BETTER ON AVERAGE, WORSE LOCALLY: INTERVAL GUARANTEES FOR ONLINE FORECAST ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

An adaptive forecaster can improve total error while worsening a recent interval. We study a causal guard that bounds the eventual excess MAE of every interval of issued multi-horizon forecasts relative to a frozen reference, including intervals whose targets are not yet observed. For absolute loss, accounting jointly for the predictions that share a target makes each interval's feasible set a translated ℓ_1 ball—exactly the forecasts that keep the guarantee for every completion of the pending targets—and under regular horizon-H feedback the history compresses into two statistics and at most H active constraints. A constructed example separates this exact accounting from any independent certificate by a factor of order $\sqrt{H}$. Replays on temperature, Exchange Rate, multivariate transit, Weather and electricity, including modern online adapters, show that unguarded adaptation often *lowers* mean error while a single interval absorbs a regression that grows 9–43-fold with aggressiveness. The certificate removes it at a mean-MAE cost that is negligible or negative on temperature and material, but tunable through the allowance, on transit, Weather and electricity. Cheap independent guards match the exact guard in error: on the audited streams exact accounting accepts 1.4–$1.7\times$ more of each candidate's correction without changing realized error, so the separation, genuine in the worst case, is operationally inert. We contribute an exact characterization of what every-interval protection allows and a measured account of what it costs.

1 INTRODUCTION

An adaptive forecaster may improve total error while causing a costly recent regression: when electricity-load or transit-demand forecasts feed a dispatch schedule, a single interval in which an updated model severely under-forecasts can trigger a reserve shortfall or a missed dispatch that its better long-run average never offsets. A guard on cumulative excess loss is not enough—it can spend gains earned long ago during the current incident. We study a stronger objective: every contiguous interval of issued forecasts must obey a user-set excess mean-absolute-error (MAE) allowance relative to a frozen reference forecaster, with a fixed burst component and a rate proportional to interval length.

The obstacle is outstanding feedback. A twelve-step forecast issued now and an eleven-step forecast issued one step later predict the same observation. Independent worst-case charges ignore this identity; treating either loss as known spends evidence before it arrives. We account for each physical target jointly, without bounded outcomes, stationarity, or a calibrated uncertainty model. This bounds the eventual losses of intervals whose targets are still partly unobserved, rather than merely detecting harm afterwards.

We make three contributions. (i) *Geometry.* A prefix constraint is exactly a translated ℓ_1 ball—necessary and sufficient for every completion of the pending targets—and every-interval protection is an intersection of such balls, so any guard that keeps the guarantee, including the cheap independent one, must issue forecasts inside this set. (ii) *State and separation.* With regular H-step feedback, completed history compresses into two scalars and at most H constraints. An alternating-reference construction separates exact accounting from arbitrary independent certificates by a factor of order $\sqrt{H}$, which a simple horizon-specific independent policy attains. The bound concerns that certificate family, not arbitrary safe predictors, and on real archives it is operationally inert: exact ac-

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.

counting accepts more of each candidate with no effect on realized error—a null we measure rather than leave as a contradiction.

(iii) *Measurement.* Unguarded adaptation exposes a systematic failure: in 57 of 117 temperature record–allowance cells it *lowers* average MAE while a single interval absorbs an excess of up to $494\times$ its burst allowance (median $36\times$), so an averaged report calls it an improvement; a published modern adapter (ADAPT-Z), multivariate transit, and two online adapters on independent Weather and electricity domains show the same localized harm at their default hyperparameters. A single cumulative budget—the standard conservative-projection object—does not confer interval protection, violating the every-interval allowance on ETTh2 at all four learning rates, whereas the shared-target account bounds it by construction. The certificate removes the localized regression at a mean-MAE cost that is negligible or negative on temperature yet materially positive on transit, Weather and electricity.

2 Related work

Delayed test-time adaptation is well developed: ST-TTC calibrates spectrally with delayed updates (Chen & Liang, 2025b), SOLID adapts heads from contextual history (Chen et al., 2024), TAFAS uses partial ground truth and gated calibration (Kim et al., 2025), and FAC studies matured-only supervision (Wang et al., 2026); causal supervision is not our novelty. EAC and TFMoE give prompt/expert traffic mechanisms (Chen & Liang, 2025a; Lee & Park, 2024); PROCEED predicts drift-driven parameter adjustments (Zhao & Shen, 2025) and ADAPT-Z adjusts features from historical gradients (Huang et al., 2026). We use these as existing candidates and claim no new backbone or delay-aware adaptation mechanism; our official-class controls are not full-pipeline reproductions.

Baseline protection is likewise established—conservative bandits (Kazerouni et al., 2017), Conservative Projection for online convex optimization (Bernasconi de Luca et al., 2021), delayed adversarial baseline safety (Hu et al., 2026), and RCL's delayed-cost reservation-plus-projection under an expert guarantee (Li et al., 2023). Neither reservation nor projection is a new principle here; our setting differs in studying exposed reference/candidate predictions, shared future target identities, and raw absolute loss with unbounded outcomes, and we claim none of their regret guarantees. Every-interval objectives are equally classical: strongly adaptive online learning controls interval regret (Daniely et al., 2015; Cutkosky, 2020), the affine interval envelope with running-minimum representation is standard in network calculus (Le Boudec & Thiran, 2001), and ℓ_1-ball projection is textbook (Duchi et al., 2008). Our object is the exact feasible geometry and bounded-history compression of partially observed, shared-target MAE liabilities under a frozen-reference allowance—not the interval objective, an optimal interval-regret claim, or the projection primitive. Sequential TTA risk monitoring (Schirmer et al., 2025), the SAFE-TS gate (Zhang, 2026), and validation-only learning-rate audits (Fujimoto & Nishi, 2026) address different properties; only the SAFE-TS gate enters, as a common-candidate control.

3 Forecasting with partially observed losses

At origin t, the reference predicts $b_t \in \mathbb{R}^{H\times N}$ and a causal candidate predicts c_t. The guard issues g_t in a feasible set defined below. Our scalar controls restrict this choice to

$$g_t = b_t + \alpha_t(c_t - b_t), \qquad 0 \le \alpha_t \le 1. \tag{1}$$

Coordinate (h, n) predicts observation $y_{t+h,n}$. Outcomes become usable only at their actual observation times. Candidate training and guard decisions must respect this filtration; an evaluation script having future labels in memory is not permission to use them for decisions.

The per-origin loss is *all-entry* MAE,

$$L_t(g) = \frac{1}{HN}\sum_{h=1}^{H}\sum_{n=1}^{N} |g_{h,n} - y_{t+h,n}|. \tag{2}$$

Define excess $d_t = L_t(g_t) - L_t(b_t)$. For a predictable nondecreasing budget B_t, the desired certificate is

$$\forall t, \qquad \sum_{s\le t} d_s \le B_t. \tag{3}$$

This concerns eventual losses of every issued prefix, including currently unresolved targets. We use $B_t = B_0 + \rho t$, with training-only scale choices. It permits average excess at most $\rho + B_0/t$. It does not imply a multiplicative percentage guarantee relative to an unknown future baseline error.

Why the metric matters. The zero-masked MAE traditionally used in traffic benchmarking has a target-dependent denominator. Our theorem does not silently substitute that metric for Eq. (2); masked MAE is reported separately (Appendix J). For transit, zero counts are genuine observations, so excluding them can even reverse the ranking of correction methods. A certificate should name the operational loss it actually controls.

4 SHARED-TARGET RISK ACCOUNTING

For an unresolved target (τ, n), let $I_{\tau,n}$ index issued forecasts that predict it. Using the corresponding scalar reference and guarded predictions, define

$$r_{\tau,n} = \frac{1}{HN} \sup_{y \in \mathbb{R}} \sum_{i \in I_{\tau,n}} \left(|g_i - y| - |b_i - y| \right). \tag{4}$$

The independent triangle reserve instead uses $\widetilde{r}_{\tau,n} = (HN)^{-1} \sum_i |g_i - b_i|$. Both are label-free; the exact reserve exploits the shared target.

Proposition 1 (Finite evaluation and comparison)**.** *Let $F(y) = \sum_i (|g_i - y| - |b_i - y|)$. Its supremum equals*

$$\max \left\{ \max_j F(b_j), \left| \sum_i (g_i - b_i) \right| \right\}. \tag{5}$$

Moreover, $0 \le r_{\tau,n} \le \widetilde{r}_{\tau,n}$.

Proof. The continuous piecewise-linear function F has upward slope jumps at g_i and downward slope jumps at b_i. A finite local maximum therefore occurs at a reference breakpoint, or on a flat interval with the same value at its endpoints. At the two tails, F equals $\pm \sum_i (g_i - b_i)$. This proves the finite formula and nonnegativity. The triangle inequality bounds each summand by $|g_i - b_i|$. □

The inequality can be strict. With reference predictions $0, 2$ and issued predictions $1, 1$ for one shared target, the independent reserve is two whereas the exact reserve is zero: their summed absolute loss never increases for any target. This is an algebraic example, not evidence that forecasting data frequently produce such cancellation.

Let D_t be the sum of realized excess contributions from all observed coordinates, normalized consistently with Eq. (2). Let $R_t = \sum_{(\tau,n) \text{ unresolved}} r_{\tau,n}$. Before issuing a new prediction, evaluate the prospective reserve $R_t(\alpha)$ after adding its coordinates. Choose

$$\alpha_t = \max\{\alpha \in [0, 1] : D_t + R_t(\alpha) \le B_t\}. \tag{6}$$

If the candidate is fully feasible, accept it. Otherwise, use conservative bisection. When a label arrives, remove its entire shared-target bucket and add the corresponding realized excess to D_t.

Theorem 1 (Pathwise anytime excess-loss budget)**.** *Assume finite predictions, arbitrary real targets, correct target identities, and causal feedback. Starting with no issued predictions, a nonnegative nondecreasing budget, and decisions satisfying $D_t + R_t(g_t) \le B_t$, every issued prefix satisfies Eq.* (3)*, regardless of the candidate's training rule or the target sequence.*

Proof. Initially the certificate is zero. Revealing a target replaces a supremum by one of the values it upper-bounds; thus $D_t + R_t$ cannot increase upon feedback. Increasing the budget preserves feasibility. Issuing the reference ($\alpha = 0$) adds identically zero excess and is feasible. The feasibility rule preserves the invariant after every issuance. For any completion of unresolved outcomes, their realized total excess is at most R_t. The eventual excess of the issued prefix is therefore at most $D_t + R_t \le B_t$. □

Proposition 2 (Maximal feasible correction along a fixed direction)**.** *The prospective certificate is convex in α. Its feasible set is an interval containing zero. With existing predictions fixed, the exact reserve gives a necessary and sufficient condition for the current prefix's budget to hold for every assignment of unresolved real targets. Thus the exact solution of Eq.* (6) *is maximally permissive along the proposed direction under that requirement.*

Proof. For any fixed target assignment, the new absolute-loss terms are convex in α; taking a supremum preserves convexity. Target coordinates can be chosen independently under the stated unrestricted-outcome assumption, so the sum of coordinate suprema is the exact worst-case pending excess. Any larger coefficient outside the feasible set admits a violating completion. Bisection approximates the endpoint from its feasible side. □

What this does not prove. A greedy maximal step need not maximize long-run improvement: spending risk now can reduce flexibility later. Exact and independent guards may follow different trajectories, so a tighter reserve at a fixed state does not guarantee better empirical MAE across trajectories. A global cumulative budget can also spend earlier gains during a later bad period; it does not protect every sliding window. With zero initial budget and zero replenishment, a generic nonzero first correction cannot be accepted. These limitations should guide comparisons, not be concealed by the term "safe."

