified universes, joint generated-regressor inference, official execution fields, an exposure-matched mapping test, and nonlinear or regime-switching density baselines.

Code, hashed derived outputs, 120 automated tests, and the reconstruction command will be provided to reviewers in an anonymized package upon request; raw provider inputs remain subject to their data license.

DOI: 10.54254/2754-1169/2026.36535

5. Conclusion

This paper presents a clone-free, two-model order-flow weighting scheme and subjects it to a negative trading test. The method is prequential, auditable, and explicit about the boundary between observable same-day predictive association and latent cognition. In the 2025 A-share case study, however, the enriched first stage does not beat rolling AR(1), the signed-association density does not beat persistence on average, and OFRW does not produce a higher annual return than response-equal, swapped, or fixed-reversal controls. Repairing model identity improves interpretability, not returns. The evidence supports retaining OFRW as a diagnostic construction, not promoting it as a validated trading signal.

References

[1] Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance, 47*(5), 1731–1764.
[2] Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance, 55*(4), 1705–1765.
[3] Bajgrowicz, P., & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics, 106*(3), 473–491.
[4] Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics, 119*(3), 861–898.
[5] Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review, 85*(5), 1313–1326.
[6] Crawford, V. P., Costa-Gomes, M. A., & Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature, 51*(1), 5–62.
[7] Zhou, H. (2022). Informed speculation with k-level reasoning. *Journal of Economic Theory, 200*, Article 105384.
[8] Cont, R., Kukanov, A., & Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics, 12*(1), 47–88.
[9] Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics, 164*(1), 130–141.
[10] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263.
[11] Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review, 25*(1), 221–247.
[12] Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703–708.