When to Apply Forecast Corrections: Opportunity, Recoverability, and Risk under Delayed Feedback

Anonymous authors
Paper under double-blind review

Abstract

Lightweight forecast corrections can improve a frozen forecaster on average, but heterogeneous effects make the correct action forecast-dependent, and the decisive loss difference is unavailable when a multi-step forecast is issued. We therefore study correction selection separately from correction generation through a matched delayed-feedback evaluation that freezes each candidate pair and distinguishes average correction capacity, retrospective opportunity, issue-time recovery, exposure, and local downside. In the primary Prompt-Z/PatchTST study, retrospective opportunity appears throughout the evaluated matrix, whereas the tested issue-time policies recover only a limited and heterogeneous share. The learned router improves over simple non-anticipating rules and the minimal linear router, but larger routers, regression-risk-aligned objectives, and simple mixtures do not yield consistent further gains under matched Validation/CAL control. Operating-point changes jointly alter recovery and intervention exposure, and positive aggregate recovery can coexist with adverse cells and local periods. A separately audited iTransformer scope check supports the same qualitative separation, while the conclusions remain bounded to the evaluated candidate pairs and open-loop replay.

1 Introduction

Lightweight correction modules can improve a frozen time-series forecaster without updating its full parameter set. An average gain, however, does not determine the correct action for a particular forecast: the corrected candidate may help one window or channel and harm another. In multi-step forecasting, that action must be chosen before its decisive evidence exists. The loss difference between the frozen and corrected forecasts becomes observable only after the complete target horizon matures. Selecting between two already-generated forecasts is therefore a delayed regression-deferral problem, distinct from the problem of generating the correction itself.

Prior work addresses the components of this problem from different directions. Delayed-feedback forecasting establishes when outcomes may enter an online state without future leakage (Joulani et al., 2013; Lau et al., 2025; Zhao & Shen, 2025; Huang et al., 2026). Test-time adaptation and lightweight correction show how a pretrained forecaster can be modified at the input, representation, parameter, or output level (Kim et al., 2025; Medeiros et al., 2025; Liang et al., 2026; Im & Kwon, 2026; Chen et al., 2024). Learning to defer and adaptive fusion treat instance-dependent choice as a separate prediction problem (Madras et al., 2018; Mao et al., 2024; Lee et al., 2025; Liu et al., 2025). In particular, regression deferral provides objectives aligned with supervised candidate risk. These strands leave a narrower question unanswered: when correction effects vary across issuances and the decisive loss difference arrives only after horizon maturity, how much of the retrospective advantage can a policy using issue-time information recover? Answering this question requires a matched evaluation that holds the candidate pair fixed and measures capacity, retrospective opportunity, issue-time recovery, and local downside on the same records.

We isolate this selection layer by freezing a forecaster, a correction generator, and the resulting Frozen/Fixed candidate pair before comparing decision policies. The evaluation separates the capacity of always applying the correction, the opportunity exposed by a retrospective Oracle, the fraction recovered from issue-time information, and the intervention risk that remains across cells and local time partitions. All policies share an issue–enqueue–mature–update replay and an immutable FIT–

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.

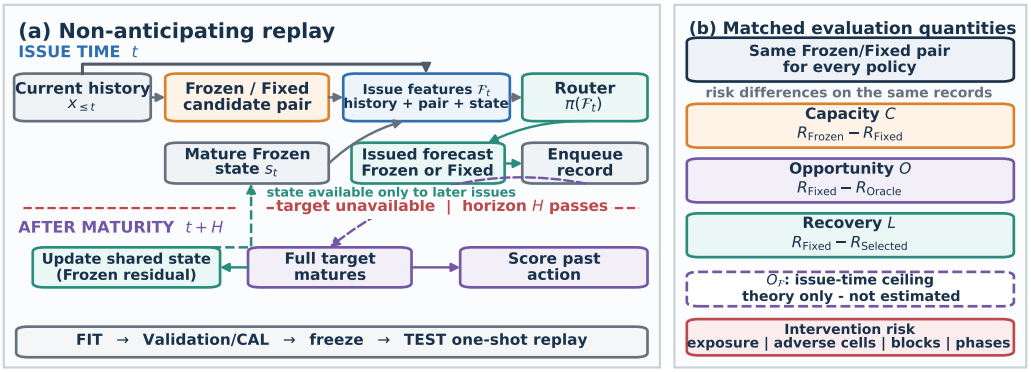


Figure 1: **Matched delayed-feedback evaluation.** At issue time, a frozen forecaster and a fixed correction generator produce the same candidate pair for every rule; routing uses only issue-time-available features. Outcomes enter the residual state after horizon maturity. The evaluation separates correction capacity, retrospective opportunity, issue-time recovery, and local intervention risk.

Validation/CAL–TEST chronology; objectives, thresholds, exposure controls, fallbacks, and mixture coefficients are frozen before TEST. The conditional mean advantage supplies the standard binary decision reference, while the unobserved information-limited opportunity keeps information restriction distinct from the realized performance of a fitted policy. Figure 1 summarizes this contract.

Across 84 matched PatchTST dataset–horizon–seed cells, retrospective opportunity appears throughout, whereas recovery by the evaluated issue-time policies is limited and heterogeneous. The selected policy exceeds the tested simple rules and minimal linear control, but larger routers, aligned objectives, and simple mixtures do not improve it consistently under matched calibration. The less restrictive Mao Natural operating point raises nominal recovery while producing more adverse cells, blocks, and phases. These controls separate policy capacity, objective alignment, operating control, and local stability rather than treating them as interchangeable explanations.

The paper makes three contributions. First, it frames the use of an already-generated correction as a matched regression-deferral problem under delayed feedback and provides an evaluation that keeps average correction capacity, retrospective opportunity, issue-time recoverability, and local intervention risk distinct. Second, the primary 84-cell study shows that widespread hindsight opportunity can coexist with limited and heterogeneous issue-time recovery: the selected policy adds value over the tested simple rules and minimal linear router, while the evaluated aligned objectives, mixtures, and larger routers provide no consistent further improvement. Third, the study shows that an operating point changes recovery jointly with intervention exposure and adverse local behavior, motivating reports that retain negative cells and periods rather than average accuracy alone. These claims remain tied to the Prompt-Z/PatchTST candidate pairs. A separately summarized 24-cell Prompt-Z/iTransformer matrix, admitted only after the candidate-reference equivalence checks passed, supports the qualitative distinction on a second backbone without pooling the two studies or claiming generator independence.

2 Related Work

Delayed feedback in online forecasting. Delayed feedback determines when an online learner may use an outcome; classical analyses characterize its effect on sequential learning rather than forecast-candidate deferral (Joulani et al., 2013). In time-series forecasting, DSOF emphasizes genuinely unknown future targets, while PROCEED and ADAPT-Z make horizon-dependent feedback availability explicit in their adaptation protocols (Lau et al., 2025; Zhao & Shen, 2025; Huang et al., 2026). These studies motivate our maturity discipline: a target may update state only after its complete horizon is observed. They do not by themselves quantify how much hindsight choice within a fixed forecast pair is recoverable when the forecasts are issued.

Frozen-forecaster correction and test-time adaptation. Forecasting systems adapt normalization, calibration, latent features, low-rank modules, or output-space components to accommodate distribution change (Kim et al., 2022; Fan et al., 2023; Liu et al., 2023; Dai et al., 2024; Kim et al., 2025; Medeiros et al., 2025; Liang et al., 2026; Im & Kwon, 2026). SOLID relates residual behavior to observed contexts, whereas ADAPT-Z adjusts features through persistent tracking (Chen et al., 2024; Huang et al., 2026). This line establishes that a useful corrected candidate can be generated in several ways. It leaves a different question when effects are heterogeneous: whether the corrected or frozen candidate should be used for the current issuance. We hold generation fixed to isolate that selection problem.

Regression deferral. Learning to defer treats expert selection as a prediction problem separate from the experts themselves (Madras et al., 2018). Mao et al. provide the closest formal reference: their two-stage regression-deferral surrogate aligns a routing score with the losses of pretrained regressors (Mao et al., 2024). We do not propose a new generic deferral objective. We instead test whether such population risk alignment translates into recoverable opportunity when the routing cost becomes observable only after a multi-step target matures, and when the operating threshold and exposure control must be frozen before TEST.

Forecast combination and issue-time context. Online ensembling and adaptive fusion show that one forecast need not dominate on every instance (Wen et al., 2023; Lee et al., 2025; Liu et al., 2025). Recent error-context modeling further motivates residual and base-prediction summaries as issue-time decision evidence, although ORCA is currently a preprint (Dai et al., 2026). Our fixed-pair protocol instead measures correction capacity, retrospective opportunity, issue-time recovery, exposure, and temporal downside on the same records under horizon-dependent maturity. The supplementary material provides a compact scope comparison; an unmarked entry is not evidence that a method categorically lacks the corresponding capability.

3 Problem Formulation

3.1 Issue time, maturity, and actions

Let B_θ be a forecaster trained before deployment and kept frozen thereafter. At issue time t, a history $x_{\le t}$ produces a frozen forecast $\hat{y}^{\mathrm{F}}_{t,c}$ for channel c and horizon H. A fixed correction generator G_ϕ produces a second candidate $\hat{y}^{\mathrm{Fix}}_{t,c}$. The routing action $a_{t,c} \in \{\mathrm{F}, \mathrm{Fix}\}$ returns the corresponding candidate. The complete target window $y_{t+1:t+H,c}$ matures only at $t+H$. We write $\mathcal{F}_t$ for the information available when the forecast is issued: the observed history, the two candidate forecasts, input summaries, and states derived exclusively from already-mature records. The current target, partial future targets, and any statistic computed from them are excluded from $\mathcal{F}_t$.

For a loss ℓ, define the realized correction advantage

$$A_{t,c} = \ell(\hat{y}^{\mathrm{F}}_{t,c}, y_{t+1:t+H,c}) - \ell(\hat{y}^{\mathrm{Fix}}_{t,c}, y_{t+1:t+H,c}). \tag{1}$$

Thus $A_{t,c} > 0$ means that applying the correction was beneficial. This sign convention is used throughout the theory, figures, and saved experiment records.

3.2 Capacity, opportunity, and recovery

Let R_{F} and R_{Fix} denote the aggregate risks of always choosing the frozen and corrected candidates. The correction capacity is

$$C = R_{\mathrm{F}} - R_{\mathrm{Fix}} = \mathbb{E}[A]. \tag{2}$$

A positive C says that always-on correction helps on average; it does not imply that the correction helps every action unit. A retrospective Oracle chooses the lower-loss candidate after the target is known. Its opportunity over always-Fixed is

$$O = R_{\mathrm{Fix}} - R_{\mathrm{O}} = \mathbb{E}[(-A)_+], \qquad (z)_+ = \max(z, 0). \tag{3}$$

For an issue-time policy $\pi(\mathcal{F}_t)$ with risk R_{L}, the recovered opportunity is

$$L = R_{\mathrm{Fix}} - R_{\mathrm{L}}, \qquad \eta = \frac{L}{O} \quad \text{when } O > 0. \tag{4}$$

Both L and η are signed. A negative value records a router that performs worse than always-Fixed; it is not clipped away. Because ratios are computed within each matched cell before aggregation, the reported mean L/O is not reconstructed by dividing two macro averages.

3.3 The issue-time optimal binary rule

Define the conditional mean advantage

$$m(\mathcal{F}_t) = \mathbb{E}[A_{t,c} \mid \mathcal{F}_t]. \tag{5}$$

Proposition 1 (Issue-time binary decision)**.** *For the fixed candidate pair, an additive loss, and no action-dependent switching cost, the minimum-risk $\mathcal{F}_t$-measurable action chooses Fixed when $m(\mathcal{F}_t) \geq 0$ and Frozen otherwise.*

The result follows by comparing the two conditional risks. Appendix A gives the short derivation. Proposition 1 is a decision reference, not a guarantee for a fitted router. It also does not say that the available feature vector is sufficient to estimate $m(\mathcal{F}_t)$ accurately.

3.4 Information-limited opportunity

The best opportunity available to any $\mathcal{F}_t$-measurable binary rule is

$$O_{\mathcal{F}} = \mathbb{E}[(-m(\mathcal{F}_t))_+]. \tag{6}$$

Conditional Jensen gives $O_{\mathcal{F}} \leq O$. Consequently,

$$O - L = (O - O_{\mathcal{F}}) + (O_{\mathcal{F}} - L). \tag{7}$$

The first term reflects opportunity hidden by the issue-time information set; the second includes estimation, model, optimization, calibration, and finite-sample effects. Our experiments observe O and L but do not estimate $O_{\mathcal{F}}$. The decomposition therefore prevents a stronger interpretation: a low empirical L/O is not, by itself, an impossibility or identifiability theorem.

3.5 Objective alignment and operating decisions

The sign rule makes objective alignment testable. Raw signed-advantage MSE targets the conditional mean, whereas Smooth-L1 targets a robust conditional location whose sign can differ under asymmetric advantage distributions. Magnitude-weighted binary classification provides another population-aligned route to the same boundary.

Proposition 2 (Magnitude-weighted classification)**.** *Let $Z = \mathbf{1}[A \geq 0]$ and weight binary cross-entropy by $|A|$. At the population optimum, thresholding its predicted probability at $1/2$ selects Fixed exactly when $m(\mathcal{F}_t) \geq 0$, provided $\mathbb{E}[|A| \mid \mathcal{F}_t] > 0$.*

Appendix A gives the algebra. Weighted BCE remains a finite-sample model rather than a guarantee. We also evaluate the two-stage surrogate of Mao et al. (2024), $\mathcal{L}_{\text{Mao}}(r) = \ell_{\text{Fix}}\,\text{softplus}(-r) + \ell_{\text{F}}\,\text{softplus}(r)$, where $r > 0$ selects Frozen. When both conditional candidate risks are positive, its population optimum is $r^*(\mathcal{F}_t) = \log\big(\mathbb{E}[\ell_{\text{Fix}} \mid \mathcal{F}_t]/\mathbb{E}[\ell_{\text{F}} \mid \mathcal{F}_t]\big)$, whose sign agrees with Proposition 1; zero-risk boundary cases are recorded in Appendix A. The natural threshold audits this population-level decision rule, while the main comparison freezes its operating point on CAL before TEST; neither population alignment nor calibration guarantees finite-sample recovery.

3.6 Local intervention risk

When a policy overrides Fixed with Frozen, it can avoid a harmful correction or discard a beneficial one. For override indicator $V_{t,c} = \mathbf{1}[a_{t,c} = \text{F}]$, define

$$H_{\text{avoid}} = \mathbb{E}[(-A_{t,c})_+ V_{t,c}], \qquad B_{\text{miss}} = \mathbb{E}[(A_{t,c})_+ V_{t,c}]. \tag{8}$$

Their difference is the policy's net gain over Fixed, $L = H_{\text{avoid}} - B_{\text{miss}}$. We report this aggregate together with correction coverage, override rate, win/tie/loss counts, the worst chronological block, and non-overlapping issuance-phase sensitivity. These are empirical exposure diagnostics. None constitutes a formal safety guarantee.

4 Matched Evaluation Framework

4.1 Holding the candidate pair fixed

The experiments separate correction generation from correction selection. PatchTST (Nie et al., 2023) supplies the frozen forecast. Prompt-Z, our bounded residual adapter, supplies the corrected candidate; its training-only checkpoint and candidate-generation contract are frozen before Validation/CAL and TEST. We do not compare generators. At a high level, the generator modifies a hidden state h_t by

$$\widetilde{h}_t = h_t + \gamma_t m_t \alpha_t \Delta h_t^{\text{raw}}, \qquad \|\alpha_t \Delta h_t^{\text{raw}}\|_2 \le \rho \|h_t\|_2, \tag{9}$$

where γ_t is a per-channel sigmoid confidence gate, m_t is the channel mask, and α_t scales the raw correction to satisfy the relative norm bound. Low-rank parameterization is used as an efficient implementation device (Hu et al., 2022); it is not the paper's main claim. Every router receives forecasts generated from the same saved backbone, correction checkpoint, and maturity-controlled replay.

4.2 Non-anticipating maturity replay

Algorithm 1 states the test-time contract. Loss is computed only after the complete target horizon arrives. Mature Frozen residuals may update the state used by later issuances, but no issued decision is revised. Sharing this policy-independent state yields an open-loop matched replay; closed-loop action-dependent trajectories lie outside the evaluation.