4.1 The complete feasible set

The scalar restriction is unnecessary. Let $F(y)$ denote the old excess function for one pending coordinate and b its new reference prediction. Define $A = \sup_y[F(y) - y - |b - y|]$ and $C = \sup_y[F(y) + y - |b - y|]$. Both are finite and computable from reference breakpoints and the constant tails.

Theorem 2 (Exact prediction-space geometry)**.** *For any new scalar prediction x, its updated unnormalized liability is*

$$\sup_y\{F(y) + |x - y| - |b - y|\} = |x - m| + k, \quad m = (C - A)/2, \quad k = (A + C)/2 \geq 0. \tag{7}$$

Collect these centers and constants for all HN new coordinates. If R_{other} is the unnormalized liability of untouched target buckets, the exact feasible set is

$$\mathcal{C}_t = \{x : \|x - m\|_1 \leq r_t\}, \qquad r_t = HN(B_t - D_t) - R_{\text{other}} - \sum_j k_j. \tag{8}$$

It is nonempty because it contains the reference. Euclidean projection of c_t onto $\mathcal{C}_t$ gives the unique minimum-change feasible forecast.

The proof is in Appendix A. Projection onto an ℓ_1 ball uses classical soft thresholding (Duchi et al., 2008); our object is the loss-derived center and radius. At the *same ledger state*, the independent reserve's feasible ball is a subset of $\mathcal{C}_t$ when nonempty. This guarantees no larger proposal distortion at that fixed state, not lower forecast error across different online trajectories. The exact projection can leave the coordinate-wise reference–candidate segment.

Implementation. Caching $F(b_j)$ and signed tails requires $O(H^2N)$ storage and work to prepare the geometry per origin; sorted projection costs $O(HN \log(HN))$. Once prepared, each scalar coefficient check costs $O(HN)$. The independent ray control has a closed-form coefficient, and independent vector projection uses a ball centered at b_t. We compare all four policies (exact or independent, ray or vector).

5 Protecting every issuance interval

The local requirement strengthens Eq. (3) to

$$\forall a \leq b, \qquad \sum_{i=a}^{b} d_i \leq B + \rho(b - a + 1), \quad B, \rho \geq 0. \tag{9}$$

It is equivalent to bounding by B the drawup of centered cumulative excess $Z_k = \sum_{i<k}(d_i - \rho)$ from its preceding minimum. This equivalence is classical; the difficulty is that recent d_i are only partially known. A prefix guard does not suffice: with $H = 1, B = 1, \rho = 0$, a first excess of -1 permits a second excess of 2 under the prefix constraint, violating the second singleton interval.

Conceptually retain a ledger for every interval start. Its realized excess plus shared-target liability is bounded by its corresponding allowance. Theorem 2 gives a feasible ball for each ledger; choose a forecast in their intersection. Feedback cannot increase any certificate, extending old intervals replenishes ρ, and the reference is feasible for the new singleton. Induction proves Eq. (9) for every ended interval, for any completion of its targets.

Theorem 3 (Bounded-delay compression)**.** *Assume unit-spaced issuance and complete, on-time H-step feedback. Before origin t, let origins $0, \dots, m-1$ be fully mature. All interval starts $a \le m$ have the same pending suffix and are dominated by the constraint with completed centered excess $Z_m - M_m$, where $M_m = \min_{0 \le k \le m} Z_k$. Together with the distinct recent starts, at most H constraints suffice. The exact next-forecast feasible set is their nonempty intersection of translated ℓ_1 balls. Storage is $O(H^2 N)$, independent of stream length.*

For a recent suffix j, let K_j sum its known, normalized contributions minus ρ per existing origin; add $Z_m - M_m$ only to the entire suffix. Its ball has center m_j from Theorem 2 and radius

$$r_j = HN(B + \rho - K_j) - \sum_q k_{j,q}. \tag{10}$$

All old pending targets lie in the new horizon, so no untouched liability remains. Appendix F proves compression and gives the independent control. The implemented exact policy takes the maximal feasible candidate-ray step using sixteen conservative bisections. A subsequent matched-objective study uses weighted Dykstra iterations with separately repaired feasible outputs and numerical primal–dual gaps (Appendix N); finite capped iterations are not assumed to be optimal. Preparing all balls costs $O(H^3 N)$ in our implementation; each ray check is $O(H^2 N)$. The independent control needs $O(HN)$ work and a closed-form ray step. The same compression extends to distinct integer issuance times with complete pending-target feedback, counting the rate per issued forecast and retaining the ledger across idle gaps (Appendix O).

Theorem 4 (Horizon separation for independent certificates)**.** *Let H be positive and even, $N = 1$, $B = 1$, and $\rho = 0$. At origin t, let every reference coordinate equal 0 for even t and 2 for odd t, and let the fixed causal candidate equal 1. The exact interval guard accepts the entire candidate for arbitrary real targets. On the constant target $y = 1$, its per-origin MAE improvement is 1, whereas any policy satisfying the independent every-interval certificate has long-run mean improvement at most $(\sqrt{4H+1} - 1)/H = O(H^{-1/2})$. A horizon-specific independent policy attains $\Theta(H^{-1/2})$. The scalar-ray subclass has the stricter bound $4/(H+2)$.*

For each target, alternating reference pairs obey $2|1-y| - |0-y| - |2-y| \le 0$. Any issuance interval has at most H unpaired terms, each with excess at most 1, so its normalized liability is at most 1. For the independent bound, canonicalize arbitrary predictions to coordinate gains $a_{t,h} \in [0, 1]$, improving their realized losses without increasing pending charges on target 1. A limit of time-averaged gains $\bar{a}_h$ must satisfy, for every $L \le H$,

$$\sum_{h=1}^{H} [2 \min(L, h) - L] \bar{a}_h \le H. \tag{11}$$

For fixed total mass $A = \sum_h \bar{a}_h$, assigning it to the earliest horizons minimizes every left-hand side. Choosing integer L nearest $(A+1)/2$ yields $A(A+2)/4 \le H$, proving the upper bound. A policy that corrects only its first $k \asymp 2\sqrt{H}$ horizons matches this order. Appendix H supplies the boundary and canonicalization arguments. This strengthens the scalar comparison to arbitrary real, time-varying predictions using the specified independent certificate (Figure 6, Appendix H); it does not establish an advantage over all safe algorithms, vanishes for co-monotone forecasts (Proposition 4), and being worst-case moves realized MAE by $<10^{-3}$.

6 Experiments: hidden localized harm and its price

Protocol and chronology. The first interval study replays 39 existing records: ETTh1 has two horizons $(24, 96)$, three seeds and four spectral learning rates; ETTh2 adds six validation-selected

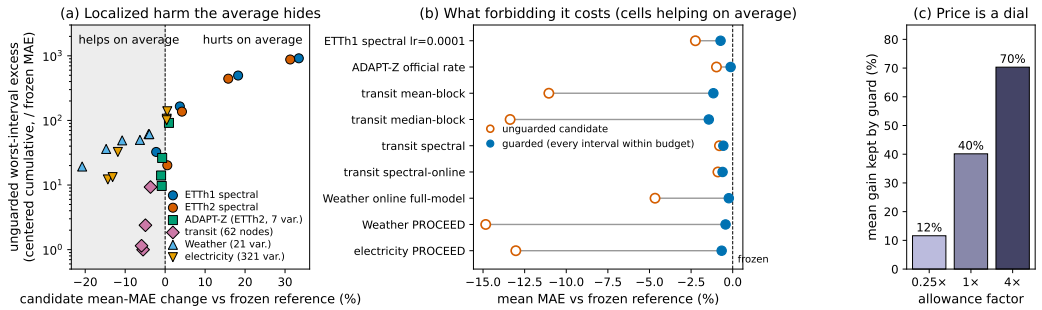


Figure 1: The empirical claim in one row, all from archived replays. (a) Every unguarded adaptive cell on five domains (ETTh1/ETTh2 online spectral calibration at four rates, ADAPT-Z at four online-rate scales, four transit adapters, Weather and electricity under full-model online updates and PROCEED; seed means): most sit left of zero—they *lower* mean MAE—while their worst issuance interval carries a centered cumulative excess of 1 to 9×10^2 frozen-MAE units. (b) For the cells that help on average, mean MAE of the unguarded candidate and of the guarded forecast relative to the frozen reference: the guard keeps every interval within its budget (§5) and retains 33% (ETTh1), 13% (ADAPT-Z), 10–71% (transit) and 3–5% (Weather, electricity) of the gain; harmful cells, where the guard also lowers mean MAE, are omitted. (c) On transit the retained fraction is set by the allowance: $12/40/70\%$ at $0.25\times/1\times/4\times$ the default burst and rate. Transit enters (a) from the original-unit study and (b,c) from the standardized study (Appendix E.3); normalization cancels units in (a).

DLinear fits; and seven-variable ETTh2 adds three iTransformer fits with three candidates (official ADAPT-Z class, head OGD, spectral calibration). This is exploratory reuse of inspected series. We then freeze a stronger-control and allowance study on all records, and separately freeze a new-domain protocol before fitting Exchange Rate (Lai et al., 2018). The latter adds three iTransformer fits with the same three candidates. We finally freeze a second pre-registered domain before any Weather forecast: nine iTransformer backbones (21 variables, three horizons $\times$ three seeds) with two contemporary online adapters (full-model updates and PROCEED (Zhao & Shen, 2025)), perturbation-checked for future-label leakage. An electricity study (321 variables, three horizons $\times$ three seeds, same adapters) adds eighteen records, giving 84 candidate records from 36 backbone fits, not independent datasets or temporal replications. The guard-value and cumulative-baseline analyses extend ETTh2 and ADAPT-Z to five seeds.

For DLinear, the original allowance is $\rho = 10^{-3}\sigma_{\text{train}}$, $B = H\rho$, on original-unit oil-temperature MAE. Multivariate iTransformer uses training-standardized all-entry MAE with $\rho = 10^{-3}$ and $B = H\rho$. We multiply both B and ρ by each of $\{0.25, 1, 4\}$, reporting every cell without selecting a test-optimal allowance. This varies one allowance scale, not burst and rate separately. The full carrier, partition, validation, and adaptation schedules are in Appendices J, D, E, M, E.1, and E.2.

Guards compared. Besides the exact and independent scalar-ray guards, we compare two classical independent vector controls—Euclidean and fixed near-horizon ($w_h = 1/h$) weighted projection onto the same independent ball (Appendix L). They allocate a common budget across coordinates rather than shrinking the whole proposal; none is a proposed forecaster. The exact interval policy in this comparison is a scalar ray; the matched-state study of Appendix N removes that decision-space mismatch. The study has $48 \times 3 \times 4 = 576$ guarded trajectories; every interval is audited offline, decisions use only mature observations, and all 35 original evaluation checks, including future-perturbation checks for every candidate family, pass.

Local harm differs from global harm. At the original allowance, all 39 temperature candidate records and both sets of prefix-guarded records violate the stronger interval constraint. This does not contradict their prefix guarantee. All 576 interval-guarded trajectories satisfy their respective allowances at tolerance 10^{-7}. Across the three allowances, unprotected candidates violate 113/117 temperature cells and 19/27 Exchange cells. The zero violations of guards do not imply zero deterioration: their allowances permit positive excess.