Algorithm 1: Non-anticipating replay with horizon maturity

1: With candidates and router architecture fixed, fit normalization and parameters on FIT records; select checkpoints without TEST.
2: Replay Validation three times, each from a reset state; carry only the third-pass residual state forward.
3: On the same Validation/CAL records, freeze the objective, requested target, threshold, fallback, and any mixture coefficient.
4: **for** TEST issue time $t = 1, 2, \ldots$ **do**
5: Pop only records whose complete H-step targets have matured.
6: Update the policy-independent residual state from those mature Frozen residuals.
7: Generate the shared $\hat{y}_t^{\text{F}}, \hat{y}_t^{\text{Fix}}$ candidate pair from current history.
8: Build $\mathcal{F}_t$ from history, the candidates, and the mature state; issue $\pi(\mathcal{F}_t)$.
9: Enqueue candidates and metadata; do not read the pending target or revise the issued action.
10: **end for**; matured TEST targets only score past actions and advance the future state.

Three invariants make comparisons meaningful. *Issue-time measurability* excludes current and future targets from $\mathcal{F}_t$; *maturity discipline* admits only complete horizons; and *split immutability* keeps training, objective choice, calibration, and fallback selection outside TEST. The appendix audits that perturbing unavailable labels cannot alter earlier actions.

4.3 Router fitting, Validation control, and one-shot TEST

Each learned router uses 76 issue-time-available features: a 64-dimensional drift state, five statistics from mature Frozen residuals, and seven summaries of the current candidate pair. No current target, partial future target, or calendar embedding enters this vector. The 5,145-parameter network is LayerNorm–64-unit GELU–scalar. Feature normalization and optimization use only the router-fit block; checkpoint selection uses a separate train-side block. Full optimizer settings appear in the appendix.

The raw-series Validation split controls the operating point; we call this use *CAL*, not a third data split. Before selection, the Validation prefix is replayed three times from reset, and only the final state is carried forward. A fixed objective–target grid is evaluated, candidates that violate the precision constraint are rejected, and the lowest-MSE accepted configuration is frozen. CAL maps the requested override target to a score threshold. TEST reuses the chosen objective, threshold, fallback, architecture, and any mixture coefficient in one pass without a cumulative override cap. Mature TEST labels may score past actions and advance the shared Frozen-residual state, but never supervise a pending action or revise a frozen choice.

4.4 Decision families

Selected Learned chooses on Validation among relative Smooth-L1, original-scale signed Smooth-L1, and magnitude-weighted BCE. CAL freezes an override threshold under requested targets in $\{0, 1, 2, 5, 10, 12, 15\}\%$ and a minimum harmful-override precision of 0.60; the policy otherwise defaults to Fixed.

The alternatives test distinct explanations. Global choice, per-channel history, and a matured-advantage exponential average probe simple non-anticipating rules. Raw signed-advantage MSE and the Mao surrogate in Section 3.5 probe objective alignment without changing candidates, features, or router capacity. Mao Natural uses the population-semantic boundary $r = 0$, whereas Mao CAL applies the common Validation control to the same fitted score. Global and per-channel convex mixtures, $\lambda\hat{y}^{\mathrm{Fix}} + (1 - \lambda)\hat{y}^{\mathrm{F}}$, probe action continuity with λ fitted before TEST. These families diagnose model complexity, loss mismatch, operating control, and binary action structure; they are not exhaustive.

4.5 Metrics and uncertainty

The primary loss is MSE, aggregated within each dataset–horizon–seed cell before the 84 cells are weighted equally. Alongside C, O, L, and L/O, we report wins, ties, losses, coverage, avoided harm, missed benefit, and local block or phase outcomes. Harmful-override precision is the fraction of Frozen overrides with negative realized advantage. Worst-block utility is the minimum over at most eight chronological TEST partitions, whereas issuance phase groups windows by index modulo H. Dataset-clustered and moving-block intervals respectively retain cross-dataset and short-range temporal variation; seed-specific summaries remain visible rather than serving as independent datasets. Further estimator details are in the appendix.

5 Experiments

5.1 Setup and evidence contract

The main matrix uses PatchTST on ECL, ETTh1, ETTh2, ETTm1, ETTm2, Traffic, and Weather, with horizons $H \in \{1, 12, 24, 48\}$ and seeds 2024, 2025, and 2026. This gives $7 \times 4 \times 3 = 84$ matched cells. Each raw series is split chronologically into 60% training, 10% validation, and 30% test rows, with look-back length 96. The Frozen/Fixed pair uses correction bound $\rho = 0.12$. The action unit is a window–channel pair; the reporting unit is a dataset–horizon–seed cell.

All headline comparisons in Table 1 reuse the same saved candidates and one-shot TEST replay; Oracle is retrospective, while Learned is Validation-frozen. Additional matched controls comprise an 84-cell exposure sweep, an 84-cell C0–C4 capacity matrix, and a separate reference-verified 24-cell iTransformer scope check admitted only after the candidate-reference equivalence checks passed. No operating point, architecture, or scope decision is selected on TEST. Full matrices, intervals, baselines, objective variants, mixtures, and phase records are reported in the supplementary material.

Table 1: Matched 84-cell TEST summary. Gains are MSE reductions; C, O, and L are reported in 10^{-3}.

Quantity / comparison	Mean	Median	W/T/L
Capacity C: Fixed over Frozen	2.774	1.673	67/0/17
Opportunity O: Oracle over Fixed	9.425	6.687	84/0/0
Recovery L: Learned over Fixed	1.469	0.279	79/0/5
Recovery ratio L/O (%)	11.06	4.90	–
Learned over Frozen	4.243	2.555	77/0/7

The dataset-clustered 95% interval for mean L is $[0.685, 2.349] \times 10^{-3}$. Ratios and relative gains are computed within cells before macro averaging. The relative-scale audit is in Table 6.

5.2 RQ1: How do average correction capacity and retrospective opportunity differ?

RQ1 compares two quantities on the same 84 candidate-pair cells. Always applying the correction improves 67 cells and degrades 17, even though its macro capacity is positive (Table 1); the aver-

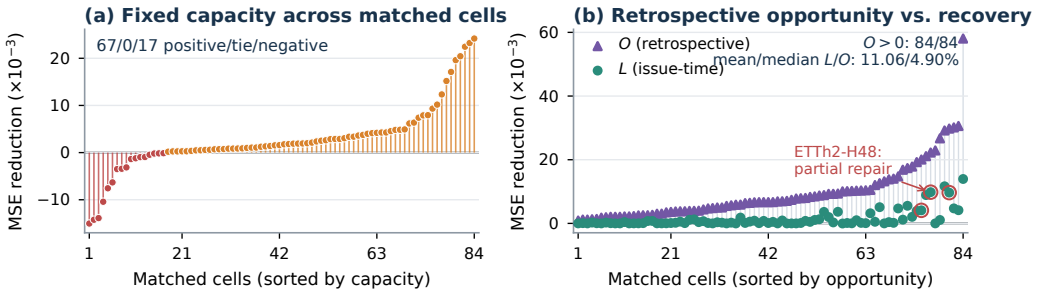


Figure 2: Matched capacity and opportunity across 84 cells. (a) Always-Fixed improves 67 cells but harms 17. (b) Retrospective Oracle opportunity is positive in every cell, while issue-time Learned recovery is smaller and can be negative. Circled ETTh2-H48 cells show partial mitigation without restoration to Frozen.

age gain of Fixed therefore shows that the generator can help, not that it should be used uniformly. The retrospective Oracle, by contrast, improves on Fixed in all 84 cells. Its mean opportunity is 9.425×10^{-3}, more than three times the mean capacity of Fixed over Frozen. Figure 2 thus locates the selection problem: harmful always-on corrections coexist with substantial within-pair hindsight headroom.

ETTh2-H48 is the clearest boundary case. Fixed is worse than Frozen in all three seeds. Learned recovers 20.43%, 32.46%, and 44.07% of the corresponding Oracle opportunity, yet remains worse than Frozen by $(6.321, 5.371, 4.427) \times 10^{-3}$ MSE. The selected operating points therefore provide partial mitigation. The prespecified exposure sweeps approach the Frozen baseline only as the policy approaches always-Frozen and do not identify whether the remaining opportunity gap arises from control, features, router class, or unavailable issue-time information. The Oracle remains retrospective and is not the information-limited quantity $O_{\mathcal{F}}$.

5.3 RQ2: How much opportunity do the evaluated issue-time policies recover?

Learned improves on Fixed in 79 cells and loses in five. Its mean within-cell recovery ratio is 11.06%, while the median is only 4.90%; the full range is -0.35% to 44.07%. Figure 3a shows the resulting long-tailed distribution rather than compressing it to a single average. The positive mean and clustered interval support nonzero aggregate recovery under this protocol, but the low median and five negative cells show that the available policy does not approach retrospective selection uniformly.

Three non-anticipating rules test whether the learned score merely reproduces a historical heuristic. Global Validation choice and per-channel history reach mean L/O of 2.23% and 1.60%, both with median $L = 0$; a matured-advantage exponential average reaches 4.68% but has negative mean worst-block utility. Learned beats them in 76, 78, and 61 cells, respectively (Figure 3b), supporting value over these rules rather than every possible heuristic.

Under matched CAL control, Raw-A MSE, magnitude-weighted BCE, and Mao exact reach mean L/O of 8.96%, 7.61%, and 9.07%; Mao's W/T/L against Selected is 22/0/62. Global and per-channel mixtures reach 3.59% and 4.84% (Figure 3c). None consistently improves on Selected. These controls narrow the explanation for the gap without ruling out other objectives or richer mixtures.

The capacity control changes only router width or depth. Linear C0 falls to 6.68% mean L/O, with a paired ΔL interval below zero relative to C1. C2–C4 reach 11.12–11.92%, but their paired intervals cross zero and C4 has seven negative cells versus five for C1. Increasing router capacity from 5,145 to 26,649 parameters therefore does not systematically close the gap; richer features or policy classes remain open.

All 24 regenerated iTransformer candidate pairs pass the reference-equivalence check. Opportunity is positive in 24/24 cells; Learned has W/T/L 21/1/2 and mean/median L/O 10.37%/3.72%. This separately summarized subset reproduces the qualitative separation but does not establish generator-independent behavior.

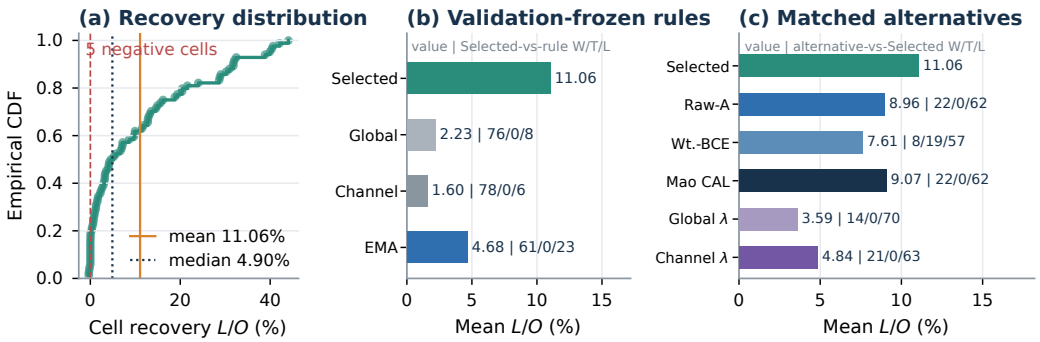


Figure 3: Recoverability diagnostics. (a) The cellwise ECDF retains five negative cells. (b) Selected Learned exceeds three Validation-frozen, non-anticipating rules; labels give Selected-versus-rule W/T/L. (c) Under matched CAL control, Raw-A MSE, magnitude-weighted BCE, Mao exact, and fit/calibration-only mixtures remain below Selected on mean L/O; labels give alternative-versus-Selected W/T/L.

Mao Natural and Mao CAL use the same fitted score but different thresholds and intervention rates. RQ3 therefore compares recovery, exposure, and local downside rather than ranking them by L/O alone.

5.4 RQ3: How do operating points trade recovery against intervention risk?

RQ3 evaluates requested Validation targets, realized TEST exposure, aggregate recovery, and adverse local partitions. Selected freezes cell-dependent targets: 14 cells choose zero percent and 32 choose 15%. These are Validation quantiles, not TEST caps. The frozen thresholds yield 9.92% mean TEST override; the zero-target cells average 0.136% override with L W/T/L 11/0/3. Selected retains 11.06% mean recovery, with five negative cells and 33 cells with a negative worst block. The worst is Traffic-H12, seed 2025, with $L = -1.751 \times 10^{-3}$.

For the same fitted Mao score, Natural and CAL make the control effect explicit. Natural reaches 12.58% mean L/O at 42.29% mean TEST override, whereas CAL reaches 9.07% at 7.83%. Their negative cell/block/phase counts are 21/57/46 and 16/41/33, respectively; Selected has 5/33/28. CAL therefore reduces exposure and adverse counts relative to Natural, but neither setting certifies safety or dominates Selected.

The fixed-score and online-prefix-cap sweeps are TEST-descriptive and cannot reselect Selected. The raw 25% target reaches 11.09% mean L/O at 27.00% exposure but has 18/49/40 negative cells/blocks/phases; at 50%, recovery falls to 7.26% and negative cells rise to 39. The hard-cap curve is likewise non-monotone. Broader intervention can therefore add local downside without monotonically recovering opportunity (Figure 4).

Across 1,785 dependent Selected phase rows, 55 cells are strictly positive, 28 contain a negative phase, and one contains zeros but no negative phase. Positive macro recovery can therefore coexist with adverse local periods.

6 Discussion, Limitations, and Conclusion

What the gap means. Capacity belongs to the candidate pair, opportunity marks hindsight action changes, recovery measures an issue-time policy, and local downside identifies adverse cells and periods. Positive capacity neither removes opportunity nor precludes adverse local outcomes; the Oracle–Learned gap is a selection problem conditional on the candidates.

What alignment and control explain. The controls narrow several explanations without identifying one cause. Learned improves on the tested simple rules, and C1 improves on C0, but larger routers, aligned objectives, and simple mixtures do not consistently improve on Selected. Mao Natural and the exposure sweeps show that operating control changes recovery, exposure, and local

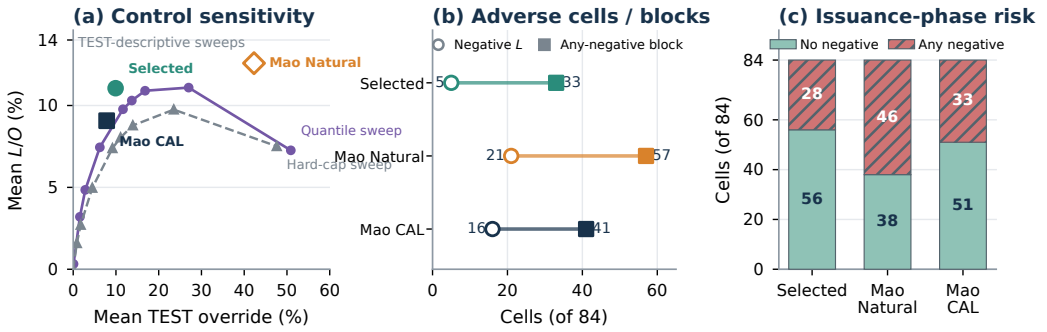


Figure 4: Operating control and local risk. (a) The prespecified quantile and online hard-cap sweeps through the 50% target are TEST-descriptive; Selected and Mao CAL are Validation-frozen, and Mao Natural is the $r = 0$ audit. (b) Counts of cells with negative aggregate recovery or at least one negative block; the worst Selected block is Traffic-H12, seed 2025, with $L = -1.751 \times 10^{-3}$. (c) Counts of cells with no negative phase versus at least one negative phase; phases are dependent views of the same 84 policies.

downside together. Because $O_{\mathcal{F}}$ is unobserved, information, representation, estimation, optimization, and calibration remain entangled in Equation (7).

Evaluation implication. Selective-correction studies should report C, O, L, L/O, exposure, and adverse cells and periods with choices frozen before TEST. Requested target, frozen threshold, realized TEST override, and online hard cap are distinct; when the corrected candidate is weak, as in ETTh2-H48, its deficits may be only partly mitigated.

Limitations. The primary study covers one Prompt-Z/PatchTST pair, seven datasets, four horizons, three seeds, and window–channel actions. The smaller, separately summarized 24-cell iTransformer matrix does not establish generator- or backbone-agnostic behavior. We neither estimate $O_{\mathcal{F}}$ nor sweep delay. C0–C4 spans 77–26,649 parameters under one feature schema, not richer representations. Open-loop matched replay omits action-dependent state trajectories, and the intervals and exposure diagnostics cover neither unseen shifts nor prospective safety.

Conclusion. For the evaluated fixed pairs, corrections that help on average still leave substantial forecast-level choice unresolved at issue time. Separating retrospective opportunity, realized recovery, intervention exposure, and local downside makes that unresolved choice measurable without treating a calibrated operating point as a deployment guarantee.

References

Mouxiang Chen, Lefei Shen, Han Fu, Zhuo Li, Jianling Sun, and Chenghao Liu. Calibration of time-series forecasting: Detecting and adapting context-driven distribution shift. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 341–352, 2024. doi: 10.1145/3637528.3671926.

Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Yuerong Xue, Shu-Tao Xia, and Zexuan Zhu. DDN: Dual-domain dynamic normalization for non-stationary time series forecasting. In *Advances in Neural Information Processing Systems*, 2024.

Xilin Dai, Yiding Liu, Hongjie Xia, Yifan Hu, Zewei Dong, Jiang-Ming Yang, and Qiang Xu. Learning the context of errors: Black-box online adaptation of time series foundation models. *arXiv preprint arXiv:2606.14222*, 2026. URL `https://arxiv.org/abs/2606.14222`.

Wei Fan, Pengyang Wang, Dongkun Wang, Dongjie Wang, Yuanchun Zhou, and Yanjie Fu. Dish-TS: A general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7522–7529, 2023.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Xiannan Huang, Shuhan Qiu, Jiayuan Du, and Chao Yang. Online time series prediction using feature adjustment. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=s4U2FWEMTU.

Jeonghwan Im and Hyuk-Yoon Kwon. COSA: Context-aware output-space adapter for test-time adaptation in time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=L7Z5wBMPrW.

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvari. Online learning under delayed feedback. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1453–1461, 2013. URL https://proceedings.mlr.press/v28/joulani13.html.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17868–17876, 2025. doi: 10.1609/aaai.v39i17.33965. URL https://ojs.aaai.org/index.php/AAAI/article/view/33965.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=cGDAkQo1C0p.

Ying-yee Ava Lau, Zhiwen Shao, and Dit-Yan Yeung. Fast and slow streams for online time series forecasting without information leakage. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=I0n3EyogMi.

Thomas L. Lee, William Toner, Rajkarn Singh, Artjom Joosen, and Martin Asenov. Lightweight online adaption for time series foundation model forecasts. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 2025. URL https://proceedings.mlr.press/v267/lee25ag.html.

Daojun Liang, Qi Li, Yinglong Wang, Jing Chen, Hu Zhang, Xiaoxiao Cui, Qizheng Wang, and Shuo Li. The forecast after the forecast: A post-processing shift in time series. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=syfWdclGE1.

Zhiding Liu, Mingyue Cheng, Zhi Li, Zhenya Huang, Qi Liu, Yanhu Xie, and Enhong Chen. Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective. In *Advances in Neural Information Processing Systems*, 2023.

Zhining Liu, Ze Yang, Xiao Lin, Ruizhong Qiu, Tianxin Wei, Yada Zhu, Hendrik Hamann, Jingrui He, and Hanghang Tong. Breaking silos: Adaptive model fusion unlocks better time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 40022–40042, 2025. URL https://proceedings.mlr.press/v267/liu25cm.html.

David Madras, Toni Pitassi, and Richard Zemel. Predict responsibly: Improving fairness and accuracy by learning to defer. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/09d37c08f7b129e96277388757530c72-Abstract.html.

Anqi Mao, Mehryar Mohri, and Yutao Zhong. Regression with multi-expert deferral. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34738–34759, 2024. URL https://proceedings.mlr.press/v235/mao24d.html.

Heitor R. Medeiros, Hossein Sharifi-Noghabi, Gabriel L. Oliveira, and Saghar Irandoust. Accurate parameter-efficient test-time adaptation for time series forecasting. *arXiv preprint arXiv:2506.23424*, 2025. URL https://arxiv.org/abs/2506.23424.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Qingsong Wen, Weiqi Chen, Liang Sun, Zhang Zhang, Liang Wang, Rong Jin, and Tieniu Tan. OneNet: Enhancing time series forecasting models under concept drift by online ensembling. In *Advances in Neural Information Processing Systems*, volume 36, pp. 69949–69980, 2023.

Lifan Zhao and Yanyan Shen. Proactive model adaptation against concept drift for online time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2020–2031, 2025. doi: 10.1145/3690624.3709210.

AI Use Statement

Generative AI tools assisted with literature organization and search, experimental-design feedback, code editing, figure preparation, numerical-result summarization, mathematical exposition, and manuscript drafting and language editing. They were not used to fabricate experimental observations or to replace execution of the reported experiments. AI-assisted code and numerical claims were checked against tests, saved records, and protocol manifests; citations were checked against source records. The authors reviewed all AI-assisted material and take responsibility for the final content.

Reproducibility Statement

Sections 3–5 specify the action unit, maturity rule, split order, policy selection, and reported quantities. Algorithm 1 gives the non-anticipating maturity-replay contract. Appendices C–**??** provide implementation details, non-anticipation audits, confidence intervals, complete cellwise matrices, scope exclusions, and the artifact/command index.

A Proofs and Decision Identities

A.1 Proof of Proposition 1

Condition on the issue-time information $\mathcal{F}_t$. The conditional risk difference between choosing Fixed and choosing Frozen is

$$\mathbb{E}[\ell(\hat{y}^{\mathrm{Fix}}, y) - \ell(\hat{y}^{\mathrm{F}}, y) \mid \mathcal{F}_t] = -\mathbb{E}[A \mid \mathcal{F}_t] \tag{10}$$

$$= -m(\mathcal{F}_t). \tag{11}$$

Fixed therefore has no larger conditional risk exactly when $m(\mathcal{F}_t) \geq 0$. Pointwise minimization of the conditional risk gives the stated measurable rule. Ties are assigned to Fixed by convention.

A.2 Information-limited opportunity

The opportunity recovered by the rule in Proposition 1, relative to always-Fixed, is

$$O_{\mathcal{F}} = \mathbb{E}[-m(\mathcal{F}_t)\mathbf{1}\{m(\mathcal{F}_t) < 0\}] = \mathbb{E}[(-m(\mathcal{F}_t))_+]. \tag{12}$$

Since $z \mapsto (-z)_+$ is convex, conditional Jensen yields

$$(-\mathbb{E}[A \mid \mathcal{F}_t])_+ \leq \mathbb{E}[(-A)_+ \mid \mathcal{F}_t]. \tag{13}$$

Taking expectations gives $O_{\mathcal{F}} \leq O$. The empirical study does not estimate $O_{\mathcal{F}}$; the quantity is retained to distinguish information restriction from finite-sample learning error.

A.3 Proof of Proposition 2

Let $Z = \mathbf{1}[A \geq 0]$ and $W = |A|$. At a fixed $\mathcal{F}_t$, the population minimizer of weighted binary cross-entropy is

$$p^*(\mathcal{F}_t) = \frac{\mathbb{E}[WZ \mid \mathcal{F}_t]}{\mathbb{E}[W \mid \mathcal{F}_t]}, \tag{14}$$

when the denominator is positive. Thresholding at one half selects Fixed if and only if

$$\mathbb{E}[WZ \mid \mathcal{F}_t] \geq \mathbb{E}[W(1-Z) \mid \mathcal{F}_t] \tag{15}$$

$$\iff \mathbb{E}[A\mathbf{1}\{A \geq 0\} \mid \mathcal{F}_t] \geq \mathbb{E}[-A\mathbf{1}\{A < 0\} \mid \mathcal{F}_t] \tag{16}$$

$$\iff m(\mathcal{F}_t) \geq 0. \tag{17}$$

The equivalence concerns the population objective. Clipped weights, finite samples, optimization, score calibration, and an exposure control can all move the fitted policy away from this reference.

A.4 Population optimum of the exact regression-deferral surrogate

Let

$$q_{\text{Fix}}(\mathcal{F}_t) = \mathbb{E}[\ell(\hat{y}^{\text{Fix}}, y) \mid \mathcal{F}_t], \qquad q_{\text{F}}(\mathcal{F}_t) = \mathbb{E}[\ell(\hat{y}^{\text{F}}, y) \mid \mathcal{F}_t]. \tag{18}$$

With the sign convention $r > 0$ selecting Frozen, the conditional Mao objective used in the matched objective audit is

$$\mathcal{J}(r \mid \mathcal{F}_t) = q_{\text{Fix}} \,\text{softplus}(-r) + q_{\text{F}} \,\text{softplus}(r). \tag{19}$$

For positive conditional risks,

$$\frac{\partial \mathcal{J}}{\partial r} = -q_{\text{Fix}}\sigma(-r) + q_{\text{F}}\sigma(r) = 0 \quad \Rightarrow \quad r^*(\mathcal{F}_t) = \log \frac{q_{\text{Fix}}(\mathcal{F}_t)}{q_{\text{F}}(\mathcal{F}_t)}. \tag{20}$$

Hence $r^* > 0$ exactly when $q_{\text{F}} < q_{\text{Fix}}$, so the zero-threshold action agrees with the Bayes selector. If one conditional risk is zero, the same result holds as a limiting score; if both vanish, either action is optimal. This is a population property of the regression-deferral surrogate of Mao et al. (2024), not a claim that finite-sample scores are calibrated or locally safe.

B Expanded Research Positioning

Table 2: Primary scope of neighboring research lines. ✓: central; ○: partial or setting-dependent; –: not the main evaluation target.

Research line	Delayed maturity	Changes forecast	Issue-time selection	Matched $C/O/L$/risk
Delayed online learning (Joulani et al., 2013)	✓	○	–	–
Online TS adaptation (Lau et al., 2025; Zhao & Shen, 2025; Huang et al., 2026)	✓	✓	○	–
Frozen correction (Kim et al., 2025; Medeiros et al., 2025; Liang et al., 2026; Im & Kwon, 2026)	○	✓	○	–
Supervised regression deferral (Madras et al., 2018; Mao et al., 2024)	–	–	✓	○
Forecast fusion / error context (Lee et al., 2025; Liu et al., 2025; Dai et al., 2026)	○	✓	✓	–
This work	✓	fixed pair	✓	✓

C Protocol and Implementation Detail

C.1 Chronological windows and maturity

Each series is divided into 60% training, 10% validation, and 30% test rows. Normalization statistics are fitted on training rows. A window is assigned to a split only when its complete target lies in that split; its look-back history may extend into the immediately preceding split. For horizon H, a record issued at t remains pending until every target in $y_{t+1:t+H}$ is available. Partial target windows do not update the residual state.

The residual state is policy independent. It is computed from the Frozen branch, so choosing Fixed or Frozen does not change the future feature stream. During TEST, matured outcomes serve only to score earlier decisions and to advance this Frozen-residual state for later issues. They do not update model parameters, recalibrate thresholds, or revise past actions. This is an open-loop matched evaluation: a closed-loop deployment may induce action-dependent states and is outside the present comparison.

C.2 Correction generator

Prompt-Z intervenes between the frozen encoder and prediction head. A rolling buffer of matured Frozen residuals supplies per-channel summaries of error magnitude, trend, dispersion, signed bias, and time since the last mature update. These summaries and a compact hidden-state summary form a 64-dimensional drift state. A rank-eight module proposes the hidden correction in Equation 9; γ_t is the per-channel sigmoid confidence gate, m_t is the channel mask, and α_t enforces the relative norm bound ρ.

For each main cell, the generator is fitted on a training-only tail block, with a training-only holdout used for checkpoint selection. Once selected, the backbone and generator remain frozen. The main matched matrix uses $\rho = 0.12$. Results from the separate historical $\rho = 0.10$ selector study are not pooled into the 84-cell headline.

C.3 Router fit and Validation control

Every learned router receives the same 76-dimensional issue-time feature vector. The 64-dimensional drift state is followed by five mature Frozen-residual statistics (absolute-error mean, slope, standard deviation, signed bias, and steps since the last mature update) and seven current-candidate summaries (Frozen mean, absolute mean, and standard deviation; correction mean, abso-

lute mean, RMS, and correction-to-Frozen RMS ratio). No current target, partial future target, or calendar embedding enters the vector.

The train-side chronology is partitioned into state warm-up, router fit, checkpoint selection, and a safety-audit block; the exact index ranges are stored in each split manifest. The headline router applies LayerNorm to the 76 features, then a 64-unit GELU hidden layer and a scalar output, for 5,145 trainable parameters. Feature normalization and router optimization use router-fit records only; checkpoint selection uses the separate train-side block. Training uses AdamW with learning rate 10^{-3}, weight decay 10^{-4}, batch size 512, at most 100 epochs, and patience 12. Relative Smooth-L1 predicts normalized relative advantage; absolute Smooth-L1 predicts signed advantage on the original MSE scale; weighted BCE predicts its sign with clipped magnitude weights. Raw-A MSE regresses unnormalized signed advantage with squared error. Mao exact instead minimizes Equation (19) using the two per-action forecast losses. These variants preserve the feature vector and candidate tensors.

CAL denotes the operating-control use of the raw-series Validation split; it is not an additional data partition. Validation/CAL evaluates requested override targets of 0, 1, 2, 5, 10, 12, and 15 percent, and each target becomes a fixed score-quantile threshold. TEST reuses the threshold without a cumulative cap, so realized TEST override can differ from the requested percentage. Nonzero candidates must reach harmful-override precision of at least 0.60; otherwise they fall back to always-Fixed. Harmful-override precision is the fraction of selected Frozen actions for which $A < 0$; it is undefined when there are no overrides. Among accepted candidates, the lowest Validation routed MSE is selected, with smaller targets and then a stable objective order breaking exact ties. All selection artifacts explicitly record that TEST was unread at FIT and control-selection time.

At a zero-percent Validation target, the threshold lies just below the minimum CAL score and produces no CAL override. A shifted TEST score can still cross it. Fourteen headline cells select this zero target; all realize a small nonzero TEST override, with mean 0.136% and recovery W/T/L 11/0/3. This does not alter the overall 79/0/5 Learned-versus-Fixed count.

For Mao exact, *Natural* applies the semantic boundary $r = 0$ directly, without a requested target, precision constraint, or fallback. *CAL* fits the same score but selects its threshold and control on Validation under the common rule above. Natural is therefore an objective-alignment audit, whereas CAL is the deployment-controlled comparison. Both are replayed on the same TEST candidate tensors.

C.4 Frozen contract and non-anticipation audits

The candidate contract freezes one candidate pair before policy action, a window–channel action unit, native checkpoint precision, and a maturity rule under which a target at TEST index t may affect candidates only from $t + H + 1$. Validation state uses three independent-reset replays. Each pass resets the residual tracker and pending queue, replays the same prefix, and carries only the third-pass state forward. Every policy then receives the same candidate tensors and the same policy-independent Frozen-residual state.

Table 3: Candidate-contract, non-anticipation, and lineage audit.

Check	Recorded result
Lineage and artifact hashes	passed
Chronological split manifest	passed
Candidate-policy invariance	passed
Future-unseen-label perturbation	passed
Validation warmup	3 independent-reset replays
Boundary-gap labels	not used

Runtime tensor-hash and perturbation audits cover three seed-2026, $H = 48$ representatives; all 84 cells retain checkpoint hashes. They confirm that unseen-label and router changes do not alter earlier artifacts or the candidate pair. The recorded environment is PyTorch 2.12.0.dev20260407, CUDA 12.8, and an NVIDIA RTX 5090; hardware affects runtime only.

Table 4: Macro recovery and dataset-clustered 95% intervals. W/T/L: Learned–Fixed (row 1), Learned–simple rule, or objective/mixture alternative–Learned.