How large is the unguarded harm, and does the certificate remove it? The violation counts above establish that unguarded harm occurs; its magnitude decides whether the certificate is worth its correction. We therefore measure, for each candidate, the worst-interval centered cumulative excess against the frozen reference (the peak draw-up of Figure 3), aggregated over seeds and allowances (Figures 1a and 2). For the online spectral-calibration candidate this harm grows monotonically with the adaptation rate: on ETTh1 (three seeds $\times$ three allowance factors $\times$ two horizons) from 49.6 at 10^{-4} to 1.40×10^3 at 10^{-1}, and on ETTh2 (five seeds) from 55.4 to 2.40×10^3 (original DLinear units), robustly across seeds. The deterioration is easy to miss on averages: in 57 of 117 cells the unguarded candidate simultaneously *lowers* overall MAE, so a mean-error report would call it an improvement while a single interval absorbs a large sustained regression. The cheapest guard, the independent ray, caps this excess at the allowance by construction (§5); its median reduction of the worst-interval excess is $180\times$ (minimum $30\times$), and its overall MAE is a median of -2.4% against the same unguarded candidate—lower, because the excursions that inflate the interval excess also inflate the mean—rising to $+1.5\%$ only at the smallest learning rate. The effect is not an artifact of a naive candidate: the official ADAPT-Z online adapter (Huang et al., 2026), run under our common carrier, shows the same pattern (five seeds), its worst-interval excess growing from 3.08 to 29.2 (standardized units) as its online rate is scaled $1\times$ to $30\times$ while its mean MAE stays at or below the frozen reference until the most aggressive setting.

A structurally different third domain sharpens both the phenomenon and its cost. On Beijing multivariate transit (62 nodes, delayed feedback), four candidates *lower* mean MAE by 3.7–5.9% yet inflict worst-interval excess of 17–160 (original units; Figure 1b reports the standardized study); here protection is not free—removing it returns 3.1–4.8% of mean MAE. Two further independent domains under contemporary adapters reproduce phenomenon and price: on Weather (21 variables) and 321-variable electricity, PROCEED (Zhao & Shen, 2025) lowers mean MAE by 0.04–0.11 standardized units (full-model online updates by 0.02–0.04 on Weather) while inflicting worst-interval excess (4–46) the certificate removes (residual violation $\leq 10^{-14}$) while retaining only 3–6% of the gain; on electricity the full-model online adapter is slightly harmful on average, yet its larger excess (32–53) is bounded just the same (Appendices E.1, E.2). Across all five domains, adaptation—useful on average or not—can hide a large localized regression the certificate provably bounds, nearly free on temperature but tunably priced on transit, Weather and electricity (Figure 1b).

A cumulative-budget baseline does not protect intervals. The natural comparator is a conservative projection under a single global budget (delayed-cost reservation plus scaling toward the reference to keep cumulative excess within $H\rho + \rho t$), the object underlying conservative bandits, Conservative Projection and RCL (Kazerouni et al., 2017; Bernasconi de Luca et al., 2021; Li et al., 2023). Run on the same ETTh2 candidate (five seeds, two horizons), it keeps its cumulative budget yet violates the every-interval allowance at every rate—its worst-interval excess exceeds the allowance by 12–24 (original units, seed mean), spending accumulated slack in bursts—whereas our certificate satisfies it (violation $\leq 10^{-14}$) at essentially equal mean MAE (2.740 vs. 2.738). A global cumulative budget therefore confers no interval guarantee; the every-interval account supplies it, here at no average-accuracy cost.

Node-wise certificates close the cross-node loophole. The certificates above bound the interval excess *averaged over stations*, so one station can absorb a large regression while the average stays within its allowance: per node, the aggregate-guarded transit forecasts leave a worst-station violation of 14–117 units. Protecting each station independently (a product of the same translated-ℓ_1 balls, so the geometry of §5 extends factor-wise) drives every station's violation to $\leq 10^{-14}$ with mean MAE within 0.12% of the aggregate guard's; the same holds on ADAPT-Z (7 variables, worst $0.28 \to 0$), Weather (21, 0.4–$2.4 \to 0$), and 321-variable electricity, where the loophole is far larger (up to $54 \to \leq 10^{-6}$), growing with dimensionality and horizon—the node-wise certificate closes it regardless. This stronger guarantee thus costs little mean MAE, though the retained-gain fraction falls by up to eight points; it recovers no extra gain, since harmful and useful signal are co-located within a coordinate—a limit of the frozen reference, not of the node-wise account.

Does exact accounting pay on real data? No—though its room is used. Theorem 4 lets exact accounting accept what independent certificates must refuse, yet on the archives the two accountings are indistinguishable in error: independent Euclidean projection has lower mean MAE than the exact ray in 11 of the 13 rows of Table 1, and matched exact and independent guards differ in

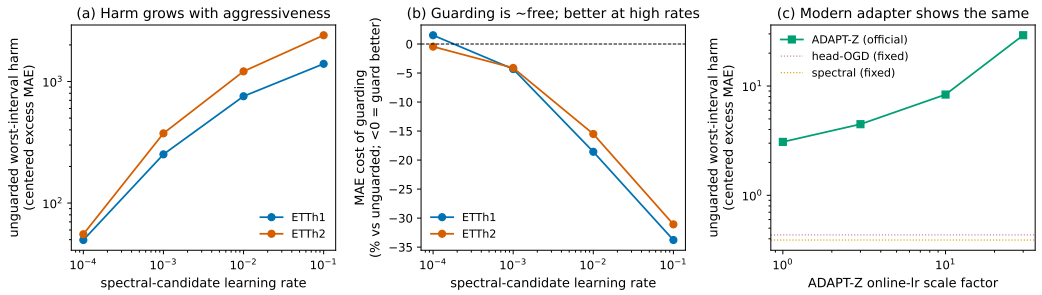


Figure 2: Unguarded localized harm and the cost of removing it. (a) Worst-interval centered cumulative excess of the online spectral-calibration candidate against the frozen reference grows with the adaptation rate on ETTh1 and ETTh2 (log–log). (b) The overall-MAE cost of the independent-ray guard relative to the same unguarded candidate; below zero the guard also lowers mean error. (c) The official ADAPT-Z online adapter exhibits the same harm growth as its online rate is scaled (log–log); fixed-rate head-OGD and spectral candidates are shown as references. ETTh1 over three seeds, three allowance factors and two horizons; ETTh2 and ADAPT-Z over five seeds; descriptive means with substantial cross-seed dispersion, no significance claim. DLinear (a,b) and standardized iTransformer (c) units are not comparable across panels.

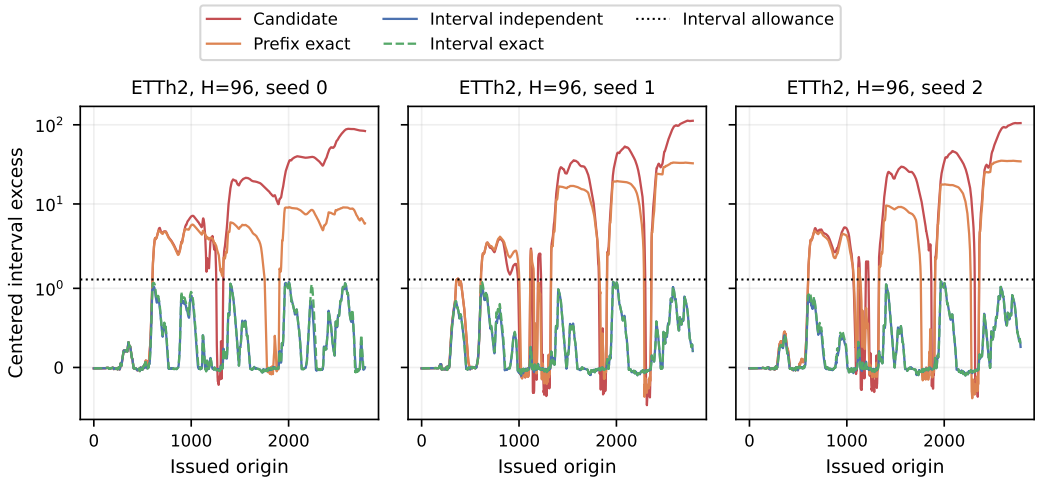


Figure 3: Original-allowance ETTh2 $H = 96$ paths, all three seeds. Centered cumulative excess is measured from its best preceding prefix, audited offline. A valid prefix guard can exceed the local allowance. The symmetric-log axis is linear within the allowance and logarithmic above it.

realized MAE by under 10^{-3} on ETTh2 and at most 4.2×10^{-4} on ADAPT-Z. The extra room is not idle. Natural forecasts are not co-monotone (per-target zero-gap fraction 0.88 on transit but 0.01 on ADAPT-Z, whose worst-case liability gap averages $0.36\, L_{\mathrm{ind}}$); both guards modify the candidate on at least 86% of issuances; exact accounting retains 1.4–$1.7\times$ more of each candidate's correction (17–32% vs. 10–20%); and natural targets reach the exact worst-case liability on 12–40% of targets (Proposition 4, Appendix H). The separation is exercised, but the extra accepted correction is error-neutral at the observed operating point, for reasons we leave open: the separation bounds what exact accounting can permit, not what it buys in accuracy.

Allowance curves. Larger allowances preserve more of the official-rate ETTh1 candidate's gain but can worsen error for a harmful candidate or at particular horizons (Figure 8, Appendix L); the original allowance is not promoted as optimal. On held-out Exchange Rate the exact and near-horizon projections trade places across allowances with no stable winner (all nine groups in Table 12). The same $\{0.25, 1, 4\}$ sweep on transit traces the cost–gain tradeoff: the aggregate guard

Setting	Frozen	Candidate	Indep. ray	Exact ray	Indep. vec.	Near vec.
ETTh1 24, 0.0001	1.21653	1.19530	1.21145	1.21144	1.20946	1.21057
ETTh1 24, 0.001	1.21653	1.19878	1.21051	1.20854	1.20748	1.20838
ETTh1 24, 0.01	1.21653	1.42197	1.21466	1.21354	1.21112	1.21889
ETTh1 24, 0.1	1.21653	1.76881	1.21500	1.21407	1.21253	1.22140
ETTh1 96, 0.0001	1.83697	1.78975	1.82681	1.82681	1.82630	1.83613
ETTh1 96, 0.001	1.83697	1.96705	1.83017	1.82975	1.82974	1.83949
ETTh1 96, 0.01	1.83697	2.18909	1.83188	1.83251	1.83057	1.84385
ETTh1 96, 0.1	1.83697	2.30598	1.83116	1.83211	1.83090	1.84479
ETTh2 24	2.22769	2.21860	2.22499	2.22494	2.22403	2.22193
ETTh2 96	3.24163	3.27704	3.24267	3.24266	3.24256	3.24345
iT-ETTh2 ADAPT-Z	0.31846	0.31536	0.31807	0.31790	0.31800	0.31764
iT-ETTh2 Head OGD	0.31846	0.31704	0.31795	0.31771	0.31776	0.31763
iT-ETTh2 Spectral	0.31846	0.31787	0.31813	0.31813	0.31811	0.31810

Table 1: Three-seed mean MAE at the original allowance; every guarded column tracks the frozen reference, keeping part of the candidate's gain where it helps and suppressing nearly all its harm where it hurts. DLinear rows use original target units; iT-ETTh2 rows use training-standardized units and are not comparable across carriers. Near vec. uses fixed $1/h$ weights.

retains $12\%/40\%/70\%$ of mean gain as the allowance grows (Figure 1c, Appendix E.3), so the price of protection is set by a tunable allowance, not fixed.

Cost and solver choice. The independent vector guard takes 0.27 ms per decision on transit histories. Two exact-set vector solvers each have lower seed-mean MAE than the matching independent vector guard in 14 of 16 settings; a hybrid solver raises tolerance attainment from 72.9% to 96.5% yet lowers MAE further in only $9/16$—numerical completion does not consistently improve accuracy (Appendix N).