Method	Mean L	Median L	Mean L/O	Dataset 95% CI	W/T/L
Selected Learned	1.469e-03	2.794e-04	11.06%	$[6.846e-04, 2.349e-03]$	79/0/5
Global choice	3.725e-04	0.000e+00	2.23%	$[0.000e+00, 9.596e-04]$	76/0/8
Per-channel history	2.511e-04	0.000e+00	1.60%	$[5.945e-05, 4.878e-04]$	78/0/6
Matured-advantage EMA	5.936e-04	4.096e-05	4.68%	$[2.640e-04, 1.026e-03]$	61/0/23
Relative Smooth-L1	1.031e-03	1.442e-04	8.04%	$[4.187e-04, 1.545e-03]$	9/19/56
Absolute Smooth-L1	1.332e-03	1.265e-04	9.74%	$[6.203e-04, 2.168e-03]$	7/47/30
Magnitude-weighted BCE	8.780e-04	7.864e-05	7.61%	$[4.160e-04, 1.321e-03]$	8/19/57
Absolute-MSE regression	1.254e-03	1.093e-04	8.96%	$[5.282e-04, 2.125e-03]$	22/0/62
Global mixture	8.919e-04	2.619e-11	3.59%	$[6.979e-05, 2.137e-03]$	14/0/70
Per-channel mixture	1.002e-03	1.126e-05	4.84%	$[1.292e-04, 2.136e-03]$	21/0/63

C.5 Matched objective-control audit

The matched objective-control study reuses the frozen PatchTST candidate pair in every one of the 84 main cells. All 84 FIT/Validation/CAL stages and all 84 one-shot TEST replays completed; every record has one TEST read, a passing candidate-identity check, and no protocol-failure flag. Its registry contains exactly 84 unique dataset–horizon–seed rows. Summary construction asserts this cardinality before producing any aggregate or table.

C.6 Additional control and scope audits

The exposure, capacity, and iTransformer scope checks use the machine-readable protocol tags `iclr2027_second_revision_budget_v1`, `iclr2027_second_revision_capacity_v1`, and `iclr2027_itransformer_matched_v2`. The budget and capacity FIT audits each pass 84 unique PatchTST cells with zero TEST reads during FIT/Validation/CAL, three independent-reset warmups, fixed feature schemas, and complete prespecified grids. Budget summaries are generated from 84 cell JSON files and distinguish raw Validation-quantile replay from the online prefix cap. Capacity summaries are generated from the same 84-cell candidate tensors for C0–C4; architecture selection on TEST is prohibited.

The reference-verified iTransformer audit first passes an eight-cell gate and then 24/24 prespecified reference and verification files. Every verification passes the frozen tolerance and every final cell records candidate-hash equivalence. Its results are summarized only within the iTransformer matrix and are never pooled with PatchTST.

C.7 Uncertainty and aggregation

Every macro statistic first reduces a TEST replay to one dataset–horizon–seed cell and then gives the 84 cells equal weight. Dataset-clustered percentile intervals resample seven whole datasets; with only seven clusters, they are descriptive rather than strong asymptotic inference. A second interval uses moving blocks of 48 forecast windows within each cell before macro aggregation. These intervals answer different questions and are not interchangeable. The seed summaries are descriptive because the three seeds for the same dataset and horizon share data.

D Matched 84-Cell Evidence

Table 4 collects macro recovery and dataset-clustered intervals. For selected Learned, the dataset-clustered 95% interval for mean L is $[6.846 \times 10^{-4}, 2.349 \times 10^{-3}]$; the moving-block interval is $[1.350 \times 10^{-3}, 1.585 \times 10^{-3}]$. The leave-one-dataset-out mean remains positive for every omitted dataset. These summaries support positive average recovery under the evaluated matrix, not independence among its 84 cells.

Table 5 is the complete 84-cell matched TEST matrix. It retains unfavorable cells and reports the Validation-frozen score and requested target beside the four TEST risks. Scores are Rel.-SL1, Abs.-SL1, or Wt.-BCE.

Table 5: Complete 84-cell matched TEST matrix. Score and requested target were selected on Validation; Oracle is retrospective. Coverage is the fraction using Fixed.

Dataset	H	Seed	Score	Val. target	Frozen	Fixed	Learned	Oracle	L/O	Cov.
ECL	1	2024	Rel.-SL1	2%	0.058234	0.057176	0.057155	0.054133	0.68%	98.6%
ECL	1	2025	Abs.-SL1	0%	0.055829	0.055574	0.055574	0.054366	0.00%	100.0%
ECL	1	2026	Rel.-SL1	2%	0.062786	0.060804	0.060771	0.057328	0.94%	98.1%
ECL	12	2024	Abs.-SL1	5%	0.120367	0.118827	0.118470	0.113831	7.16%	96.6%
ECL	12	2025	Wt.-BCE	2%	0.121968	0.119114	0.119024	0.115294	2.35%	97.3%
ECL	12	2026	Abs.-SL1	1%	0.121344	0.118458	0.118437	0.115940	0.87%	99.3%
ECL	24	2024	Abs.-SL1	2%	0.135136	0.134146	0.134096	0.132743	3.54%	98.3%
ECL	24	2025	Abs.-SL1	0%	0.135151	0.134081	0.134081	0.133182	0.00%	100.0%
ECL	24	2026	Abs.-SL1	10%	0.134114	0.133546	0.133430	0.129936	3.20%	93.6%
ECL	48	2024	Abs.-SL1	15%	0.154600	0.155969	0.155120	0.149335	12.80%	85.9%
ECL	48	2025	Abs.-SL1	1%	0.160718	0.160244	0.160119	0.155528	2.66%	99.3%
ECL	48	2026	Wt.-BCE	15%	0.157533	0.157257	0.156463	0.149263	9.93%	87.0%
ETTh1	1	2024	Wt.-BCE	5%	0.196075	0.178971	0.177896	0.152252	4.02%	95.9%
ETTh1	1	2025	Abs.-SL1	15%	0.744224	0.750539	0.736591	0.692477	24.02%	83.9%
ETTh1	1	2026	Abs.-SL1	0%	0.304375	0.280189	0.280189	0.257271	0.00%	100.0%
ETTh1	12	2024	Wt.-BCE	15%	0.513541	0.514479	0.509782	0.499719	31.82%	88.7%
ETTh1	12	2025	Rel.-SL1	15%	0.762821	0.762985	0.762214	0.760297	28.66%	85.1%
ETTh1	12	2026	Wt.-BCE	5%	0.500184	0.477731	0.473592	0.447134	13.53%	90.7%
ETTh1	24	2024	Abs.-SL1	5%	0.455598	0.435167	0.434990	0.421113	1.26%	93.6%
ETTh1	24	2025	Abs.-SL1	10%	0.416548	0.412557	0.412577	0.406675	-0.34%	89.9%
ETTh1	24	2026	Abs.-SL1	0%	0.476343	0.472971	0.472971	0.459758	0.00%	100.0%
ETTh1	48	2024	Abs.-SL1	15%	0.412055	0.412183	0.410877	0.403432	14.93%	89.0%
ETTh1	48	2025	Rel.-SL1	15%	0.451406	0.458972	0.454256	0.428822	15.64%	91.4%
ETTh1	48	2026	Abs.-SL1	1%	0.475246	0.462908	0.462878	0.452714	0.30%	99.5%
ETTh2	1	2024	Abs.-SL1	15%	0.133018	0.129968	0.129771	0.126211	5.24%	84.7%
ETTh2	1	2025	Wt.-BCE	2%	0.093157	0.085268	0.085194	0.078591	1.11%	97.5%
ETTh2	1	2026	Abs.-SL1	15%	0.089441	0.089636	0.089385	0.087693	12.93%	81.3%
ETTh2	12	2024	Abs.-SL1	15%	0.250592	0.247700	0.246438	0.243758	32.02%	86.0%
ETTh2	12	2025	Abs.-SL1	15%	0.178885	0.178282	0.177871	0.176092	18.78%	85.7%
ETTh2	12	2026	Rel.-SL1	10%	0.251236	0.250483	0.250403	0.249277	6.70%	86.8%
ETTh2	24	2024	Rel.-SL1	12%	0.263476	0.277364	0.265705	0.248175	39.94%	80.2%
ETTh2	24	2025	Abs.-SL1	15%	0.215551	0.213642	0.213162	0.206955	7.19%	85.4%
ETTh2	24	2026	Abs.-SL1	0%	0.212322	0.211080	0.211091	0.204004	-0.16%	100.0%
ETTh2	48	2024	Rel.-SL1	5%	0.248492	0.258912	0.254814	0.238849	20.43%	89.7%
ETTh2	48	2025	Abs.-SL1	15%	0.289458	0.304495	0.294829	0.274714	32.46%	81.7%
ETTh2	48	2026	Abs.-SL1	15%	0.246455	0.260691	0.250882	0.238434	44.07%	78.3%
ETTm1	1	2024	Abs.-SL1	2%	0.071101	0.060927	0.060924	0.056119	0.06%	97.3%
ETTm1	1	2025	Abs.-SL1	15%	0.213299	0.193689	0.189758	0.174437	20.42%	83.8%
ETTm1	1	2026	Abs.-SL1	0%	0.063130	0.058215	0.058215	0.050818	0.00%	100.0%
ETTm1	12	2024	Abs.-SL1	5%	0.230457	0.227938	0.227019	0.220986	13.22%	94.4%
ETTm1	12	2025	Abs.-SL1	15%	0.293282	0.293027	0.287499	0.275656	31.82%	79.6%
ETTm1	12	2026	Abs.-SL1	15%	0.294229	0.295190	0.286300	0.274070	42.09%	81.3%
ETTm1	24	2024	Abs.-SL1	0%	0.399597	0.398698	0.398698	0.390878	0.00%	100.0%
ETTm1	24	2025	Rel.-SL1	12%	0.720837	0.719021	0.718767	0.717712	19.37%	84.9%
ETTm1	24	2026	Wt.-BCE	12%	0.721284	0.716442	0.715733	0.712057	16.17%	84.2%
ETTm1	48	2024	Rel.-SL1	12%	0.398830	0.397964	0.395909	0.380112	11.51%	84.2%
ETTm1	48	2025	Abs.-SL1	1%	0.377166	0.373810	0.373682	0.363679	1.26%	98.2%
ETTm1	48	2026	Abs.-SL1	0%	0.736252	0.731995	0.731982	0.729696	0.59%	98.4%
ETTm2	1	2024	Wt.-BCE	10%	0.037223	0.035107	0.034879	0.030104	4.55%	89.1%
ETTm2	1	2025	Abs.-SL1	0%	0.035870	0.034556	0.034556	0.032353	0.00%	100.0%
ETTm2	1	2026	Abs.-SL1	0%	0.035375	0.033979	0.033979	0.032170	0.00%	100.0%
ETTm2	12	2024	Abs.-SL1	2%	0.092695	0.092418	0.092416	0.090338	0.07%	96.8%
ETTm2	12	2025	Abs.-SL1	15%	0.164056	0.163357	0.162609	0.160781	29.02%	81.3%
ETTm2	12	2026	Rel.-SL1	15%	0.114075	0.114534	0.113406	0.110615	28.80%	84.8%
ETTm2	24	2024	Abs.-SL1	2%	0.143118	0.141358	0.140962	0.137358	9.89%	95.2%
ETTm2	24	2025	Abs.-SL1	0%	0.174823	0.174457	0.174456	0.168966	0.02%	100.0%
ETTm2	24	2026	Wt.-BCE	2%	0.125051	0.124214	0.124131	0.121651	3.24%	96.9%
ETTm2	48	2024	Abs.-SL1	0%	0.168868	0.167284	0.167289	0.165202	-0.24%	100.0%
ETTm2	48	2025	Abs.-SL1	5%	0.165217	0.164500	0.164436	0.162423	3.06%	91.8%
ETTm2	48	2026	Rel.-SL1	1%	0.170632	0.168739	0.168727	0.165667	0.37%	99.1%
Traffic	1	2024	Wt.-BCE	15%	0.248366	0.242190	0.241716	0.233872	5.70%	84.4%
Traffic	1	2025	Abs.-SL1	15%	0.244673	0.240408	0.239231	0.223547	6.98%	83.3%
Traffic	1	2026	Wt.-BCE	15%	0.250722	0.243327	0.242823	0.231287	4.19%	83.8%
Traffic	12	2024	Abs.-SL1	15%	0.420892	0.411567	0.410571	0.401274	9.67%	83.0%
Traffic	12	2025	Wt.-BCE	15%	0.416938	0.410579	0.410615	0.400516	-0.35%	84.3%
Traffic	12	2026	Wt.-BCE	12%	0.422968	0.418845	0.418612	0.413165	4.11%	86.9%
Traffic	24	2024	Rel.-SL1	5%	0.417646	0.412835	0.412707	0.404923	1.61%	95.3%
Traffic	24	2025	Rel.-SL1	15%	0.415662	0.411912	0.411794	0.405226	1.77%	84.3%
Traffic	24	2026	Rel.-SL1	12%	0.424096	0.419542	0.419237	0.410258	3.29%	89.6%
Traffic	48	2024	Rel.-SL1	10%	0.442261	0.439647	0.439507	0.433273	2.20%	92.0%
Traffic	48	2025	Abs.-SL1	0%	0.501597	0.486416	0.486416	0.479492	-0.00%	100.0%
Traffic	48	2026	Wt.-BCE	10%	0.438367	0.436434	0.435645	0.429749	11.79%	85.5%
Weather	1	2024	Rel.-SL1	12%	0.206012	0.202399	0.201318	0.193915	12.74%	88.0%
Weather	1	2025	Rel.-SL1	12%	0.195340	0.191148	0.190745	0.177311	2.92%	91.7%
Weather	1	2026	Wt.-BCE	12%	0.185496	0.185005	0.184449	0.181196	14.60%	82.3%
Weather	12	2024	Abs.-SL1	15%	0.457746	0.456894	0.453283	0.448009	40.65%	78.3%
Weather	12	2025	Wt.-BCE	15%	0.400067	0.392136	0.391580	0.385674	8.60%	86.0%
Weather	12	2026	Abs.-SL1	0%	0.478412	0.455180	0.455105	0.444772	0.71%	99.8%
Weather	24	2024	Rel.-SL1	15%	0.434229	0.435430	0.433599	0.429521	30.99%	71.5%
Weather	24	2025	Rel.-SL1	1%	0.458270	0.458084	0.457925	0.447743	1.54%	98.2%
Weather	24	2026	Wt.-BCE	15%	0.437124	0.440261	0.438259	0.430988	21.60%	86.7%
Weather	48	2024	Abs.-SL1	15%	0.560139	0.557778	0.554598	0.547259	30.23%	71.8%
Weather	48	2025	Abs.-SL1	15%	0.560345	0.563769	0.558674	0.551139	40.34%	77.5%
Weather	48	2026	Wt.-BCE	15%	0.536846	0.540345	0.536697	0.530962	38.88%	61.8%

Absolute MSE reductions remain the headline because they correspond directly to the four matched risks. Table 6 addresses cross-dataset scale by normalizing within each cell before macro aggregation. The relative summaries preserve the same bounded interpretation and are not ratios of macro averages.

Table 6: Cell-first relative-scale audit for the 84 PatchTST cells. Every percentage is computed within a cell before macro aggregation.

Quantity (%)	Mean	Median	Minimum	Maximum
C/R_{F}	1.3706	0.7204	-5.7763	14.3084
O/R_{Fix}	3.8433	2.7064	0.1820	14.9292
L/R_{Fix}	0.4617	0.1213	-0.0087	4.2035
$(R_{\mathrm{F}} - R_{\mathrm{L}})/R_{\mathrm{F}}$	1.8362	0.9444	-2.5439	14.3122

E Simple Baselines, Objectives, and Mixtures

The three simple rules share the matched candidates and non-anticipating maturity stream with Learned. Global choice makes one Validation decision for the complete cell; per-channel history makes a Validation-frozen channel decision; the EMA uses matured advantage only. Table 7 reports the complete recovery matrix. A zero is retained when a Validation-selected rule reduces to always-Fixed.

Table 7: Complete simple-rule recovery matrix. Entries are within-cell L/O; every baseline and the Learned operating point were frozen before TEST.

Dataset	H	Seed	Learned	Global	Per-channel	Matured EMA
ECL	1	2024	0.68%	0.00%	0.00%	-0.29%
ECL	1	2025	0.00%	0.00%	0.00%	0.16%
ECL	1	2026	0.94%	0.00%	0.00%	0.06%
ECL	12	2024	7.16%	0.00%	0.00%	-0.02%
ECL	12	2025	2.35%	0.00%	0.00%	0.06%
ECL	12	2026	0.87%	0.00%	0.00%	-0.14%
ECL	24	2024	3.54%	0.00%	0.00%	0.14%
ECL	24	2025	0.00%	0.00%	0.00%	-0.51%
ECL	24	2026	3.20%	0.00%	0.00%	-0.01%
ECL	48	2024	12.80%	0.00%	0.00%	0.02%
ECL	48	2025	2.66%	0.00%	0.00%	0.01%
ECL	48	2026	9.93%	0.00%	0.00%	-0.16%
ETTh1	1	2024	4.02%	0.00%	0.00%	-0.15%
ETTh1	1	2025	24.02%	0.00%	8.60%	16.88%
ETTh1	1	2026	0.00%	0.00%	0.00%	1.54%
ETTh1	12	2024	31.82%	0.00%	0.00%	17.29%
ETTh1	12	2025	28.66%	0.00%	0.00%	7.99%
ETTh1	12	2026	13.53%	0.00%	0.00%	-0.08%
ETTh1	24	2024	1.26%	0.00%	0.00%	-3.25%
ETTh1	24	2025	-0.34%	0.00%	0.00%	-6.82%
ETTh1	24	2026	0.00%	0.00%	0.00%	-0.24%
ETTh1	48	2024	14.93%	0.00%	0.00%	-9.05%
ETTh1	48	2025	15.64%	0.00%	0.00%	-5.74%
ETTh1	48	2026	0.30%	0.00%	0.00%	0.00%
ETTh2	1	2024	5.24%	0.00%	0.00%	6.96%
ETTh2	1	2025	1.11%	0.00%	0.00%	0.03%
ETTh2	1	2026	12.93%	0.00%	9.11%	13.31%
ETTh2	12	2024	32.02%	0.00%	0.00%	35.17%
ETTh2	12	2025	18.78%	0.00%	0.00%	15.82%
ETTh2	12	2026	6.70%	0.00%	0.00%	-1.23%
ETTh2	24	2024	39.94%	0.00%	0.00%	51.09%
ETTh2	24	2025	7.19%	0.00%	0.00%	1.22%
ETTh2	24	2026	-0.16%	0.00%	0.00%	-1.21%
ETTh2	48	2024	20.43%	51.93%	27.78%	2.80%
ETTh2	48	2025	32.46%	0.00%	0.00%	6.37%
ETTh2	48	2026	44.07%	63.96%	13.59%	3.94%
ETTm1	1	2024	0.06%	0.00%	0.00%	-0.10%
ETTm1	1	2025	20.42%	0.00%	0.00%	27.84%
ETTm1	1	2026	0.00%	0.00%	0.00%	1.71%
ETTm1	12	2024	13.22%	0.00%	0.00%	-0.00%
ETTm1	12	2025	31.82%	0.00%	0.00%	9.87%
ETTm1	12	2026	42.09%	0.00%	0.00%	2.18%
ETTm1	24	2024	0.00%	0.00%	0.00%	1.31%
ETTm1	24	2025	19.37%	0.00%	0.00%	-0.47%
ETTm1	24	2026	16.17%	0.00%	0.00%	-2.63%
ETTm1	48	2024	11.51%	0.00%	0.00%	0.00%
ETTm1	48	2025	1.26%	0.00%	0.00%	-0.00%
ETTm1	48	2026	0.59%	0.00%	0.00%	-0.03%
ETTm2	1	2024	4.55%	0.00%	0.00%	3.73%
ETTm2	1	2025	0.00%	0.00%	0.00%	0.83%
ETTm2	1	2026	0.00%	0.00%	0.00%	0.25%
ETTm2	12	2024	0.07%	0.00%	0.00%	4.08%
ETTm2	12	2025	29.02%	0.00%	0.00%	29.03%
ETTm2	12	2026	28.80%	0.00%	0.00%	19.95%
ETTm2	24	2024	9.89%	0.00%	0.00%	16.19%
ETTm2	24	2025	0.02%	0.00%	0.00%	9.09%
ETTm2	24	2026	3.24%	0.00%	0.00%	7.13%
ETTm2	48	2024	-0.24%	0.00%	0.00%	3.60%
ETTm2	48	2025	3.06%	0.00%	0.00%	3.07%
ETTm2	48	2026	0.37%	0.00%	0.00%	-0.68%
Traffic	1	2024	5.70%	0.00%	0.00%	-0.31%
Traffic	1	2025	6.98%	0.00%	0.00%	0.99%
Traffic	1	2026	4.19%	0.00%	0.00%	1.38%
Traffic	12	2024	9.67%	0.00%	0.00%	-0.67%
Traffic	12	2025	-0.35%	0.00%	0.00%	1.71%
Traffic	12	2026	4.11%	0.00%	0.00%	1.92%
Traffic	24	2024	1.61%	0.00%	0.00%	5.47%
Traffic	24	2025	1.77%	0.00%	0.00%	6.84%
Traffic	24	2026	3.29%	0.00%	0.00%	5.39%
Traffic	48	2024	2.20%	0.00%	0.00%	18.01%
Traffic	48	2025	-0.00%	0.00%	0.00%	-0.82%
Traffic	48	2026	11.79%	0.00%	0.00%	1.73%

Continued on next page

Table 7 (continued)

Dataset	H	Seed	Learned	Global	Per-channel	Matured EMA
Weather	1	2024	12.74%	0.00%	5.50%	-0.08%
Weather	1	2025	2.92%	0.00%	0.00%	-0.07%
Weather	1	2026	14.60%	0.00%	6.51%	0.39%
Weather	12	2024	40.65%	0.00%	6.46%	0.95%
Weather	12	2025	8.60%	0.00%	0.00%	0.00%
Weather	12	2026	0.71%	0.00%	0.00%	-0.81%
Weather	24	2024	30.99%	0.00%	0.00%	14.48%
Weather	24	2025	1.54%	0.00%	0.00%	0.03%
Weather	24	2026	21.60%	33.83%	13.88%	16.53%
Weather	48	2024	30.23%	0.00%	0.00%	0.63%
Weather	48	2025	40.34%	0.00%	23.21%	14.72%
Weather	48	2026	38.88%	37.29%	19.37%	17.24%

The objective-and-mixture comparison changes either the score objective or the action family while preserving the feature vector, hidden capacity, chronological split, and TEST prohibition. The global and per-channel mixtures estimate their coefficients before TEST. Table 8 reports all seven methods for all 84 cells. It is a diagnostic matrix rather than a post-TEST model-selection menu.

Table 8: Complete objective-and-mixture matrix. Entries are within-cell L/O. Sel. is the Validation-selected Learned policy.

Dataset	H	Seed	Sel.	Rel.-SL1	Abs.-SL1	Wt.-BCE	Abs.-MSE	λ_g	λ_c
ECL	1	2024	0.68%	0.68%	0.00%	0.00%	0.00%	0.00%	-0.76%
ECL	1	2025	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-0.26%
ECL	1	2026	0.94%	0.94%	-0.00%	0.00%	0.01%	0.00%	-0.01%
ECL	12	2024	7.16%	8.59%	7.16%	2.83%	6.18%	4.26%	8.31%
ECL	12	2025	2.35%	1.61%	0.90%	2.35%	2.30%	0.00%	2.01%
ECL	12	2026	0.87%	0.00%	0.87%	0.00%	-0.00%	0.00%	0.14%
ECL	24	2024	3.54%	1.99%	3.54%	3.72%	3.10%	0.00%	0.35%
ECL	24	2025	0.00%	0.00%	0.00%	0.00%	0.02%	-0.00%	0.61%
ECL	24	2026	3.20%	2.23%	3.20%	0.87%	4.49%	5.74%	6.27%
ECL	48	2024	12.80%	8.96%	12.80%	9.32%	12.75%	23.96%	27.01%
ECL	48	2025	2.66%	1.69%	2.66%	1.80%	4.50%	9.21%	11.57%
ECL	48	2026	9.93%	8.98%	6.49%	9.93%	6.69%	13.92%	18.69%
ETTh1	1	2024	4.02%	3.87%	1.31%	4.02%	0.04%	-0.01%	-8.27%
ETTh1	1	2025	24.02%	15.08%	24.02%	9.16%	20.82%	10.88%	10.27%
ETTh1	1	2026	0.00%	0.00%	0.00%	-0.04%	0.13%	0.00%	0.00%
ETTh1	12	2024	31.82%	31.92%	40.94%	31.82%	34.13%	-0.00%	14.94%
ETTh1	12	2025	28.66%	28.66%	25.63%	20.52%	29.35%	0.00%	2.45%
ETTh1	12	2026	13.53%	6.34%	10.95%	13.53%	10.97%	0.00%	0.00%
ETTh1	24	2024	1.26%	-1.11%	1.26%	0.48%	0.23%	-0.42%	-7.13%
ETTh1	24	2025	-0.34%	0.15%	-0.34%	-1.39%	-0.39%	-11.76%	0.98%
ETTh1	24	2026	0.00%	0.00%	0.00%	0.00%	-0.00%	-0.00%	-0.82%
ETTh1	48	2024	14.93%	-0.01%	14.93%	0.28%	0.24%	5.81%	5.47%
ETTh1	48	2025	15.64%	15.64%	0.43%	0.04%	-0.00%	17.32%	14.84%
ETTh1	48	2026	0.30%	0.01%	0.30%	-0.33%	-0.21%	0.00%	-3.82%
ETTh2	1	2024	5.24%	3.38%	5.24%	1.36%	4.41%	0.00%	0.00%
ETTh2	1	2025	1.11%	0.09%	0.26%	1.11%	0.75%	-0.00%	-14.54%
ETTh2	1	2026	12.93%	6.49%	12.93%	6.80%	9.84%	7.48%	15.19%
ETTh2	12	2024	32.02%	21.42%	32.02%	28.94%	11.48%	-0.00%	-2.02%
ETTh2	12	2025	18.78%	15.03%	18.78%	13.31%	5.40%	-27.55%	-20.35%
ETTh2	12	2026	6.70%	6.70%	7.35%	5.61%	-0.02%	-0.00%	-0.02%
ETTh2	24	2024	39.94%	39.94%	39.58%	19.20%	44.01%	47.58%	47.58%
ETTh2	24	2025	7.19%	5.62%	7.19%	0.76%	0.06%	-0.00%	-2.57%
ETTh2	24	2026	-0.16%	-4.04%	-0.16%	0.00%	-1.64%	-12.86%	-9.48%
ETTh2	48	2024	20.43%	20.43%	20.77%	4.59%	31.98%	51.93%	51.93%
ETTh2	48	2025	32.46%	8.58%	32.46%	14.14%	32.38%	50.49%	50.49%
ETTh2	48	2026	44.07%	9.54%	44.07%	0.14%	41.67%	63.96%	63.96%
ETTm1	1	2024	0.06%	-0.00%	0.06%	-0.12%	0.06%	-0.00%	-0.00%
ETTm1	1	2025	20.42%	15.19%	20.42%	20.30%	9.85%	-0.00%	-29.04%
ETTm1	1	2026	0.00%	0.00%	0.00%	-0.00%	-0.00%	1.15%	1.51%
ETTm1	12	2024	13.22%	7.41%	13.22%	0.03%	0.61%	-4.90%	-4.83%
ETTm1	12	2025	31.82%	28.76%	31.82%	31.85%	31.55%	4.76%	19.91%
ETTm1	12	2026	42.09%	33.68%	42.09%	32.92%	39.02%	6.04%	10.97%
ETTm1	24	2024	0.00%	0.00%	0.00%	0.00%	0.00%	-0.00%	-0.75%
ETTm1	24	2025	19.37%	19.37%	-0.09%	-0.36%	0.33%	-0.00%	-5.00%
ETTm1	24	2026	16.17%	6.75%	10.53%	16.17%	14.72%	-0.00%	-1.75%
ETTm1	48	2024	11.51%	11.51%	0.00%	1.58%	1.49%	-4.85%	2.75%
ETTm1	48	2025	1.26%	-0.07%	1.26%	-0.02%	0.04%	-27.47%	3.62%
ETTm1	48	2026	0.59%	0.04%	0.59%	0.93%	3.95%	0.00%	-22.59%
ETTm2	1	2024	4.55%	4.22%	4.74%	4.55%	5.01%	1.73%	2.74%
ETTm2	1	2025	0.00%	-0.01%	0.00%	-0.00%	-0.00%	-0.00%	-0.00%
ETTm2	1	2026	0.00%	0.00%	0.00%	0.00%	0.00%	-0.00%	0.47%
ETTm2	12	2024	0.07%	0.67%	0.07%	-0.01%	0.01%	0.00%	2.14%
ETTm2	12	2025	29.02%	24.63%	29.02%	28.17%	25.17%	0.00%	0.00%
ETTm2	12	2026	28.80%	28.80%	31.18%	22.77%	27.67%	11.72%	11.72%
ETTm2	24	2024	9.89%	1.40%	9.89%	6.53%	14.08%	-28.11%	-15.39%
ETTm2	24	2025	0.02%	-0.01%	0.02%	-0.08%	5.85%	-6.67%	-6.61%
ETTm2	24	2026	3.24%	0.09%	0.08%	3.24%	2.73%	0.00%	0.00%
ETTm2	48	2024	-0.24%	-0.07%	-0.24%	-0.01%	-0.69%	0.00%	-13.48%
ETTm2	48	2025	3.06%	-3.17%	3.06%	0.00%	-0.50%	-22.78%	-20.39%
ETTm2	48	2026	0.37%	0.37%	0.11%	0.00%	0.20%	-7.02%	-23.56%
Traffic	1	2024	5.70%	2.39%	0.22%	5.70%	1.48%	-0.00%	-0.17%
Traffic	1	2025	6.98%	3.01%	6.98%	4.62%	0.00%	1.05%	3.21%
Traffic	1	2026	4.19%	2.07%	-0.71%	4.19%	-0.01%	0.09%	0.30%
Traffic	12	2024	9.67%	1.81%	9.67%	5.43%	2.69%	0.00%	0.85%
Traffic	12	2025	-0.35%	-4.30%	-5.08%	-0.35%	-8.29%	0.00%	-2.07%
Traffic	12	2026	4.11%	1.65%	3.11%	4.11%	1.77%	-0.00%	-0.16%
Traffic	24	2024	1.61%	1.61%	0.01%	1.44%	0.00%	-0.00%	1.38%
Traffic	24	2025	1.77%	1.77%	1.43%	0.00%	0.00%	-0.00%	2.63%
Traffic	24	2026	3.29%	3.29%	2.20%	0.01%	1.45%	0.00%	0.30%
Traffic	48	2024	2.20%	2.20%	1.09%	-0.00%	3.02%	-0.00%	-0.49%
Traffic	48	2025	-0.00%	0.00%	-0.00%	0.00%	0.00%	0.00%	-1.88%
Traffic	48	2026	11.79%	12.78%	11.73%	11.79%	12.26%	1.17%	-1.07%
Weather	1	2024	12.74%	12.74%	0.02%	15.12%	13.27%	-0.62%	10.85%
Weather	1	2025	2.92%	2.92%	0.63%	0.35%	2.35%	-0.00%	1.02%
Weather	1	2026	14.60%	3.13%	0.32%	14.60%	12.16%	-0.17%	10.26%
Weather	12	2024	40.65%	35.97%	40.65%	39.22%	41.93%	-0.00%	7.98%
Weather	12	2025	8.60%	0.50%	5.71%	8.60%	4.97%	0.00%	6.83%
Weather	12	2026	0.71%	1.42%	0.71%	0.01%	0.00%	0.00%	0.14%
Weather	24	2024	30.99%	30.99%	21.58%	26.06%	26.10%	20.34%	22.22%
Weather	24	2025	1.54%	1.54%	1.00%	0.15%	-0.01%	0.31%	10.98%
Weather	24	2026	21.60%	31.44%	26.31%	21.60%	27.58%	33.83%	41.21%
Weather	48	2024	30.23%	17.02%	30.23%	28.93%	30.58%	-1.96%	-2.50%
Weather	48	2025	40.34%	28.45%	40.34%	35.09%	40.83%	27.11%	53.49%
Weather	48	2026	38.88%	31.67%	44.43%	38.88%	41.77%	37.29%	45.79%

The matched objective audit adds Raw-A MSE and Mao exact to the comparison. Table 9 gives the cell-first aggregates; Table 10 retains all 84 cells, including negative recovery and near-zero CAL fallbacks. Mao CAL wins/ties/loses against Selected in 22/0/62 cells. Natural has higher nominal mean recovery but is not a Validation-controlled operating point.