7 Limitations

Our characterization assumes finite predictions, correct target identities, and the stated all-entry loss; bounded-state compression further requires complete pending-target feedback. Replays use dense issuance (a tested extension permits distinct integer issuance times). Missing labels, multiple issuances per tick, variable horizons with out-of-order completion, and target-dependent masked denominators are not covered. The node-wise certificate closes the aggregation gap at no accuracy cost but recovers no extra adaptation, and neither is a statement of downstream operational safety.

The separation applies to independent interval certificates, not all safe algorithms, and no regret optimality is established. The two temperature series are related, and the modern-adapter studies are official-class controls rather than full reproductions. More faithful comparators, end-to-end costs and a justification for the chosen allowances remain necessary; current results do not support a state-of-the-art forecasting claim.

8 Conclusion

Shared future observations pin down the exact feasible geometry of delayed absolute-loss protection, which regular feedback compresses to bounded state; a scoped construction shows when exact accounting avoids a horizon-dependent acceptance loss—room natural streams use without changing realized error. Across five domains and two modern adapters, adaptation that helps on average can still inflict a large localized regression the certificate provably removes—at a cost an allowance sweep shows to be tunable. Which guarded action to issue, and when it repays its cost in accuracy, remain open.

AI USE STATEMENT

In our work, we mainly use GPT-5 for writing enhancements, primarily to improve grammar and text clarity.

REPRODUCIBILITY STATEMENT

Proofs are in Appendices A, F, G and H; protocols, seeds and allowances are in Appendices D, E, J and M. The artifact records the frozen experiment settings, the official third-party sources with their licenses and Git hashes, prediction archives, decision paths, per-prefix certificates, source/data hashes and tests; every table is generated from completed runs. The Beijing transit data cannot be redistributed and are available only under the data owner's access conditions.

ETHICS STATEMENT

Inputs are aggregated sensor and station counts. Raw AFC records are not included in a public release, and redistribution requires appropriate rights and privacy review. The every-interval certificate is a mathematical forecasting property; it is not a certification of downstream operational safety or equitable performance at every station.

REFERENCES

Martino Bernasconi de Luca, Edoardo Vittori, Francesco Trovò, and Marcello Restelli. Conservative online convex optimization. In *ECML PKDD*, 2021. URL https://ecmlpkdd-storage.s3.eu-central-1.amazonaws.com/preprints/2021/sub_12975_17.pdf.

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 *KDD*, 2024. URL https://arxiv.org/abs/2310.14838.

Wei Chen and Yuxuan Liang. Expand and compress: Exploring tuning principles for continual spatio-temporal graph forecasting. In *ICLR*, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/cb2266111eadcfa2c02187ace64e2183-Abstract-Conference.html.

Wei Chen and Yuxuan Liang. Learning with calibration: Exploring test-time computing of spatio-temporal forecasting. In *NeurIPS*, 2025b. URL https://arxiv.org/abs/2506.00635.

Ashok Cutkosky. Parameter-free, dynamic, and strongly-adaptive online learning. In *ICML*, volume 119, pp. 2250–2259, 2020. URL https://proceedings.mlr.press/v119/cutkosky20a.html.

Amit Daniely, Alon Gonen, and Shai Shalev-Shwartz. Strongly adaptive online learning. In *ICML*, volume 37, pp. 1405–1411, 2015. URL https://proceedings.mlr.press/v37/daniely15.html.

John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *ICML*, pp. 272–279, 2008. URL https://stanford.edu/~jduchi/projects/DuchiShSiCh08.pdf.

Takumi Fujimoto and Hiroaki Nishi. When does online adaptation pay on the edge? A leakage-free evaluation of warmup, learning-rate selection, and resource trade-offs for time-series forecasting. *arXiv:2609.01126*, 2026. URL https://arxiv.org/abs/2609.01126.

Ting Hu, Luanda Cai, and Emmanouil-Vasileios Vlatakis-Gkaragkounis. Prudent-Banker: No extra fees for baseline safety in adversarial bandits with and without delays. *arXiv:2605.23351*, 2026.

Xiannan Huang, Shuhan Qiu, Jiayuan Du, and Chao Yang. Online time series prediction using feature adjustment. In *ICLR*, 2026. URL https://arxiv.org/abs/2509.03810v2.

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi-Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. In *NeurIPS*, 2017. URL https://papers.nips.cc/paper_files/paper/2017/hash/bdc4626aa1d1df8e14d80d345b2a442d-Abstract.html.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *AAAI*, 2025. URL https://arxiv.org/abs/2501.04970.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *SIGIR*, 2018. URL https://arxiv.org/abs/1703.07015v3.

Jean-Yves Le Boudec and Patrick Thiran. *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*, volume 2050 of *Lecture Notes in Computer Science*. Springer, 2001. URL https://leboudec.github.io/netcal/.

Sanghyun Lee and Chanyoung Park. Continual traffic forecasting via mixture of experts. *arXiv:2406.03140*, 2024.

Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren. Robust learning for smoothed online convex optimization with feedback delay. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/36848567d39a5128e671ad04a6075374-Abstract-Conference.html.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. *arXiv:2310.06625v4*, 2024. URL https://arxiv.org/abs/2310.06625v4.

Mona Schirmer, Metod Jazbec, Christian A. Naesseth, and Eric Nalisnick. Monitoring risks in test-time adaptation. In *NeurIPS*, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/746960ad49ddb47248970a0e1404230c-Paper-Conference.pdf.

Ryan J. Tibshirani. Dykstra's algorithm, ADMM, and coordinate descent: Connections, insights, and extensions. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/5ef698cd9fe650923ea331c15af3b160-Abstract.html.

Haochun Wang, Ruichen Xu, Georgios Kementzidis, Karen Cho, Sebastian Ramirez Villarreal, and Yuefan Deng. Towards principled test-time adaptation for time series forecasting. *arXiv:2605.17250*, 2026. URL https://arxiv.org/abs/2605.17250.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *AAAI*, 2023. URL https://arxiv.org/abs/2205.13504.

Yuchen Zhang. SAFE-TS: Causal selective residual adaptation with a frozen fallback. Research software repository, 2026. URL https://github.com/litlelight/SAFE-TS. Commit 4b678bc, accessed 18 September 2026; not represented as a peer-reviewed publication.

Lifan Zhao and Yanyan Shen. Proactive model adaptation against concept drift for online time series forecasting. In *KDD*, 2025. URL https://arxiv.org/abs/2412.08435v5.

Haoyi Zhou. Electricity transformer temperature dataset. Official dataset repository, 2020. URL https://github.com/zhouhaoyi/ETDataset.

Appendix roadmap. The appendices fall into four groups. **Proofs:** the exact prediction-space geometry (A), bounded-history compression (F) and its irregular-issuance extension (O), the horizon separation (G) with the co-monotone collapse proposition and exact-vs-independent operational-divergence measurements (H). **Protocols and data provenance:** frozen protocols and source hashes for the DLinear (J, D), stronger-control including Weather and electricity (E, E.1, E.2), allowance-sweep (E.3) and Exchange (M) studies. **Complete results:** every interval-replay cell (I), the allowance-curve tables (L) and the earlier ETTh1 prefix grid with the SAFE-TS gate (K). **Solvers and cost:** matched-state exact projection, complete solver trajectories and guard latency (N). Earlier exploratory studies (proposal selection, prefix-only budgets, an adapter-architecture matrix) and the full numerical audits are archived in the artifact rather than reproduced here.

Notation. For one physical target with horizon index h:

H	forecast horizon	N	nodes / variables
y	realized target	b_h	frozen reference forecast
g_h	issued (guarded) forecast	ρ	per-step excess-MAE rate
B	burst allowance ($B{=}H\rho$ default)	L_{ex}	exact shared-target liability
L_{ind}	independent marginal liability		

An interval of length ℓ is allowed centered excess MAE $B + \rho\ell$ relative to the reference b; the exact account treats forecasts sharing a target jointly, the independent account marginally.

A Proof of the exact prediction-space geometry

Since $|x - y| = \max(x - y, y - x)$, exchanging the supremum and a finite maximum gives

$$\sup_y [F(y) + |x - y| - |b - y|] = \max(x + A, -x + C) = |x - (C - A)/2| + (A + C)/2.$$

Let $S = \sum_i (g_i - b_i)$ for the old forecasts. The left tail of the expression defining A is $S - b$, and its right tail tends to minus infinity. The expression defining C has right tail $b - S$ and tends to minus infinity on the left. Finite maxima can occur at old reference breakpoints or at b, the only downward slope jumps. Thus both constants are finite and available without a target value. Their tail lower bounds imply $A + C \geq 0$.

Summing this equality over new coordinates, then adding realized excess and untouched liabilities, proves Eq. (8). Under the requirement of robustness to every completion, unresolved target coordinates may take arbitrary values: this is a quantifier over assignments, not a stochastic independence assumption. Therefore this sum is the exact worst-case excess. Issuing b_t adds zero excess, so the previous invariant makes b_t feasible and $r_t \geq 0$.

Strict convexity of $\frac{1}{2}\|x - c_t\|_2^2$ gives a unique projection. The classical ℓ_1-ball solution is

$$g_t = m + \mathrm{sign}(c_t - m) \max\{|c_t - m| - \lambda, 0\},$$

coordinate-wise, with $\lambda = 0$ for a feasible candidate and otherwise chosen so the projected ℓ_1 norm equals r_t. Sorting the absolute coordinates obtains this threshold (Duchi et al., 2008). Substituting any point of this ball in the ledger invariant proves its anytime certificate. At a common old state, the triangle reserve bounds the exact reserve for every proposed x, so its nonempty feasible set is contained in the exact ball. Euclidean distance to the candidate is thus no greater for the exact projection at that state. This conclusion does not compare ledgers generated by different earlier decisions.

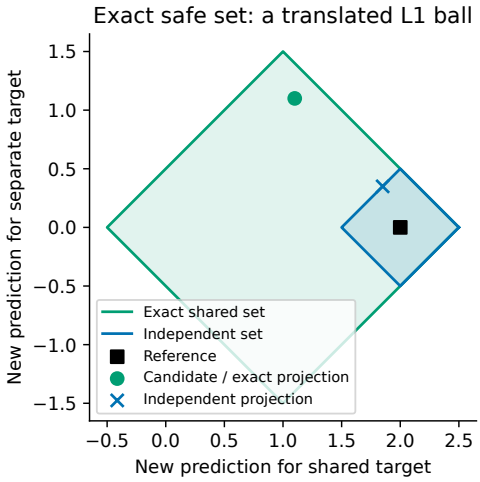


Figure 4: Algebraic illustration, not an empirical result. One old prediction has reference 0 and issued value 1 for the first target. The new reference is $(2, 0)$ and the available unnormalized budget is 1.5. The exact feasible ball is centered at $(1, 0)$ with radius 1.5; the independent ball is centered at $(2, 0)$ with radius 0.5. Candidate $(1.1, 1.1)$ is feasible only in the exact set. These sets share the same old ledger.

B ADDITIONAL THEORY BOUNDARIES

Proposition 3 (Zero budget precludes the first deviation)**.** *Starting with no issued predictions and* $B_t = 0$ *for every* t*, any policy satisfying Eq.* (3) *for every real target sequence must issue the reference at every time.*

Proof. Suppose there is a first deviation. All earlier excess functions are zero. For each coordinate of the new prediction choose its target to attain $|g - b|$, the supremum of its absolute-loss difference. The summed prefix excess is strictly positive, contradicting its zero budget. □

Thus a nontrivial pathwise guarantee requires an allowance or a different assumption; universal cost-free improvement is unavailable here.

For fixed quantile $q \in (0, 1)$, pinball loss satisfies $\rho_q(y - g) = \frac{1}{2}|y - g| + (q - \frac{1}{2})(y - g)$. A bucket's exact excess supremum is therefore $\frac{1}{2} \sup_y F(y) + (\frac{1}{2} - q) \sum_i (g_i - b_i)$, and the same ledger proof applies with a fixed normalization. This extension has not been evaluated empirically. Squared loss is different: its bucket excess equals $\sum_i (g_i^2 - b_i^2) - 2y \sum_i (g_i - b_i)$, whose supremum is infinite unless the signed correction sum is zero. The MAE result cannot simply be relabeled as an MSE guarantee.

C ALGORITHM

Algorithm 1 Shared-target risk guard: vector rule and scalar control

Require: Immutable predictor, causal candidate, horizon H, budget B_t
Initialize target buckets, realized excess $D = 0$
for each forecast origin t **do**
  for each newly observed target (τ, n) **do**
    Replace its reserve by actual excess in D; delete its pending bucket
  end for
  Obtain reference b_t and candidate c_t without future labels
  Compute m, k and radius r_t using Theorem 2
  if vector projection policy **then**
    $g_t \leftarrow m + \mathrm{Proj}_{\|\cdot\|_1 \le r_t}(c_t - m)$
  else
    Find largest feasible $\alpha_t \in [0, 1]$ along $b_t + \alpha_t(c_t - b_t)$
    $g_t \leftarrow b_t + \alpha_t(c_t - b_t)$
  end if
  Issue g_t and add its target contributions
  Record predictable certificate $D + R(g_t) \le B_t$
end for

The scalar implementation uses sixteen bisection iterations when full acceptance is infeasible. The offline auditor can inspect every eventual loss, but those labels are not inputs to the issuing rule. The certificate is asserted with numerical tolerances and outputs are retained in double precision for the risk-control replay.

D VALIDATION-SELECTED ETTH2 REPLICATION

The ETTh2 protocol was fixed before its test outcomes were inspected. Six backbone fits (three seeds at each of two horizons) choose their checkpoints by validation MAE. Separate causal validation calibration selects a learning rate by unguarded MAE; all choose 10^{-4}. Thus fixed and validation-selected policies coincide. A single spectral candidate is shared by all protectors within a fit. No test trajectory selects the budget, candidate rate, checkpoint, or reported subset of cells. For $H = 24$ the guards do not bind; for $H = 96$ the validation-selected rate is naturally harmful (unprotected MAE rises from 3.2416 to 3.2770) and both protectors reduce it to 3.2530 while still worsening average MAE relative to freezing, within the allowed budget. Table 2 and Figure 5 give all seeds' summaries and prefix paths.

	$H = 24$		$H = 96$	
Method	MAE $\pm$ seed SD	Violation	MAE $\pm$ seed SD	Violation
Frozen	2.2277 ± 0.0195	0.00	3.2416 ± 0.0003	0.00
Unguarded	2.2186 ± 0.0175	0.00	3.2770 ± 0.0033	77.70
Hedge	2.2205 ± 0.0163	0.00	3.2556 ± 0.0023	15.38
Raw + indep. ray	2.2186 ± 0.0175	0.00	3.2530 ± 0.0008	0.00
Raw + exact ray	2.2186 ± 0.0175	0.00	3.2530 ± 0.0009	0.00
Raw + indep. vector	2.2186 ± 0.0175	0.00	3.2531 ± 0.0009	0.00
Raw + exact vector	2.2186 ± 0.0175	0.00	3.2530 ± 0.0011	0.00
Hedge + indep. ray	2.2205 ± 0.0163	0.00	3.2529 ± 0.0008	0.00
Hedge + exact vector	2.2205 ± 0.0163	0.00	3.2529 ± 0.0009	0.00

Table 2: ETTh2 all-entry MAE, mean and initialization SD of three seeds. Violation is the maximum positive prefix-budget violation across those seeds in summed MAE units. All guards satisfy the constraint, yet can have higher MAE than freezing because the budget permits positive excess. The two omitted Hedge-projection variants are also retained in the artifact.

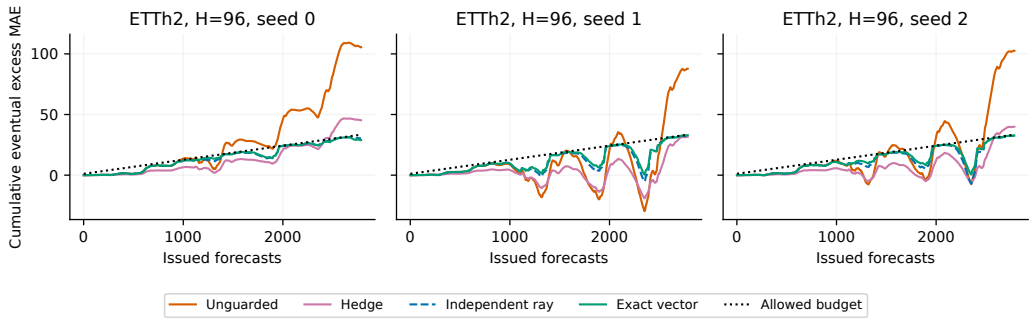


Figure 5: All three ETTh2 $H = 96$ seed trajectories at the validation-selected rate. Natural updates exceed the budget without injected failure. Both protectors respect the allowance but still have positive cumulative excess. Their curves nearly overlap; additional permissiveness does not produce a material improvement on this series. Curves are audited offline.

E STRONGER CARRIER AND OFFICIAL FEATURE-ADAPTATION CLASS

We froze a further exploratory protocol after the DLinear studies. It uses all seven ETTh2 variables, input 96, horizon 48, three seeds, and a common iTransformer (Liu et al., 2024) with two layers, width/feed-forward 128, eight heads, GELU, and zero dropout. The upstream 60/10/30% chronological partitions use windows entirely inside each partition. Training uses Adam 10^{-3}, batch 32, at most ten epochs, patience three, and the lowest validation-MSE checkpoint. Normalization uses training-only column moments.

We extract the official ADAPT-Z class at commit `f5b70e1`; only its final return is changed to expose already-computed forecasts for auditing. Three validation warmup passes, test batch 24, historical feature-gradient accumulation, adapter rate 3×10^{-4}, and head rate 3×10^{-5} follow that class. Frozen warm-adapter predictions separate initialization from online adaptation. Head-only OGD and spectral calibration also receive three validation passes; their test updates use the same fully mature batch schedule. No candidate uses guard decisions as training labels.

There are explicit reproduction boundaries. The harness aliases an upstream directory spelling mismatch and loads audited configuration directly. It uses the same internal normalization during base training and adaptation, where the upstream entry points set different defaults. Thus this is an official-class common-protocol control, not the authors' full benchmark. Tests verify equivalence of the ordinary and zero-perturbation feature models, batch independence, and unchanged prediction prefixes when unobserved future observations are perturbed. ETTh2 was already inspected, so it is not a new untouched-domain claim. Per-candidate outcomes are in Table 3.

Proposal	Unguarded	Violation	Indep. ray	Exact ray	Indep. vec.	Exact vec.
Frozen	0.31846	0.00	–	–	–	–
Warm adapter	0.31716	2.81	–	–	–	–
ADAPT-Z class	0.31536	0.78	0.31582	0.31579	0.31586	0.31582
Head OGD	0.31704	0.00	0.31705	0.31705	0.31705	0.31705
Spectral	0.31787	0.00	0.31787	0.31787	0.31787	0.31787

Table 3: Mean training-standardized MAE over three iTransformer fits. Violation is the largest unguarded positive prefix-budget violation across seeds. All four protectors have zero checked violation. The budget uses $\rho = 10^{-3}$, $B_0 = 48\rho$ in fixed training-standardized units; it does not certify unweighted MAE in the seven original units. Warm adapter freezes its validation-trained parameters and uses zero historical gradient.

The feature-adaptation candidate improves mean standardized MAE, yet incurs a small prefix violation. Protection retains most of that improvement. Exact-ray MAE is 0.31579 versus 0.31582 for the independent ray, while exact vector projection gives 0.31582. This small difference across seeds

sharing one series does not establish a material practical advantage. All three candidate families and all 36 guarded trajectories are retained, including the largely nonbinding head and spectral controls.

E.1 WEATHER INDEPENDENT-DOMAIN RESULTS

Complete per-cell outcomes for the pre-registered Weather study (21 variables, iTransformer backbone, three horizons $\times$ three seeds, two contemporary online adapters: full-model online gradient and PROCEED), standardized all-entry MAE, $\rho = 10^{-3}$, $B = H\rho$, no tuning. "Unguarded harm" is the worst-interval centered excess MAE of the adaptive candidate; "agg. viol." the aggregate guard's worst-interval violation; "agg. var. viol." the aggregate account's worst per-variable violation, driven to the "node viol." column by the node-wise certificate. Both adapters lower mean MAE while inflicting large localized harm the aggregate certificate removes while retaining only 3–6% of the mean gain (Weather is a priced domain, like transit); the node-wise certificate additionally closes the cross-variable loophole with mean MAE within 0.11% of the aggregate guard's. Future-perturbation probes are bit-identical (no label leakage). Every cell is listed in Table 4.

H	seed	adapter	frozen MAE	cand. MAE	unguarded harm	agg. MAE	agg. viol.	agg. var. viol.	node viol.
24	2021	Online	0.4761	0.4494	19.45	0.4746	0e+00	0.787	0e+00
24	2021	Proceed	0.4761	0.3816	8.50	0.4733	0e+00	0.385	0e+00
24	2022	Online	0.4771	0.4392	25.94	0.4754	0e+00	0.775	0e+00
24	2022	Proceed	0.4771	0.3697	7.86	0.4741	0e+00	0.368	0e+00
24	2023	Online	0.4732	0.4478	25.86	0.4717	0e+00	1.037	0e+00
24	2023	Proceed	0.4732	0.3788	11.21	0.4703	0e+00	0.529	0e+00
48	2021	Online	0.5781	0.5555	37.07	0.5768	0e+00	1.036	0e+00
48	2021	Proceed	0.5781	0.4900	18.40	0.5756	0e+00	0.610	0e+00
48	2022	Online	0.5786	0.5537	37.23	0.5773	0e+00	1.118	0e+00
48	2022	Proceed	0.5786	0.4937	24.59	0.5762	0e+00	0.709	0e+00
48	2023	Online	0.5790	0.5544	32.19	0.5778	0e+00	1.172	0e+00
48	2023	Proceed	0.5790	0.4959	19.78	0.5766	0e+00	0.713	0e+00
96	2021	Online	0.6798	0.6519	38.82	0.6785	0e+00	1.757	0e+00
96	2021	Proceed	0.6798	0.6045	22.63	0.6776	0e+00	0.910	0e+00
96	2022	Online	0.6734	0.6496	39.29	0.6721	0e+00	1.383	0e+00
96	2022	Proceed	0.6734	0.6059	37.30	0.6713	0e+00	1.200	0e+00
96	2023	Online	0.6799	0.6512	46.45	0.6787	0e+00	2.438	0e+00
96	2023	Proceed	0.6799	0.6041	39.87	0.6778	0e+00	1.035	0e+00

Table 4: All 18 Weather adaptive-candidate cells (Online full-model gradient, PROCEED). Standardized units; violations $\leq 10^{-14}$ shown as `0e+00`. Descriptive per-cell values, no significance claim.