Table 9: Matched objective audit over the 84 cells. Intervals resample datasets; negative cells retain $L < 0$.

Method	Mean L/O	Median L/O	Dataset 95% CI	Neg. cells
Selected	11.06%	4.90%	[6.66%, 15.67%]	5
Raw-A	8.96%	2.32%	[4.79%, 13.57%]	14
Wt.-BCE	7.61%	1.69%	[4.40%, 11.75%]	13
Mao Natural	12.58%	5.89%	[5.92%, 19.88%]	21
Mao CAL	9.07%	2.05%	[4.53%, 13.83%]	16

Table 10: Complete matched objective-control recovery matrix. Natural uses the zero score boundary; CAL freezes its threshold and override control on Validation.

Dataset	H	Seed	Selected	Raw-A	Wt.-BCE	Mao Nat.	Mao CAL	Nat. over.	CAL over.
ECL	1	2024	0.68%	0.00%	0.00%	1.01%	0.00%	25.77%	0.00%
ECL	1	2025	0.00%	0.00%	0.00%	2.71%	-0.00%	45.41%	0.00%
ECL	1	2026	0.94%	0.01%	0.00%	2.99%	0.00%	20.98%	0.00%
ECL	12	2024	7.16%	6.18%	2.83%	10.22%	6.57%	22.40%	2.92%
ECL	12	2025	2.35%	2.30%	2.35%	2.11%	-0.00%	15.84%	0.00%
ECL	12	2026	0.87%	-0.00%	0.00%	-0.79%	-0.01%	7.71%	0.00%
ECL	24	2024	3.54%	3.10%	3.72%	4.07%	2.00%	9.16%	0.76%
ECL	24	2025	0.00%	0.02%	0.00%	0.54%	-0.00%	8.57%	0.00%
ECL	24	2026	3.20%	4.49%	0.87%	5.80%	2.10%	54.06%	1.09%
ECL	48	2024	12.80%	12.75%	9.32%	28.71%	7.18%	73.36%	10.32%
ECL	48	2025	2.66%	4.50%	1.80%	14.57%	2.55%	63.65%	1.23%
ECL	48	2026	9.93%	6.69%	9.93%	19.17%	2.55%	66.31%	3.45%
ETTh1	1	2024	4.02%	0.04%	4.02%	-0.90%	-0.08%	22.79%	0.13%
ETTh1	1	2025	24.02%	20.82%	9.16%	29.94%	10.65%	60.41%	14.69%
ETTh1	1	2026	0.00%	0.13%	-0.04%	-5.39%	0.07%	30.30%	0.02%
ETTh1	12	2024	31.82%	34.13%	31.82%	35.77%	33.15%	40.75%	15.38%
ETTh1	12	2025	28.66%	29.35%	20.52%	35.12%	13.96%	45.18%	14.98%
ETTh1	12	2026	13.53%	10.97%	13.53%	4.17%	12.76%	38.55%	8.32%
ETTh1	24	2024	1.26%	0.23%	0.48%	-2.89%	-0.61%	39.11%	2.39%
ETTh1	24	2025	-0.34%	-0.39%	-1.39%	-17.49%	-0.46%	64.30%	2.67%
ETTh1	24	2026	0.00%	-0.00%	0.00%	-0.23%	0.00%	29.81%	0.01%
ETTh1	48	2024	14.93%	0.24%	0.28%	23.19%	17.76%	34.63%	11.17%
ETTh1	48	2025	15.64%	-0.00%	0.04%	23.66%	12.25%	51.98%	9.73%
ETTh1	48	2026	0.30%	-0.21%	-0.33%	-4.78%	0.00%	39.22%	0.01%
ETTh2	1	2024	5.24%	4.41%	1.36%	5.18%	0.80%	36.00%	1.82%
ETTh2	1	2025	1.11%	0.75%	1.11%	0.71%	0.52%	49.88%	11.77%
ETTh2	1	2026	12.93%	9.84%	6.80%	15.84%	8.41%	47.08%	14.79%
ETTh2	12	2024	32.02%	11.48%	28.94%	37.77%	26.14%	38.03%	15.66%
ETTh2	12	2025	18.78%	5.40%	13.31%	10.55%	0.33%	34.61%	0.02%
ETTh2	12	2026	6.70%	-0.02%	5.61%	5.99%	6.72%	19.00%	6.01%
ETTh2	24	2024	39.94%	44.01%	19.20%	54.42%	40.46%	79.40%	22.56%
ETTh2	24	2025	7.19%	0.06%	0.76%	6.97%	0.06%	53.35%	0.03%
ETTh2	24	2026	-0.16%	-1.64%	0.00%	-6.29%	-2.15%	73.95%	0.22%
ETTh2	48	2024	20.43%	31.98%	4.59%	54.57%	30.76%	88.65%	17.26%
ETTh2	48	2025	32.46%	32.38%	14.14%	57.27%	26.86%	88.44%	15.20%
ETTh2	48	2026	44.07%	41.67%	0.14%	63.07%	33.77%	95.47%	13.87%
ETTm1	1	2024	0.06%	0.06%	-0.12%	-2.12%	-0.00%	14.33%	0.00%
ETTm1	1	2025	20.42%	9.85%	20.30%	15.16%	7.45%	50.25%	6.85%
ETTm1	1	2026	0.00%	-0.00%	-0.00%	-1.08%	0.00%	24.65%	0.00%
ETTm1	12	2024	13.22%	0.61%	0.03%	11.59%	9.50%	47.58%	1.96%
ETTm1	12	2025	31.82%	31.55%	31.85%	35.15%	31.13%	57.75%	19.08%
ETTm1	12	2026	42.09%	39.02%	32.92%	41.04%	33.76%	59.89%	17.70%
ETTm1	24	2024	0.00%	0.00%	0.00%	21.09%	0.22%	42.32%	0.82%
ETTm1	24	2025	19.37%	0.33%	-0.36%	15.90%	24.05%	70.08%	19.50%
ETTm1	24	2026	16.17%	14.72%	16.17%	20.09%	10.16%	58.81%	20.33%
ETTm1	48	2024	11.51%	1.49%	1.58%	18.03%	17.51%	56.34%	22.49%
ETTm1	48	2025	1.26%	0.04%	-0.02%	6.70%	0.06%	46.90%	0.11%
ETTm1	48	2026	0.59%	3.95%	0.93%	1.54%	0.00%	50.24%	0.00%
ETTm2	1	2024	4.55%	5.01%	4.55%	2.73%	4.60%	28.81%	4.40%
ETTm2	1	2025	0.00%	-0.00%	-0.00%	-12.23%	-0.09%	53.50%	0.02%
ETTm2	1	2026	0.00%	0.00%	0.00%	-0.63%	-0.02%	20.12%	0.00%
ETTm2	12	2024	0.07%	0.01%	-0.01%	-2.33%	0.01%	26.44%	0.02%
ETTm2	12	2025	29.02%	25.17%	28.17%	32.10%	29.11%	33.40%	17.47%
ETTm2	12	2026	28.80%	27.67%	22.77%	34.07%	26.35%	54.71%	15.92%
ETTm2	24	2024	9.89%	14.08%	6.53%	5.03%	10.88%	43.93%	7.80%
ETTm2	24	2025	0.02%	5.85%	-0.08%	11.93%	3.74%	54.01%	2.61%
ETTm2	24	2026	3.24%	2.73%	3.24%	-5.86%	3.66%	41.58%	1.60%
ETTm2	48	2024	-0.24%	-0.69%	-0.01%	-1.97%	-0.11%	33.96%	20.79%
ETTm2	48	2025	3.06%	-0.50%	0.00%	1.35%	-0.19%	28.13%	0.03%
ETTm2	48	2026	0.37%	0.20%	0.00%	-12.86%	0.06%	44.63%	0.01%
Traffic	1	2024	5.70%	1.48%	5.70%	-2.54%	1.46%	33.86%	15.46%
Traffic	1	2025	6.98%	0.00%	4.62%	6.56%	4.23%	41.63%	15.94%
Traffic	1	2026	4.19%	-0.01%	4.19%	-1.41%	-0.49%	25.92%	10.85%
Traffic	12	2024	9.67%	2.69%	5.43%	3.20%	0.91%	23.11%	5.52%
Traffic	12	2025	-0.35%	-8.29%	-0.35%	-2.26%	-1.22%	25.78%	13.40%
Traffic	12	2026	4.11%	1.77%	4.11%	2.93%	2.74%	15.96%	10.41%
Traffic	24	2024	1.61%	0.00%	1.44%	3.96%	1.56%	32.58%	1.22%
Traffic	24	2025	1.77%	0.00%	0.00%	-0.41%	0.00%	23.39%	0.00%
Traffic	24	2026	3.29%	1.45%	0.01%	3.80%	1.40%	24.12%	1.88%
Traffic	48	2024	2.20%	3.02%	-0.00%	0.81%	1.18%	26.23%	1.99%
Traffic	48	2025	-0.00%	0.00%	0.00%	-34.21%	0.00%	20.26%	0.00%
Traffic	48	2026	11.79%	12.26%	11.79%	7.89%	0.00%	42.83%	0.00%
Weather	1	2024	12.74%	13.27%	15.12%	17.15%	15.71%	44.73%	21.38%
Weather	1	2025	2.92%	2.35%	0.35%	5.60%	0.55%	32.15%	0.65%
Weather	1	2026	14.60%	12.16%	14.60%	12.66%	12.67%	62.73%	14.20%
Weather	12	2024	40.65%	41.93%	39.22%	44.71%	41.14%	46.54%	20.72%
Weather	12	2025	8.60%	4.97%	8.60%	7.48%	6.91%	36.65%	10.87%
Weather	12	2026	0.71%	0.00%	0.01%	3.96%	1.13%	17.44%	0.30%
Weather	24	2024	30.99%	26.10%	26.06%	38.75%	31.60%	44.74%	21.57%
Weather	24	2025	1.54%	-0.01%	0.15%	25.20%	-0.02%	45.10%	0.03%
Weather	24	2026	21.60%	27.58%	21.60%	40.21%	29.62%	69.52%	23.98%
Weather	48	2024	30.23%	30.58%	28.93%	28.93%	28.79%	42.53%	22.69%
Weather	48	2025	40.34%	40.83%	35.09%	45.19%	33.50%	61.10%	17.35%
Weather	48	2026	38.88%	41.77%	38.88%	47.16%	41.69%	57.05%	29.12%

E.1 Router-capacity control

The C0–C4 control changes only router width or depth. C0 is linear; C1 is the 5,145-parameter headline model; C2 and C3 use one hidden layer of width 128 and 256; C4 uses two 128-unit hidden layers. All architectures receive the same 76 features, candidate tensors, objective set, chronological splits, and CAL rule. Table 11 reports the prespecified 84-cell comparison. C0 is worse than C1 under the paired dataset-clustered interval. The larger architectures have similar mean recovery, but their paired intervals cross zero and their adverse-cell counts do not improve uniformly. They therefore do not support selecting a larger router from TEST.

Table 11: Matched router-capacity control on 84 PatchTST cells. Architectures share candidates, 76 issue-time features, objectives, splits, and CAL rules. ΔL is paired against C1; brackets are dataset-clustered 95% intervals, in 10^{-3}.

Arch.	Params	Mean L/O (%)	Median L/O (%)	Neg. L	Neg. block	Neg. phase	ΔL vs. C1
C0	77	6.68	0.05	21	52	46	-0.523 [-0.832, -0.248]
C1	5,145	11.06	4.90	5	33	28	–
C2	10,137	11.12	6.09	5	35	25	-0.012 [-0.062, 0.023]
C3	20,121	11.39	6.82	8	34	30	0.012 [-0.044, 0.062]
C4	26,649	11.92	6.85	7	34	25	0.067 [-0.100, 0.190]

F Intervention and Issuance-Phase Diagnostics

Table 12 separates the requested Validation target from realized TEST override. Avoided harm and missed benefit use Equation (8); their difference equals L up to saved numeric precision. The worst-block column is descriptive because the blocks are read after the policy has been frozen and evaluated.

Table 12: Complete selected-policy intervention matrix. Requested Validation targets were converted to frozen thresholds; TEST override and utility columns are outcomes.