E.2 ELECTRICITY INDEPENDENT-DOMAIN RESULTS

A second independent domain: electricity (321 variables), three horizons $(24, 48, 96) \times$ three seeds under the same two adapters and standardized allowance, evaluated inline (guard metrics computed from in-memory forecasts, so no multi-GB archive is stored). PROCEED lowers mean MAE (by 0.04–0.05 at every horizon) with its localized excess removed while retaining about 5% of its gain (a priced domain), whereas the full-model online adapter is slightly harmful on average yet inflicts a much larger excess (32–53) that the certificate removes just the same—at, here, a small *improvement* in mean MAE, since a harmful candidate inflates both the interval excess and the mean. The aggregate account's worst per-variable violation is far larger than on the lower-dimensional domains and grows with horizon (up to $36/48/54$ at H=$24/48/96$, versus ≤ 2.4 on Weather); the node-wise certificate drives it to $\leq 10^{-6}$ at unchanged mean MAE, so the cross-variable loophole grows with both dimensionality and horizon while its closure stays free. The residual is $\leq 10^{-6}$ here versus $\leq 10^{-14}$ on the lower-dimensional domains because it is floating-point accumulation over 321 per-node solves, not a weakening of the guarantee; the exact-rational feasibility audit is unchanged. (The $\leq 10^{-14}$ at Eq.-level in §6 refers to the aggregate guard's interval violation, a distinct quantity from this per-variable node-wise residual.) Table 5 lists every cell.

H	seed	adapter	frozen MAE	cand. MAE	unguarded harm	agg. MAE	agg. var. viol.	node viol.
24	2021	Online	0.3317	0.3326	31.90	0.3313	33.82	0e+00
24	2021	Proceed	0.3317	0.2851	4.04	0.3292	9.98	0e+00
24	2022	Online	0.3323	0.3339	37.98	0.3319	36.47	0e+00
24	2022	Proceed	0.3323	0.2844	4.38	0.3298	9.51	0e+00
24	2023	Online	0.3312	0.3327	35.95	0.3308	31.29	0e+00
24	2023	Proceed	0.3312	0.2831	3.74	0.3287	8.74	0e+00
48	2021	Online	0.3602	0.3623	52.69	0.3599	47.67	6e-07
48	2021	Proceed	0.3602	0.3166	4.48	0.3579	11.40	0e+00
48	2022	Online	0.3619	0.3637	47.96	0.3616	47.92	8e-07
48	2022	Proceed	0.3619	0.3115	3.67	0.3594	7.66	5e-08
48	2023	Online	0.3600	0.3618	49.37	0.3597	43.84	2e-06
48	2023	Proceed	0.3600	0.3117	6.21	0.3576	12.25	0e+00
96	2021	Online	0.3925	0.3939	36.89	0.3922	54.17	2e-07
96	2021	Proceed	0.3925	0.3428	15.96	0.3901	22.78	0e+00
96	2022	Online	0.3929	0.3942	34.92	0.3926	54.21	8e-07
96	2022	Proceed	0.3929	0.3454	10.17	0.3906	20.26	0e+00
96	2023	Online	0.3916	0.3935	46.72	0.3913	52.11	1e-06
96	2023	Proceed	0.3916	0.3496	12.16	0.3893	22.98	0e+00

Table 5: All eighteen electricity (321 variables, $H \in \{24, 48, 96\}$) adaptive-candidate cells (Online full-model gradient, PROCEED). Standardized units; violations $\leq 10^{-14}$ shown as `0e+00`. Descriptive per-cell values, no significance claim.

E.3 Allowance sweep on a priced domain

The priced-domain cost is a point on a tunable curve, not a fixed verdict. On transit (aggregate guard, eight cells: four candidates $\times$ two seeds), scaling the allowance by $\{0.25, 1, 4\}$—the same factors as the ETT budget study, applied to existing forecasts—traces the cost–gain tradeoff: a larger permitted burst retains more of the adapters' mean gain (Table 6). Here "gain retained" is the fraction of the candidate's mean-MAE gain that survives guarding; it is a different caliber from the mean-MAE-cost percentages in §6 (fraction of total MAE), so the two need not coincide. The last column reports the *realized* worst-interval excess (not the violation, which is 0 throughout): the three factors are distinct guarded trajectories, so this measured quantity is not monotone in the allowance—at the widest allowance no interval reaches the larger budget. The provable within-budget guarantee holds at every factor; the "cost" of protection is the operating point one selects.

factor	permitted burst B	gain retained	guarded worst-int. excess
0.25	0.096	11.6%	0.016
1	0.386	40.1%	0.015
4	1.543	70.3%	0.000

Table 6: Transit allowance sweep (aggregate guard, standardized units, mean over eight cells: four candidates $\times$ two seeds). Larger permitted burst retains more mean gain; the realized guarded worst-interval excess stays within budget.

F Proof and implementation of interval compression

This proves Theorem 3. For any start $a \leq m$, all forecasts before m are fully observed, and all forecasts from m onwards are the same partially observed suffix. Its completed centered contribution is $Z_m - Z_a$. Taking the largest over $a \leq m$ gives $Z_m - M_m$, computed using only completed losses. Remaining starts are $m+1, \ldots, t$; because at most $H-1$ older origins are pending, there are at most H distinct constraints including the compressed one. Equation (10) follows by summing their coordinate liabilities. All pending old targets lie among the new H targets under regular feedback, hence there is no untouched term. The reference adds identically zero excess and is feasible for every constraint by induction. Every ended interval keeps its certificate: later feedback only replaces an upper bound by a value it bounds. This proves the claimed simultaneous eventual guarantee. For

unrestricted target assignments, coordinate suprema and the separate interval conditions make the intersection exact.

For the independent control, bound each old origin's eventual excess by $u_i = \text{known}_i + (HN)^{-1} \sum_{\text{pending}} |g_i - b_i|$. Take the largest suffix sum of $u_i - \rho$, including the empty suffix zero, and add $Z_m - M_m$ only to the full recent suffix. Call this maximum U. Then all constraints reduce to $\|g - b\|_1/(HN) \le B + \rho - U$. This is one ball centered at b; a scalar ray has a closed-form coefficient. Observation settlement touches at most HN old coordinates; suffix sums and the new charge give $O(HN)$ work. Storing all pending forecasts uses $O(H^2N)$ space. Its feasible ball is contained in the exact intersection at a common history, without implying accuracy dominance along different histories.

The exact implementation streams physical target buckets. For each bucket it evaluates old absolute-loss differences at all old/new reference breakpoints, and takes reverse cumulative sums to obtain all suffix functions. Queries from excluded older forecasts are harmless additional points: the actual downward breakpoints and tails remain included. The two tail formulae then give every suffix's center and constant. There are $O(H^3N)$ arithmetic operations and $O(H^2N)$ stored values overall. Bisection checks every ball and rounds toward feasibility. The code rejects missing or out-of-order observations instead of silently extending the claim.

G Proof and numerical check of the horizon separation

Fix any issuance interval of length L. At each physical target the contributing origins form a contiguous alternating sequence of references $0, 2$. Pairing terms gives nonpositive excess by convexity of absolute loss. An unpaired term has supremum 1. The contributor counts are the convolution of length-L and length-H all-ones sequences. For even H, the number of odd counts equals H if $L < H$ is odd, L if $L < H$ is even, and H if $L \ge H$. Thus at most H terms remain unpaired, proving normalized interval excess at most 1 for every target assignment. Partial observations cannot raise the robust upper bound, so the full candidate remains feasible at every issuance.

On target 1, an origin of age j has j observed coordinates with total normalized excess $-j\alpha/H$, and independent remaining reserve $(H - j)\alpha/H$. With $K = H/2$, the most recent K origins therefore require $\sum_{j=0}^{K-1}(1 - 2j/H)\alpha_{t-j} \le 1$. Let $W = (H + 2)/4$ be the sum of weights. Summing for $t = K - 1, \dots, T - 1$ counts each interior coefficient with weight W. For $T \ge 2K - 1$, bounding the $2(K - 1)$ boundary coefficients by one gives

$$\frac{1}{T}\sum_{t=0}^{T-1} \alpha_t \le \frac{T - K + 1}{WT} + \frac{2(K - 1)}{T} = \frac{4}{H + 2} + O(H/T).$$

Improvement equals α_t on this target. Equal-magnitude corrections make the independent Euclidean ball projection a uniform shrinkage, so it also obeys this bound. Arbitrary horizon-specific safe decisions need not.

The constructed simulation uses $H \in \{4, 8, 16, 32, 64\}$, $T = 32H$, and the same $B = 1, \rho = 0$. At $H = 64$, exact acceptance is one throughout; independent mean acceptance over the second half is 0.060606, matching $4/66$ up to conservative numerical rounding. Both satisfy every checked interval. This mechanism check was specified after 22 of the 39 real-data records were completed; it is an exploratory construction, not a preregistered empirical discovery or evidence about typical reference errors.

H Separation beyond uniform horizon shrinkage

Appendix G proves the exact guard's full acceptance and the scalar-ray bound $4/(H+2)$; this section extends the bound to arbitrary independent certificates. For an arbitrary real scalar prediction g, reference $b \in \{0, 2\}$, and target 1, let $a = \max\{0, 1 - |g - 1|\}$ and $g' = b + a(1 - b)$. The canonical forecast improves the actual gain to a and has pending triangle charge $|g' - b| = a \le |g - b|$. Consequently, replacing each coordinate by its canonical forecast decreases both observed excess and independent pending reserve at every time on this target trajectory. It preserves all independent interval certificates. This is a comparison of realized sequences for an upper bound, not an assumption that the original policy knows future outcomes.

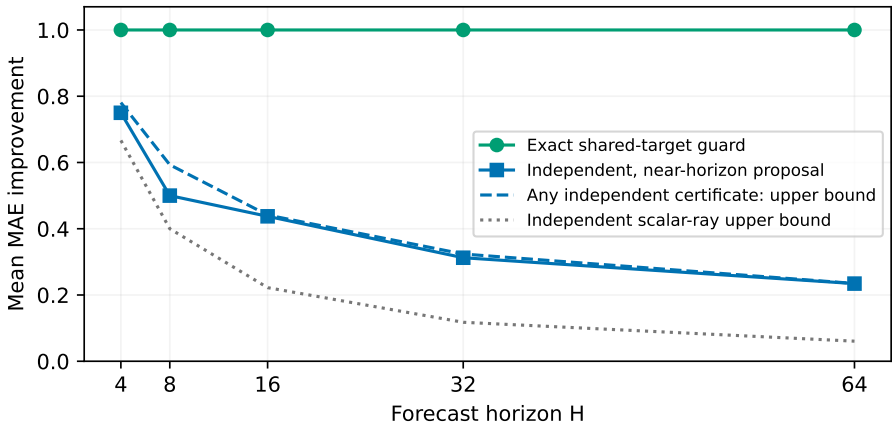


Figure 6: Constructed mechanism check, not a forecasting benchmark. Exact accounting accepts the full candidate; correcting near horizons beats uniform scalar shrinkage yet stays within the $O(H^{-1/2})$ bound. Both guards satisfy every interval allowance; the construction was specified after inspecting real-data interval results.

It suffices to consider $a_{t,h} \in [0,1]$. At issuance t, origin $t-j$ has observed its first j coordinates. Protecting the last $L \le H$ origins requires

$$\sum_{j=0}^{L-1} \sum_{h=1}^{H} (1 - 2\mathbf{1}\{h \le j\}) a_{t-j,h} \le H.$$

Sum this inequality for $t = L-1, \dots, T-1$. Relative to counting every coefficient at all origins, only at most $2(L-1)$ boundary origins differ. For fixed H, L their contribution divided by T vanishes. Along a subsequence attaining the limsup average gain, take a further convergent subsequence of the horizon-wise averages in the compact cube $[0,1]^H$. Its limit satisfies $\sum_h [2\min(L,h) - L]\bar{a}_h \le H$ simultaneously for all L.