Dataset	H	Seed	Val. target	Override	Avoided harm	Missed benefit	L	Worst block
ECL	1	2024	2%	1.39%	7.033e-05	4.978e-05	2.054e-05	7.732e-06
ECL	1	2025	0%	0.00%	4.679e-08	1.697e-08	3.353e-08	0.000e+00
ECL	1	2026	2%	1.86%	7.858e-05	4.591e-05	3.266e-05	2.174e-05
ECL	12	2024	5%	3.39%	6.860e-04	3.282e-04	3.578e-04	1.706e-04
ECL	12	2025	2%	2.69%	3.406e-04	2.507e-04	8.996e-05	-1.229e-04
ECL	12	2026	1%	0.74%	7.722e-05	5.530e-05	2.191e-05	1.387e-06
ECL	24	2024	2%	1.68%	1.729e-04	1.233e-04	4.964e-05	2.423e-05
ECL	24	2025	0%	0.00%	2.472e-08	0.000e+00	2.980e-08	0.000e+00
ECL	24	2026	10%	6.36%	8.999e-04	7.845e-04	1.154e-04	-1.922e-04
ECL	48	2024	15%	14.11%	1.566e-03	7.169e-04	8.488e-04	4.273e-04
ECL	48	2025	1%	0.68%	1.924e-04	6.726e-05	1.252e-04	1.974e-05
ECL	48	2026	15%	13.03%	1.222e-03	4.275e-04	7.940e-04	5.119e-04
ETTh1	1	2024	5%	4.07%	3.409e-03	2.334e-03	1.074e-03	-1.323e-03
ETTh1	1	2025	15%	16.12%	3.064e-02	1.669e-02	1.395e-02	7.980e-03
ETTh1	1	2026	0%	0.00%	5.570e-07	0.000e+00	5.364e-07	0.000e+00
ETTh1	12	2024	15%	11.34%	5.857e-03	1.160e-03	4.697e-03	2.076e-04
ETTh1	12	2025	15%	14.85%	1.420e-03	6.495e-04	7.704e-04	1.184e-04
ETTh1	12	2026	5%	9.33%	8.602e-03	4.463e-03	4.139e-03	1.919e-03
ETTh1	24	2024	5%	6.39%	1.529e-03	1.351e-03	1.777e-04	-1.186e-03
ETTh1	24	2025	10%	10.10%	6.886e-04	7.088e-04	-2.018e-05	-1.915e-04
ETTh1	24	2026	0%	0.04%	1.830e-06	1.181e-06	6.557e-07	-3.125e-06
ETTh1	48	2024	15%	10.99%	2.427e-03	1.121e-03	1.307e-03	3.619e-04
ETTh1	48	2025	15%	8.63%	7.646e-03	2.930e-03	4.716e-03	7.818e-04
ETTh1	48	2026	1%	0.45%	7.360e-05	4.293e-05	3.064e-05	-3.954e-05
ETTh2	1	2024	15%	15.29%	6.965e-04	4.995e-04	1.970e-04	3.906e-05
ETTh2	1	2025	2%	2.53%	1.585e-04	8.428e-05	7.427e-05	-2.915e-06
ETTh2	1	2026	15%	18.65%	5.813e-04	3.303e-04	2.510e-04	1.379e-04
ETTh2	12	2024	15%	13.98%	1.652e-03	3.903e-04	1.262e-03	2.708e-06
ETTh2	12	2025	15%	14.26%	9.377e-04	5.266e-04	4.112e-04	-3.202e-04
ETTh2	12	2026	10%	13.15%	3.081e-04	2.273e-04	8.082e-05	-9.241e-07
ETTh2	24	2024	12%	19.84%	1.430e-02	2.643e-03	1.166e-02	2.030e-03
ETTh2	24	2025	15%	14.63%	1.300e-03	8.192e-04	4.806e-04	-1.822e-04
ETTh2	24	2026	0%	0.01%	0.000e+00	1.111e-05	-1.110e-05	-5.628e-05
ETTh2	48	2024	5%	10.30%	5.199e-03	1.100e-03	4.098e-03	1.972e-04
ETTh2	48	2025	15%	18.29%	1.401e-02	4.348e-03	9.666e-03	1.147e-03
ETTh2	48	2026	15%	21.73%	1.336e-02	3.546e-03	9.809e-03	1.389e-03
ETTm1	1	2024	2%	2.66%	3.976e-06	1.327e-06	2.649e-06	1.085e-06
ETTm1	1	2025	15%	16.24%	6.263e-03	2.332e-03	3.931e-03	2.629e-03
ETTm1	1	2026	0%	0.01%	1.257e-07	4.548e-09	1.267e-07	0.000e+00
ETTm1	12	2024	5%	5.58%	2.157e-03	1.237e-03	9.194e-04	-6.294e-04
ETTm1	12	2025	15%	20.41%	8.231e-03	2.703e-03	5.528e-03	9.357e-04
ETTm1	12	2026	15%	18.66%	1.098e-02	2.089e-03	8.890e-03	5.121e-03
ETTm1	24	2024	0%	0.00%	3.387e-08	8.459e-09	5.960e-08	-6.767e-08
ETTm1	24	2025	12%	15.14%	4.905e-04	2.369e-04	2.536e-04	1.215e-04
ETTm1	24	2026	12%	15.79%	1.169e-03	4.597e-04	7.092e-04	2.242e-04
ETTm1	48	2024	12%	15.77%	4.546e-03	2.491e-03	2.056e-03	1.553e-04
ETTm1	48	2025	1%	1.82%	3.590e-04	2.313e-04	1.277e-04	-1.296e-04
ETTm1	48	2026	0%	1.56%	3.500e-05	2.148e-05	1.353e-05	-1.750e-06
ETTm2	1	2024	10%	10.88%	5.790e-04	3.513e-04	2.277e-04	9.496e-05
ETTm2	1	2025	0%	0.01%	1.446e-08	1.104e-08	7.451e-09	-1.809e-09
ETTm2	1	2026	0%	0.00%	1.862e-08	1.570e-08	3.725e-09	-1.256e-07
ETTm2	12	2024	2%	3.17%	1.171e-04	1.157e-04	1.386e-06	-6.809e-05
ETTm2	12	2025	15%	18.66%	1.079e-03	3.316e-04	7.474e-04	4.720e-04
ETTm2	12	2026	15%	15.18%	1.705e-03	5.766e-04	1.129e-03	5.337e-04
ETTm2	24	2024	2%	4.82%	1.008e-03	6.128e-04	3.956e-04	-8.348e-04
ETTm2	24	2025	0%	0.04%	3.910e-06	2.763e-06	1.147e-06	-7.569e-06
ETTm2	24	2026	2%	3.06%	2.894e-04	2.064e-04	8.301e-05	-8.123e-05
ETTm2	48	2024	0%	0.04%	1.479e-06	6.485e-06	-5.007e-06	-2.643e-05
ETTm2	48	2025	5%	8.15%	4.381e-04	3.745e-04	6.366e-05	-5.022e-04
ETTm2	48	2026	1%	0.93%	7.246e-05	6.099e-05	1.144e-05	-7.926e-05
Traffic	1	2024	15%	15.57%	2.232e-03	1.757e-03	4.745e-04	2.604e-04
Traffic	1	2025	15%	16.67%	4.329e-03	3.152e-03	1.177e-03	3.437e-04
Traffic	1	2026	15%	16.18%	3.089e-03	2.585e-03	5.044e-04	1.974e-04
Traffic	12	2024	15%	16.95%	2.410e-03	1.414e-03	9.957e-04	4.043e-04
Traffic	12	2025	15%	15.66%	3.946e-03	3.981e-03	-3.567e-05	-1.751e-03
Traffic	12	2026	12%	13.05%	9.574e-04	7.240e-04	2.333e-04	4.962e-05
Traffic	24	2024	5%	4.72%	3.432e-04	2.155e-04	1.277e-04	-1.032e-06
Traffic	24	2025	15%	15.69%	1.068e-03	9.501e-04	1.180e-04	-5.008e-04
Traffic	24	2026	12%	10.39%	8.695e-04	5.642e-04	3.053e-04	1.115e-04
Traffic	48	2024	10%	8.03%	5.866e-04	4.463e-04	1.403e-04	-6.634e-05
Traffic	48	2025	0%	0.00%	0.000e+00	5.762e-08	-2.980e-08	-3.292e-07
Traffic	48	2026	10%	14.45%	2.628e-03	1.840e-03	7.883e-04	-1.911e-04
Weather	1	2024	12%	12.01%	2.039e-03	9.583e-04	1.081e-03	7.937e-05
Weather	1	2025	12%	8.27%	8.590e-04	4.556e-04	4.034e-04	1.798e-04
Weather	1	2026	12%	17.66%	1.041e-03	4.846e-04	5.561e-04	1.609e-05
Weather	12	2024	15%	21.66%	5.240e-03	1.629e-03	3.611e-03	1.101e-03
Weather	12	2025	15%	14.03%	1.579e-03	1.023e-03	5.557e-04	2.176e-04

Continued on next page

Table 12 (continued)

Dataset	H	Seed	Val. target	Override	Avoided harm	Missed benefit	L	Worst block
Weather	12	2026	0%	0.19%	8.935e-05	1.511e-05	7.427e-05	0.000e+00
Weather	24	2024	15%	28.47%	3.258e-03	1.426e-03	1.832e-03	4.982e-04
Weather	24	2025	1%	1.76%	4.837e-04	3.247e-04	1.591e-04	-1.133e-04
Weather	24	2026	15%	13.34%	2.693e-03	6.899e-04	2.003e-03	5.101e-04
Weather	48	2024	15%	28.17%	5.969e-03	2.789e-03	3.180e-03	7.239e-04
Weather	48	2025	15%	22.45%	6.886e-03	1.791e-03	5.095e-03	-1.535e-04
Weather	48	2026	15%	38.18%	5.755e-03	2.108e-03	3.648e-03	1.120e-04

Table 13 compares Selected with representative points from the prespecified raw-quantile and online-prefix-cap sweeps. Raw quantile thresholds are frozen on Validation and reused on TEST. The online prefix cap limits cumulative Frozen actions after every issue time using only current and past scores. Both sweeps are TEST-descriptive and cannot be used to replace the headline policy. The non-monotone recovery and rising adverse counts at broader exposure are the relevant findings, not the best point on either TEST curve.

Table 13: Selected operating point and prespecified TEST-descriptive sensitivity points. Requested Validation targets, realized TEST override, and online prefix hard caps are distinct. The complete 0–100% grid is retained in the machine-readable supplement.

Family	Control	TEST override (%)	Mean L/O (%)	Median L/O (%)	Neg. L	Neg. block	Neg. phase
Selected	cell-specific CAL	9.92	11.06	4.90	5	33	28
Selected-score natural	score $= 0$	43.41	11.32	5.09	22	55	44
Quantile raw	15%	16.81	10.90	6.40	16	40	36
Quantile raw	25%	27.00	11.09	5.20	18	49	40
Quantile raw	50%	50.90	7.26	3.28	39	60	54
Online hard cap	15%	13.95	8.78	5.25	15	41	38
Online hard cap	25%	23.52	9.75	4.84	18	47	40
Online hard cap	50%	47.61	7.51	2.89	39	59	53

The issuance-phase sensitivity reuses each recorded selected policy without fitting or selecting it again. The worst chronological block divides a TEST cell into at most eight contiguous, approximately equal-length partitions and takes the minimum recovery. This differs from the 48-forecast-window moving blocks used only for interval estimation. For horizon H, issues are partitioned by index modulo H, producing non-overlapping target windows within a phase. Table 14 compresses the 1,785 phase records to one row per fitted policy. Phase rows from one cell share data, candidates, and parameters and are never treated as independent replicates. The matured-advantage EMA used by the simple baseline applies the globally fixed update $0.10\,A_{\text{new}} + 0.90\,\text{EMA}_{\text{old}}$.

Table 14: Complete non-overlapping issuance-phase sensitivity. The same recorded policy is reused for every phase; no phasewise reselection occurs.

Dataset	H	Seed	Positive phases	Median L	Min. L	Max. L	Worst phase
ECL	1	2024	100.0%	2.055e-05	2.055e-05	2.055e-05	0
ECL	1	2025	100.0%	2.983e-08	2.983e-08	2.983e-08	0
ECL	1	2026	100.0%	3.267e-05	3.267e-05	3.267e-05	0
ECL	12	2024	100.0%	3.705e-04	2.259e-04	4.563e-04	5
ECL	12	2025	100.0%	9.185e-05	2.790e-05	1.547e-04	6
ECL	12	2026	91.7%	1.808e-05	-1.126e-05	5.850e-05	4
ECL	24	2024	100.0%	4.542e-05	2.696e-06	1.164e-04	17
ECL	24	2025	4.2%	0.000e+00	0.000e+00	5.930e-07	0
ECL	24	2026	91.7%	6.653e-05	-3.001e-05	3.373e-04	12
ECL	48	2024	100.0%	7.084e-04	3.662e-04	1.669e-03	32
ECL	48	2025	100.0%	1.412e-04	2.796e-05	1.939e-04	13
ECL	48	2026	100.0%	6.329e-04	1.161e-04	2.219e-03	38
ETTh1	1	2024	100.0%	1.074e-03	1.074e-03	1.074e-03	0
ETTh1	1	2025	100.0%	1.395e-02	1.395e-02	1.395e-02	0
ETTh1	1	2026	100.0%	5.570e-07	5.570e-07	5.570e-07	0
ETTh1	12	2024	100.0%	3.755e-03	1.908e-03	9.358e-03	7
ETTh1	12	2025	100.0%	7.587e-04	4.733e-04	1.123e-03	5
ETTh1	12	2026	75.0%	4.654e-03	-3.700e-03	1.112e-02	9
ETTh1	24	2024	66.7%	2.312e-04	-1.586e-03	1.993e-03	3
ETTh1	24	2025	50.0%	-3.702e-06	-6.917e-04	8.573e-04	17
ETTh1	24	2026	29.2%	0.000e+00	-1.300e-05	1.436e-05	23
ETTh1	48	2024	68.8%	1.206e-03	-7.134e-04	5.144e-03	29
ETTh1	48	2025	85.4%	4.812e-03	-1.309e-03	1.237e-02	17
ETTh1	48	2026	58.3%	1.075e-05	-2.047e-04	5.492e-04	29
ETTh2	1	2024	100.0%	1.970e-04	1.970e-04	1.970e-04	0
ETTh2	1	2025	100.0%	7.426e-05	7.426e-05	7.426e-05	0
ETTh2	1	2026	100.0%	2.510e-04	2.510e-04	2.510e-04	0
ETTh2	12	2024	100.0%	1.255e-03	1.050e-03	1.432e-03	4
ETTh2	12	2025	100.0%	4.071e-04	2.895e-04	5.565e-04	6
ETTh2	12	2026	58.3%	5.810e-05	-1.770e-04	3.523e-04	7
ETTh2	24	2024	100.0%	1.165e-02	1.095e-02	1.233e-02	12
ETTh2	24	2025	83.3%	3.893e-04	-2.341e-04	1.512e-03	21
ETTh2	24	2026	0.0%	0.000e+00	-8.978e-05	0.000e+00	16
ETTh2	48	2024	100.0%	4.189e-03	2.224e-03	5.045e-03	8
ETTh2	48	2025	100.0%	9.704e-03	8.389e-03	1.072e-02	10
ETTh2	48	2026	100.0%	9.494e-03	6.707e-03	1.390e-02	22
ETTm1	1	2024	100.0%	2.648e-06	2.648e-06	2.648e-06	0
ETTm1	1	2025	100.0%	3.931e-03	3.931e-03	3.931e-03	0
ETTm1	1	2026	100.0%	1.212e-07	1.212e-07	1.212e-07	0
ETTm1	12	2024	100.0%	8.143e-04	1.836e-04	1.861e-03	2
ETTm1	12	2025	100.0%	5.568e-03	4.949e-03	6.259e-03	2
ETTm1	12	2026	100.0%	8.908e-03	8.264e-03	9.505e-03	8
ETTm1	24	2024	4.2%	0.000e+00	-2.030e-07	8.129e-07	17
ETTm1	24	2025	100.0%	2.428e-04	4.711e-05	4.624e-04	23
ETTm1	24	2026	100.0%	6.950e-04	5.590e-04	9.234e-04	3
ETTm1	48	2024	100.0%	1.489e-03	1.600e-04	6.246e-03	22
ETTm1	48	2025	47.9%	-1.408e-05	-2.124e-04	8.886e-04	24
ETTm1	48	2026	62.5%	7.851e-06	-3.162e-05	7.959e-05	11
ETTm2	1	2024	100.0%	2.278e-04	2.278e-04	2.278e-04	0
ETTm2	1	2025	100.0%	3.416e-09	3.416e-09	3.416e-09	0
ETTm2	1	2026	100.0%	2.923e-09	2.923e-09	2.923e-09	0
ETTm2	12	2024	58.3%	2.187e-06	-3.229e-05	2.875e-05	1
ETTm2	12	2025	100.0%	7.488e-04	6.919e-04	8.155e-04	9
ETTm2	12	2026	100.0%	1.105e-03	1.036e-03	1.246e-03	6
ETTm2	24	2024	100.0%	4.307e-04	1.824e-04	6.087e-04	22
ETTm2	24	2025	45.8%	-8.381e-09	-9.226e-06	1.669e-05	9

Continued on next page

Table 14 (continued)

Dataset	H	Seed	Positive phases	Median L	Min. L	Max. L	Worst phase
ETTm2	24	2026	91.7%	6.529e-05	-2.750e-05	2.651e-04	6
ETTm2	48	2024	20.8%	-2.698e-07	-4.871e-05	9.287e-06	8
ETTm2	48	2025	62.5%	1.147e-04	-2.793e-04	3.642e-04	24
ETTm2	48	2026	52.1%	8.587e-07	-8.531e-05	1.144e-04	36
Traffic	1	2024	100.0%	4.745e-04	4.745e-04	4.745e-04	0
Traffic	1	2025	100.0%	1.177e-03	1.177e-03	1.177e-03	0
Traffic	1	2026	100.0%	5.044e-04	5.044e-04	5.044e-04	0
Traffic	12	2024	100.0%	9.841e-04	5.209e-04	1.521e-03	7
Traffic	12	2025	33.3%	-1.382e-04	-6.240e-04	5.135e-04	9
Traffic	12	2026	100.0%	2.103e-04	3.203e-05	4.130e-04	3
Traffic	24	2024	100.0%	1.128e-04	1.085e-05	2.519e-04	21
Traffic	24	2025	66.7%	1.536e-04	-3.721e-04	4.675e-04	14
Traffic	24	2026	100.0%	2.768e-04	1.355e-04	6.082e-04	23
Traffic	48	2024	79.2%	1.129e-04	-1.625e-04	5.132e-04	16
Traffic	48	2025	0.0%	0.000e+00	-1.165e-06	0.000e+00	27
Traffic	48	2026	95.8%	7.470e-04	-4.897e-05	1.748e-03	45
Weather	1	2024	100.0%	1.081e-03	1.081e-03	1.081e-03	0
Weather	1	2025	100.0%	4.034e-04	4.034e-04	4.034e-04	0
Weather	1	2026	100.0%	5.560e-04	5.560e-04	5.560e-04	0
Weather	12	2024	100.0%	3.692e-03	3.073e-03	4.029e-03	7
Weather	12	2025	75.0%	4.737e-04	-2.859e-04	1.443e-03	8
Weather	12	2026	100.0%	7.077e-05	1.001e-05	1.560e-04	10
Weather	24	2024	100.0%	1.774e-03	1.498e-03	2.261e-03	0
Weather	24	2025	83.3%	1.277e-04	-1.587e-04	4.445e-04	17
Weather	24	2026	100.0%	1.977e-03	1.588e-03	2.479e-03	23
Weather	48	2024	100.0%	3.391e-03	9.411e-04	5.473e-03	8
Weather	48	2025	100.0%	5.045e-03	4.146e-03	6.209e-03	14
Weather	48	2026	100.0%	3.626e-03	2.719e-03	4.637e-03	12

Table 15 aligns recovery with realized TEST exposure and local downside. Table 16 gives the full Mao Natural/CAL block and issuance-phase records. "Adverse block" and "adverse phase" count a cell once when at least one corresponding partition has negative recovery; the complete matrix also reports how many such partitions occur inside each cell.