Let $A = \sum_h \bar{a}_h$. For fixed mass, moving it to earlier horizons weakly reduces each constraint because its coefficients are nondecreasing in h. Fill the first $\lfloor A \rfloor$ coordinates with one and the next with the remainder. For integer $L \le \lfloor A \rfloor + 1$, its constraint value is $L(A+1-L)$. An integer nearest $(A+1)/2$ is valid for $0 < A \le H$ and yields at least $A(A+2)/4$. Thus $A \le \sqrt{4H+1} - 1$, proving the limsup bound, including for arbitrary noncanonical forecasts by domination.

For the matching order, take the largest integer $k \le H$ with $\lfloor (k+1)^2/4 \rfloor \le H$. Propose 1 on the first k horizons and the reference on the rest, then use the causal independent interval guard. On target 1, an origin of age j contributes $[k - 2\min(j,k)]/H$ to its independent certificate. The largest recent-suffix sum is $\lfloor (k+1)^2/4 \rfloor / H \le 1$. Older completed origins only reduce it. The proposal is therefore accepted from initialization, achieving $k/H = \Theta(H^{-1/2})$. On other target sequences the guard may shrink it; its pathwise guarantee is unchanged.

We verify the analytical bound against the time-average linear program for even $H = 2, \dots, 40$, and simulate the stronger independent proposal for $H \in \{4, 8, 16, 32, 64\}$ over $32H$ origins. At $H = 64$, it improves MAE by $15/64 = 0.234375$, compared with scalar-ray limit $4/66 = 0.060606$ and exact gain 1. The analytical arbitrary-policy bound is 0.234863; the finite linear program gives 0.234375. This stronger comparison was declared after all 39 real-data records completed. It changes neither their settings nor the absence of consistent exact-versus-independent accuracy superiority.

When the separation vanishes: co-monotone collapse. The construction above manufactures a sign change of $g_h - b_h$ at every horizon. The following shows this interleaving is necessary for any gap; whether that explains the absence of a real-data advantage is measured below.

Proposition 4 (Co-monotone collapse)**.** *Fix one target and its H pending forecasts with references b_h and issued values g_h. Write the exact (shared-target) and independent (marginal) liabilities as*

$L_{\mathrm{ex}} = \sup_y \sum_h (|g_h - y| - |b_h - y|)$ *and* $L_{\mathrm{ind}} = \sum_h \sup_y (|g_h - y| - |b_h - y|) = \sum_h |g_h - b_h|$. *If* $\mathrm{sign}(g_h - b_h)$ *is constant in* h*, then* $L_{\mathrm{ex}} = L_{\mathrm{ind}}$*; the exact and independent feasible sets for that target coincide and no separation occurs.*

Proof. Assume $g_h \geq b_h$ for all h (the case $g_h \leq b_h$ is symmetric under $y \mapsto -y$ about $\max_h b_h$). For any $y \leq \min_h b_h \leq g_h$, $|g_h - y| - |b_h - y| = (g_h - y) - (b_h - y) = g_h - b_h = |g_h - b_h|$, so $\sum_h (|g_h - y| - |b_h - y|) = \sum_h |g_h - b_h|$ for every such y and hence $L_{\mathrm{ex}} \geq \sum_h |g_h - b_h| = L_{\mathrm{ind}}$. The reverse inequality $L_{\mathrm{ex}} \leq L_{\mathrm{ind}}$ holds by subadditivity of the supremum. □

Thus the independent account can exceed the exact one only when the candidate crosses the reference across a target's horizons. Over 200 random co-monotone draws at each $H \in \{4, \ldots, 64\}$ the numerically evaluated gap is at most 1.4×10^{-14}, confirming the zero-gap case.

We then measured the collapse on the real forecast archives directly, and the answer has two parts. First, natural forecasts are *not* uniformly co-monotone: the per-target zero-gap fraction is 0.88 on transit (62 nodes, H=12; mean sign changes V=0.28) but only $0.39/0.18$ on ETTh2 DLinear (H=24/96) and 0.01 on the ADAPT-Z streams (H=48, V=8.7: the online adapter crosses the frozen reference repeatedly). Where the forecasts do cross, the worst-case per-target liability gap $L_{\mathrm{ind}} - L_{\mathrm{ex}}$ is genuine and can be large—mean $0.36\, L_{\mathrm{ind}}$ and 95th percentile $0.86\, L_{\mathrm{ind}}$ on ADAPT-Z. So the co-monotone collapse alone does not explain the empirical null. Second, the extra room is used yet leaves error unchanged. Replaying the exact- and independent-guarded streams from the archives, the two accountings issue *different* forecasts on about half the steps (0.51 on ETTh2); individual steps differ by up to 0.27–0.32 of the target scale, yet those differences average only 6.6–8.5×10^{-3} and the realized MAE differs by under 10^{-3} (2.23681 vs. 2.23679 at H=24). The test is sharpest on ADAPT-Z, the domain with the largest liability gap ($0.36\, L_{\mathrm{ind}}$): there the two guards issue different forecasts on 99.2% of steps, and although per-step differences reach 0.12–0.24 of the target scale, they average 1.8×10^{-3} and the realized MAE differs by at most 4.2×10^{-4} across three seeds—the inertness is therefore in the *consequence*, not in the size of the individual corrections. Nor is the room idle: both guards modify the candidate on 95.5–99.96% of ADAPT-Z and 86–94% of ETTh2 issuances; exact accounting retains 17–23% of the candidate correction versus 10–16% for the independent guard on ADAPT-Z (22–32% versus 15–20% on ETTh2); and natural targets realize the full exact worst-case liability on 12–17% (ADAPT-Z) and 29–40% (ETTh2) of targets, consuming 27–46% of the summed exact reserve. The null is therefore not explained by worst-case room that natural targets never reach; what the data show is narrower—the additional accepted correction is error-neutral at this operating point, for reasons left open. The $O(\sqrt{H})$ separation is therefore a genuine worst-case property that is *operationally inert* at the observed near-frontier operating point, with co-monotone collapse as its clean zero-gap special case (dominant on transit). A general $o(\sqrt{H})$ bound under stationarity with bounded crossing magnitude, and a bound on the realized (as opposed to worst-case) gap, are left open. Scripts: `verify_collapse_theorem.py`, `measure_gap_real.py`, `measure_operational_divergence.py`, `measure_opdiv_adaptz.py`, `measure_binding.py`.

I COMPLETE INTERVAL REPLAY RESULTS

Setting	Frozen	Candidate	Prefix I	Prefix E	Interval I	Interval E
$H = 24$, 0.0001	1.21653	1.19530	1.19530	1.19530	1.21145	1.21144
$H = 24$, 0.001	1.21653	1.19878	1.19898	1.19899	1.21051	1.20854
$H = 24$, 0.01	1.21653	1.42197	1.22498	1.22530	1.21466	1.21354
$H = 24$, 0.1	1.21653	1.76881	1.22550	1.22562	1.21500	1.21407
$H = 96$, 0.0001	1.83697	1.78975	1.79103	1.79103	1.82681	1.82681
$H = 96$, 0.001	1.83697	1.96705	1.84287	1.84334	1.83017	1.82975
$H = 96$, 0.01	1.83697	2.18909	1.84262	1.84508	1.83188	1.83251
$H = 96$, 0.1	1.83697	2.30598	1.84262	1.84594	1.83116	1.83211

Table 7: All ETTh1 interval cells, three-seed mean MAE in original units. I/E mean independent/exact ray. No learning rate is selected by test error. The largest rates diagnose sensitivity and are not advertised as well-tuned deployment settings.

Setting	Candidate	Prefix I	Prefix E	Interval I	Interval E
etth1, 24, 0.0001	17.097	17.097	17.097	0.000	0.000
etth1, 24, 0.001	29.301	24.720	24.720	0.000	0.000
etth1, 24, 0.01	661.780	20.988	19.871	0.000	0.000
etth1, 24, 0.1	1583.593	23.102	16.566	0.000	0.000
etth1, 96, 0.0001	92.200	92.200	92.200	0.000	0.000
etth1, 96, 0.001	542.454	92.286	95.669	0.000	0.000
etth1, 96, 0.01	1077.324	66.709	73.012	0.000	0.000
etth1, 96, 0.1	1374.384	64.468	36.260	0.000	0.000
etth2, 24, unprotected	20.345	20.345	20.345	0.000	0.000
etth2, 96, unprotected	111.797	34.095	34.102	0.000	0.000
strong, 48, adapt_z	4.189	4.189	4.189	0.000	0.000
strong, 48, head_ogd	0.382	0.382	0.382	0.000	0.000
strong, 48, spectral	0.323	0.323	0.323	0.000	0.000

Table 8: Maximum positive interval-budget violation across three seeds, in summed per-origin MAE units. Strong-carrier rows use standardized units; ETTh1/ETTh2 DLinear rows use their original target units. Zero is a numerical constraint check, not a statistical significance statement.

All 39 candidate records, 78 new forecast trajectories, paths, initialization SDs, per-seed errors, acceptance coefficients, and H/$4H$ window diagnostics are retained. Strong candidates share backbone fits, so they are not independent replications. The synthetic separation does not alter these real-data comparisons or their budgets.

J ORIGINAL FORECASTING PROTOCOLS

Transit carrier and candidates. The Beijing transit study (62 stations, delayed feedback) uses an existing graph-state forecaster with twelve input and twelve output steps, width 64, and a common deployment checkpoint per seed; two seeds are archived. Ten initial days are followed by eight one-day evaluation blocks. This is an already known development stream; no untouched test-city claim is made. The four adaptive candidates are: (i) online spectral calibration, the unmodified `FRPlusModule` of the official ST-TTC engine at commit `ddba3fe` (Git blob hash verified, Apache license preserved; four frequency groups, zero-initialized amplitude/phase corrections, Adam at 10^{-4}), updating only after all twelve targets of an older sample mature; (ii) its block-feedback version, which takes one chronological training pass over the previous scored block and then freezes during the next block; and (iii)–(iv) previous-block mean and median residual correction. The module receives normalized frozen-carrier forecasts and is scored after inverse scaling. The guard observes individual target coordinates as they mature, and both reserve variants receive exactly the same feedback. Reusing cached frozen-backbone predictions is exact for these candidates and does not retrain the backbone. The budget is $\rho = 10^{-3}\sigma_{\text{init}}$, $B_0 = H\rho$, with σ_{init} the

initial-series standard deviation. Zero-masked MAE (§3) is kept separate from the all-entry loss the certificate controls, because the two reverse the candidates' ranking: on the first seed, all-entry MAE orders them median block (4.582) < mean block (4.726) < online spectral (5.266) < block spectral (5.271), whereas masked MAE orders them exactly the other way (6.617, 6.619, 7.089, 7.146 for online spectral, block spectral, mean, median).

Temperature protocol, declared before running. We use the public ETTh1 oil-temperature series (Zhou, 2020), with 8,640/2,880/2,880 chronological training/validation/test observations. An official DLinear model (Zeng et al., 2023) uses input length 168, $H \in \{24, 96\}$, training-only normalization, Adam at 10^{-3}, batch size 64, and the lowest validation-MAE checkpoint among 30 epochs. For each of three seeds, online spectral calibration uses all rates $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. Rates are a sensitivity grid, never selected on the test set; the largest rates deliberately stress candidate optimization, and no output corruption is applied. All test windows end within the test partition. The budget is $\rho = 10^{-3}\sigma_{\text{train}}$, $B_0 = H\rho$. ETTh2 follows the same split, backbone, budgets, seeds and horizons under a protocol fixed before inspecting its test results (Appendix D).