Table 15: Aggregate recovery, exposure, and local downside. Adverse-block and adverse-phase columns count cells with at least one negative partition.

Method	Mean L/O	Mean over.	Neg. cells	Adv. block	Adv. phase	Mean worst block	Mean min phase
Selected	11.06%	9.92%	5	33	28	2.837e-04	9.086e-04
Mao Natural	12.58%	42.29%	21	57	46	1.922e-05	6.223e-04
Mao CAL	9.07%	7.83%	16	41	33	1.507e-04	7.163e-04

Table 16: Complete Mao local-risk matrix. Worst block and phase minimum are recoveries relative to always-Fixed; counts retain every adverse partition.

Dataset	H	Seed	Nat. worst blk.	CAL worst blk.	Nat. min ph.	CAL min ph.	Nat. n_b^-	CAL n_b^-	Nat. n_p^-	CAL n_p^-
ECL	1	2024	-3.225e-05	0.000e+00	3.064e-05	1.695e-08	1	0	0	0
ECL	1	2025	1.312e-05	-2.601e-07	3.275e-05	-3.252e-08	0	1	0	1
ECL	1	2026	1.297e-05	-5.964e-10	1.039e-04	1.288e-08	0	1	0	0
ECL	12	2024	2.890e-04	1.426e-04	3.228e-04	2.139e-04	0	0	0	0
ECL	12	2025	-6.301e-04	-7.235e-07	-1.287e-04	-6.713e-07	3	4	3	4
ECL	12	2026	-5.650e-05	-8.836e-07	-4.864e-05	-3.615e-06	7	3	11	2
ECL	24	2024	-1.378e-05	-3.490e-06	-2.924e-05	-1.232e-05	1	2	2	1
ECL	24	2025	-1.523e-05	-5.208e-08	-4.080e-05	-1.120e-07	5	2	15	4
ECL	24	2026	-1.651e-04	-1.287e-04	-1.761e-04	-2.694e-05	1	1	6	2
ECL	48	2024	3.629e-04	9.352e-05	8.536e-04	1.134e-04	0	0	0	0
ECL	48	2025	-5.528e-06	2.232e-05	6.144e-05	3.455e-05	1	0	0	0
ECL	48	2026	-4.417e-04	2.930e-05	-6.909e-04	1.101e-05	1	0	7	0
ETTh1	1	2024	-3.053e-03	-1.440e-04	-2.407e-04	-2.069e-05	3	4	1	1
ETTh1	1	2025	1.185e-02	1.705e-03	1.738e-02	6.183e-03	0	0	0	0
ETTh1	1	2026	-4.440e-03	0.000e+00	-1.236e-03	1.550e-05	5	0	1	0
ETTh1	12	2024	-1.448e-04	1.135e-04	4.343e-06	1.347e-03	1	0	0	0
ETTh1	12	2025	1.559e-04	1.196e-04	4.712e-04	2.604e-04	0	0	0	0
ETTh1	12	2026	-1.057e-03	9.246e-04	-1.052e-02	-3.006e-03	4	0	4	2
ETTh1	24	2024	-2.040e-03	-5.027e-04	-3.039e-03	-3.176e-04	5	5	15	18
ETTh1	24	2025	-2.317e-03	-2.186e-04	-2.952e-03	-3.004e-04	7	6	19	15
ETTh1	24	2026	-9.339e-04	-9.559e-08	-5.961e-04	-2.863e-07	4	1	13	1
ETTh1	48	2024	-1.813e-04	2.732e-04	-1.794e-03	-3.304e-04	1	0	7	8
ETTh1	48	2025	6.500e-04	1.097e-03	-2.626e-03	-3.843e-04	0	0	3	2
ETTh1	48	2026	-1.491e-03	-2.508e-06	-2.960e-03	-1.503e-05	5	1	34	1
ETTh2	1	2024	8.065e-05	4.653e-06	1.947e-04	3.012e-05	0	0	0	0
ETTh2	1	2025	-4.354e-04	-1.361e-04	4.752e-05	3.444e-05	2	2	0	0
ETTh2	1	2026	1.668e-04	7.663e-05	3.077e-04	1.633e-04	0	0	0	0
ETTh2	12	2024	-7.782e-06	4.096e-05	1.343e-03	8.589e-04	1	0	0	0
ETTh2	12	2025	-4.106e-04	0.000e+00	5.618e-05	0.000e+00	3	0	0	0
ETTh2	12	2026	-9.584e-05	2.407e-05	-2.640e-04	-2.697e-05	1	0	4	2
ETTh2	24	2024	7.076e-03	2.341e-03	1.534e-02	1.107e-02	0	0	0	0
ETTh2	24	2025	-7.634e-04	-1.544e-07	-4.742e-04	-6.934e-07	2	1	7	2
ETTh2	24	2026	-2.958e-03	-1.376e-03	-3.262e-03	-4.569e-04	5	2	15	15
ETTh2	48	2024	4.984e-03	4.109e-04	6.258e-03	2.921e-03	0	0	0	0
ETTh2	48	2025	2.720e-03	1.683e-03	1.482e-02	6.344e-03	0	0	0	0
ETTh2	48	2026	6.600e-03	5.021e-04	7.229e-03	4.966e-03	0	0	0	0
ETTm1	1	2024	-1.603e-04	-1.881e-07	-1.018e-04	-2.351e-08	8	1	1	1
ETTm1	1	2025	1.525e-03	7.787e-04	2.918e-03	1.435e-03	0	0	0	0
ETTm1	1	2026	-2.247e-04	-1.369e-07	-7.999e-05	8.761e-08	8	1	1	0
ETTm1	12	2024	-5.147e-04	-3.769e-04	6.791e-05	2.737e-04	2	2	0	0
ETTm1	12	2025	2.823e-03	1.496e-03	4.966e-03	4.694e-03	0	0	0	0
ETTm1	12	2026	4.470e-03	3.076e-03	7.565e-03	6.375e-03	0	0	0	0
ETTm1	24	2024	5.536e-04	-2.447e-05	-1.097e-04	-3.745e-05	0	2	2	6
ETTm1	24	2025	-2.277e-04	1.175e-04	-1.869e-04	1.580e-05	1	0	7	0
ETTm1	24	2026	3.511e-04	2.601e-04	6.269e-04	3.167e-04	0	0	0	0
ETTm1	48	2024	-2.683e-03	-1.235e-04	-1.654e-03	3.357e-04	2	1	9	0
ETTm1	48	2025	-1.270e-05	-1.282e-05	-2.661e-03	-1.563e-05	1	3	25	14
ETTm1	48	2026	-2.236e-04	0.000e+00	-1.307e-03	0.000e+00	2	0	22	0
ETTm2	1	2024	-1.735e-04	4.672e-05	1.364e-04	2.303e-04	2	0	0	0
ETTm2	1	2025	-6.563e-04	-9.537e-06	-2.693e-04	-1.979e-06	8	3	1	1
ETTm2	1	2026	-4.747e-05	-1.984e-06	-1.132e-05	-2.871e-07	6	2	1	1
ETTm2	12	2024	-2.621e-04	-3.169e-06	-1.327e-04	-1.520e-06	5	2	10	5
ETTm2	12	2025	5.065e-04	4.310e-04	7.704e-04	7.097e-04	0	0	0	0
ETTm2	12	2026	6.865e-04	5.532e-04	1.242e-03	9.612e-04	0	0	0	0
ETTm2	24	2024	-1.785e-03	-6.305e-04	-2.595e-04	2.493e-04	4	4	5	0
ETTm2	24	2025	-6.607e-04	3.129e-06	3.576e-04	7.737e-05	2	0	0	0
ETTm2	24	2026	-9.055e-04	-3.111e-05	-5.645e-04	1.274e-05	5	1	18	0
ETTm2	48	2024	-4.285e-04	-3.067e-04	-4.161e-04	-2.604e-04	4	3	26	23
ETTm2	48	2025	-7.784e-04	-3.524e-05	-4.555e-04	-3.942e-05	4	2	22	20
ETTm2	48	2026	-1.931e-03	-1.324e-07	-1.335e-03	0.000e+00	6	1	31	0
Traffic	1	2024	-4.873e-04	-8.749e-05	-2.114e-04	1.214e-04	5	2	1	0
Traffic	1	2025	1.635e-05	3.789e-04	1.105e-03	7.139e-04	0	0	0	0
Traffic	1	2026	-5.182e-04	-4.057e-04	-1.697e-04	-5.855e-05	5	4	1	1

Continued on next page

Table 16 (continued)

Dataset	H	Seed	Nat. worst blk.	CAL worst blk.	Nat. min ph.	CAL min ph.	Nat. n_b^-	CAL n_b^-	Nat. n_p^-	CAL n_p^-
Traffic	12	2024	-7.092e-04	-2.392e-04	-3.748e-04	-3.668e-04	1	3	4	5
Traffic	12	2025	-2.239e-03	-1.908e-03	-7.468e-04	-6.707e-04	5	4	9	8
Traffic	12	2026	-1.378e-04	-7.697e-05	-1.095e-04	-9.721e-05	2	2	2	2
Traffic	24	2024	-5.573e-04	-1.006e-04	-5.006e-04	-8.501e-05	2	1	7	3
Traffic	24	2025	-9.590e-04	0.000e+00	-4.839e-04	0.000e+00	3	0	13	0
Traffic	24	2026	-2.138e-04	3.034e-05	-1.006e-04	3.872e-06	2	0	2	0
Traffic	48	2024	-4.480e-04	-2.135e-05	-6.051e-04	-6.537e-05	2	2	20	9
Traffic	48	2025	-4.504e-03	0.000e+00	-4.960e-03	0.000e+00	8	0	48	0
Traffic	48	2026	-1.006e-03	0.000e+00	-7.247e-04	0.000e+00	3	0	16	0
Weather	1	2024	1.867e-04	4.079e-04	1.455e-03	1.333e-03	0	0	0	0
Weather	1	2025	-2.299e-04	-3.530e-07	7.749e-04	7.556e-05	1	1	0	0
Weather	1	2026	-9.858e-05	1.102e-04	4.822e-04	4.826e-04	1	0	0	0
Weather	12	2024	1.281e-03	1.313e-03	3.401e-03	3.283e-03	0	0	0	0
Weather	12	2025	-4.078e-04	-2.212e-05	-9.761e-04	-4.096e-04	1	1	5	3
Weather	12	2026	-5.486e-04	-2.329e-06	-1.029e-03	4.146e-05	2	2	4	0
Weather	24	2024	2.843e-04	9.991e-05	2.001e-03	1.506e-03	0	0	0	0
Weather	24	2025	-4.697e-04	-3.840e-05	1.301e-04	-4.050e-05	1	3	0	7
Weather	24	2026	4.913e-04	5.139e-05	2.581e-03	1.780e-03	0	0	0	0
Weather	48	2024	2.317e-04	5.225e-04	6.602e-05	9.421e-04	0	0	0	0
Weather	48	2025	2.630e-04	5.214e-04	4.747e-03	3.374e-03	0	0	0	0
Weather	48	2026	-1.180e-04	-1.700e-04	3.643e-03	3.315e-03	1	1	0	0

G External and Backbone Context

External online methods are included only to establish that the corrected candidate is nontrivial. Their available cells and adaptation procedures do not satisfy the fixed-pair contract used by the main 84-cell study, so their errors are not used to compute C, O, L, or L/O. Weather is unavailable for all four comparisons, and SOLID and OGD additionally lack Traffic-H48.

Table 17: Available seed-2026 PatchTST pairs with external online methods. Positive mean values favor the Validation-selected Prompt-Z policy; W/T/L uses that policy's perspective.

Method	Cells	Mean (%)	Median (%)	W/T/L
OneNet	24	31.528	25.325	23/0/1
PROCEED	24	6.132	-0.330	11/0/13
SOLID	23	4.845	-0.142	11/0/12
OGD	23	7.286	0.773	18/0/5

A separate earlier three-warmup study contains 28 seed-2026 iTransformer cells and reports a mean gain of 3.225% over Frozen with 27/0/1 wins/ties/losses. It does not contain the retrospective Oracle and does not share the present three-seed FIT–Validation/CAL–TEST lineage, so it is supplementary candidate-quality evidence only. CMoS and SOFTS artifacts are likewise excluded from the headline study because no complete fixed-pair opportunity, objective, and issuance-phase contract was frozen for them. These exclusions are why the main conclusion is not described as backbone- or generator-agnostic.

An earlier lineage attempt did not pool that historical study. All eight planned ECL/Traffic FIT/Validation/CAL cells completed, but TEST produced only one passing ECL-H1 record before ECL-H12 failed the frozen-reference equivalence check. The observed Fixed MSE was 0.0964570269 versus the recorded reference 0.0966273260, an absolute difference of 1.703×10^{-4}. Execution stopped at the guard, and that partial protocol contributes no performance evidence.

A reference-regenerated protocol repaired the lineage under a new protocol tag. Its eight-cell ECL/Traffic gate passed before the prespecified 24-cell ECL/ETTh2/Traffic/Weather matrix was run for horizons 12 and 48 and seeds 2024–2026. All 24 references and verifications pass, and every final record preserves candidate equivalence. Table 19 reports the separate result: opportunity is positive in all 24 cells, while Learned has 21/1/2 W/T/L against Fixed and mean/median L/O of 10.37%/3.72%. This supports the qualitative decomposition on the tested iTransformer subset, not a pooled or generator-agnostic claim.

Table 18: iTransformer protocol disposition. A failed legacy lineage is retained and not mixed with the reference-regenerated protocol.

Audit item	Recorded outcome
Legacy lineage	halted after ECL-H12 Fixed mismatch; no aggregate
Reference-regenerated gate	8/8 references and TEST verifications passed
Prespecified scope	24 cells
Completed scope	24/24 with candidate equivalence
Aggregation rule	iTransformer only; no pooling with PatchTST

Table 19: Separate 24-cell reference-verified iTransformer summary after all candidate-reference checks passed. C, O, and L are in 10^{-3}; no statistic is pooled with PatchTST.

Quantity	Definition	Mean	Median	W/T/L
Capacity C	$R_{\mathrm{F}} - R_{\mathrm{Fix}}$	4.375	2.416	18/0/6
Opportunity O	$R_{\mathrm{Fix}} - R_{\mathrm{O}}$	9.028	8.024	24/0/0
Recovery L	$R_{\mathrm{Fix}} - R_{\mathrm{L}}$	0.931	0.249	21/1/2
Recovery ratio L/O	cell-first	10.37%	3.72%	–

H Boundary Cases and Interpretation

ETTh2-H48 separates recovery from correction capacity and operating control. Learned recovers part of the hindsight opportunity in every seed, but its MSE remains above Frozen. Across the prespecified raw-quantile and hard-cap curves, broader override raises recovery; the curves approach Frozen only as actions approach always-Frozen, and no evaluated operating point beats Frozen in all three seeds. This establishes partial mitigation under the tested controls. It does not identify calibration, feature sufficiency, router class, or unavailable issue-time information as the unique cause of the remaining gap.

Table 20: ETTh2-H48 boundary case retained in full.

Seed	Frozen	Fixed	Learned	Oracle	L/O
2024	0.248492	0.258912	0.254814	0.238849	20.43%
2025	0.289458	0.304495	0.294829	0.274714	32.46%
2026	0.246455	0.260691	0.250882	0.238434	44.07%

The most adverse selected-policy chronological block is Traffic-H12, seed 2025, with $L = -1.751 \times 10^{-3}$. Its issuance-phase view is also non-uniform: one third of phases are positive and the worst phase has $L = -6.240 \times 10^{-4}$. These two partitions are related but not identical. The block statistic follows consecutive issues; the phase statistic removes target overlap within each residue class.