Controls in the prefix grid (Appendix K). Half-strength scales the candidate by one half. Delayed Hedge receives a sample's two expert losses only at full maturity, clips MAE$/\sigma_{\text{train}}$ to $[0, 1]$ and uses $\eta = \sqrt{8\log(2)/T}$; it is an auxiliary weighting rule with no raw-MAE or delay-free regret guarantee attributed to it. The official SAFE-TS gate uses its default 2,000 bootstrap draws and 0.05 lower quantile on paired relative gains; we supply the most recent eight fully mature issuance blocks of length H, require two blocks before activating, and otherwise freeze. These are our declared integration settings, not the full SAFE-TS benchmark, and adjacent blocks may share targets, so a bootstrap independence interpretation is limited.

Audit. The guard receives only mature observations. A separate offline evaluator computes every prefix's eventual excess using archived targets and asserts the budget inequality numerically. Unit tests cover exact liability versus all breakpoints, cancellation, large arbitrary targets, zero budgets, and every-prefix safety. Source, data, and deployment hashes link forecasts to their runs. Seed-level full-trajectory summaries avoid treating overlapping forecasts as independent replicates.

K ETTh1 PREFIX SENSITIVITY GRID WITH THE SAFE-TS GATE

This earlier prefix-budget study (before the every-interval objective) is retained because it is the only place the official SAFE-TS gate enters as a common-candidate control. "Half-strength" scales the spectral candidate by one half; "Hedge" is an archived causal expert-weighted proposal.

H	LR	n	Frozen	Unguarded	Half	Hedge	Gate	Indep.	Shared
24	10^{-4}	3	1.217	1.195	1.201	1.200	1.217	1.195	1.195
24	10^{-3}	3	1.217	1.199	1.189	1.188	1.217	1.199	1.199
24	10^{-2}	3	1.217	1.422	1.251	1.218	1.217	1.225	1.225
24	10^{-1}	3	1.217	1.769	1.368	1.238	1.217	1.225	1.226
96	10^{-4}	3	1.837	1.790	1.796	1.798	1.837	1.791	1.791
96	10^{-3}	3	1.837	1.967	1.842	1.845	1.854	1.843	1.843
96	10^{-2}	3	1.837	2.189	1.862	1.823	1.837	1.843	1.845
96	10^{-1}	3	1.837	2.306	1.886	1.832	1.837	1.843	1.846

Table 9: ETTh1 all-entry MAE, mean of three seeds for every horizon/rate cell. All 24 cells are shown. Gate is the official SAFE-TS gate on our common spectral candidate. The largest rates are an optimization sensitivity test; they are not presented as well-tuned deployment settings.

At the official spectral rate 10^{-4}, mean MAE improves from 1.217 to 1.195 for $H = 24$, and from 1.837 to about 1.791 for guarded $H = 96$ (Table 9). At larger rates, adaptation becomes harmful without injected noise. Both reserve guards constrain every evaluated prefix; unprotected, half-strength, Hedge, and the bootstrap gate have violations in some cells. The gate often freezes entirely, thereby losing useful adaptation; it does not itself claim our deterministic budget property.

These results do not establish an accuracy advantage of exact accounting: for $H = 96$ at rate 10^{-1}, independent and shared MAE are about 1.843 and 1.846. Final average error, accepted correction magnitude, and the anytime budget are three different outcomes (Figure 7).

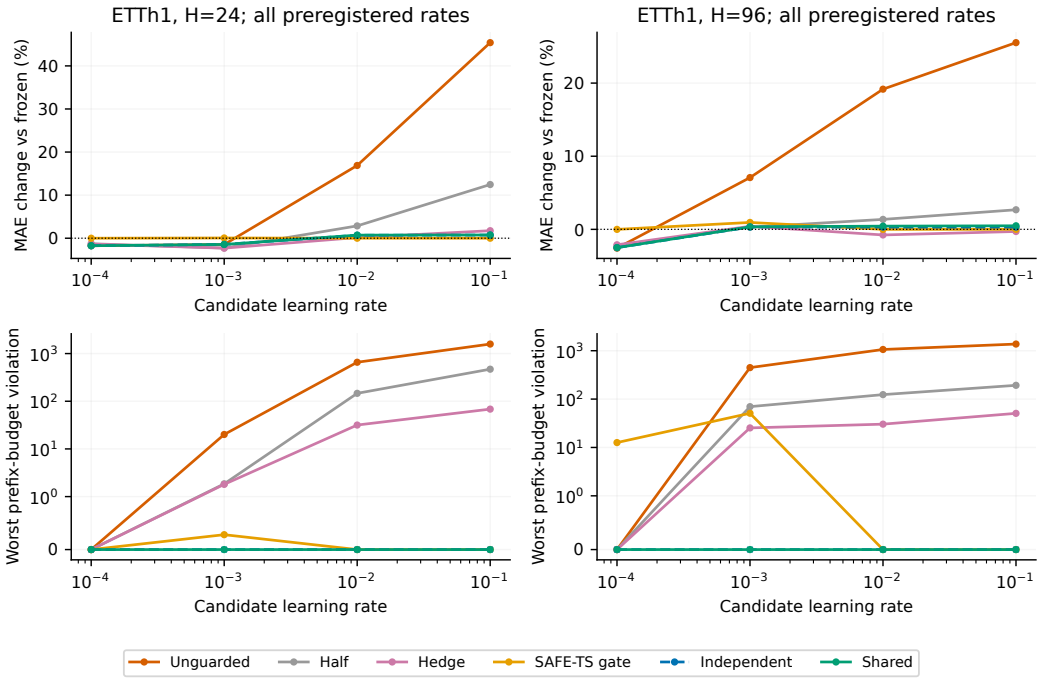


Figure 7: Complete ETTh1 sensitivity grid. Top: seed-mean relative MAE. Bottom: largest positive prefix violation across three seeds, with a symmetric-log axis that retains zero. Both reserve controls satisfy the budget, but their accuracy advantage over each other is absent. This is conditional evidence on one series and two horizons, not population-level generalization from independent datasets.

L Stronger independent controls and allowance curves

The two independent vector controls compared in §6 use the same independent feasible ball centered at b_t with allowance Q_t and solve

$$\min_{\|x-b_t\|_1 \le Q_t} \frac{1}{2} \sum_{h,n} w_h (x_{h,n} - c_{h,n})^2, \qquad w_h = 1 \text{ or } w_h = 1/h, \tag{12}$$

with weights fixed before replay. The 39 temperature records are replayed under three allowance factors and four guards (468 policy trajectories, including 78 scalar runs reused bit-for-bit from the initial interval study); Exchange adds 108, for 576 in total. These counts reflect policies sharing candidate forecasts, not additional fits. Both allowance components scale together, every candidate rate, seed and factor is retained, and no endpoint is selected as an optimized deployment budget. Source/protocol hashes are preserved, shared physical targets match exactly in all 36 source archives, and initialization SDs and paired errors are archived without treating overlapping windows as independent samples. For fixed positive weights, write $v = c - b$. The weighted projection is $x = b + \mathrm{sign}(v) \max\{|v| - \lambda/w, 0\}$ coordinate-wise. Sort $w_j|v_j|$ in descending order. For an active set of the first k coordinates, $\lambda = (\sum_{j \le k} |v_j| - Q)/(\sum_{j \le k} 1/w_j)$. Choose the consistent active set, or $\lambda = 0$ for a feasible proposal. This is a standard weighted quadratic projection, with $O(HN \log(HN))$ sorting after the independent ledger update. Outputs use the declared numerical feasibility tolerance. Tests check KKT conditions, the unweighted special case, arbitrary-target interval safety, and unchanged prefixes under future perturbations. No test outcome chooses the weights.

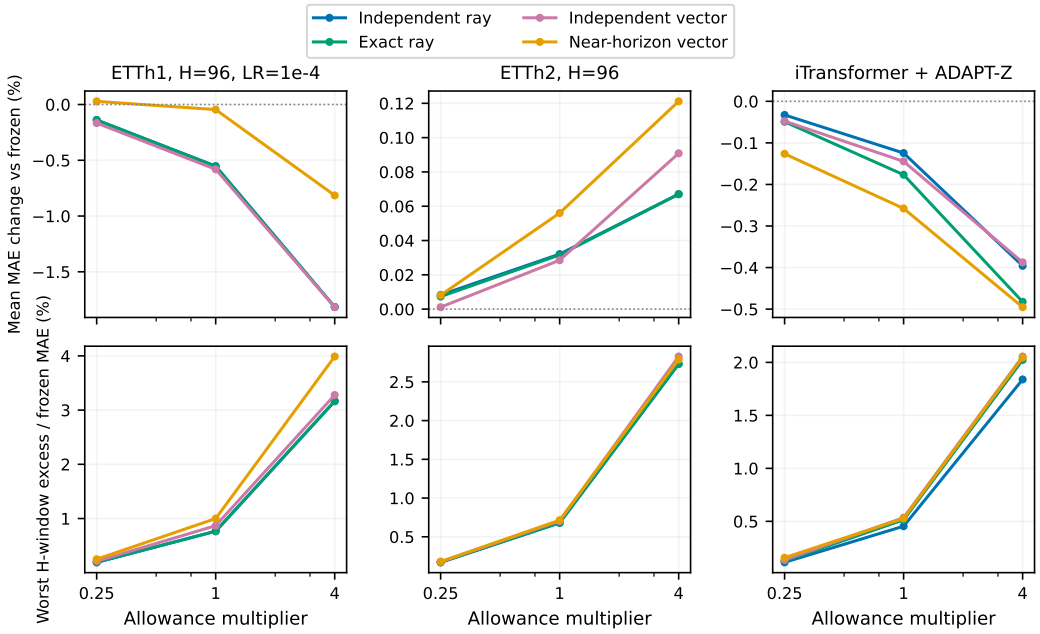


Figure 8: Allowance curves for the official-rate long-horizon ETTh1 candidate, validation-selected ETTh2 failure, and feature-adaptation candidate. Top: three-seed mean MAE relative to the frozen model. Bottom: worst H-window mean excess across seeds, scaled by the mean frozen MAE for display only. The input allowances still use training-only scales. Larger allowance need not improve accuracy or yield the same harm–gain tradeoff across candidates.

Setting	Frozen	Candidate	Indep. ray	Exact ray	Indep. vec.	Near vec.
ETTh1 24, 0.0001	1.21653	1.19530	1.21518	1.21514	1.21418	1.21452
ETTh1 24, 0.001	1.21653	1.19878	1.21496	1.21425	1.21391	1.21317
ETTh1 24, 0.01	1.21653	1.42197	1.21600	1.21557	1.21473	1.21620
ETTh1 24, 0.1	1.21653	1.76881	1.21608	1.21559	1.21506	1.21693
ETTh1 96, 0.0001	1.83697	1.78975	1.83442	1.83441	1.83391	1.83749
ETTh1 96, 0.001	1.83697	1.96705	1.83519	1.83501	1.83449	1.83695
ETTh1 96, 0.01	1.83697	2.18909	1.83563	1.83547	1.83499	1.83797
ETTh1 96, 0.1	1.83697	2.30598	1.83545	1.83517	1.83496	1.83858
ETTh2 24	2.22769	2.21860	2.22684	2.22681	2.22599	2.22226
ETTh2 96	3.24163	3.27704	3.24190	3.24187	3.24167	3.24190
iT-ETTh2 ADAPT-Z	0.31846	0.31536	0.31836	0.31831	0.31831	0.31806
iT-ETTh2 Head OGD	0.31846	0.31704	0.31833	0.31825	0.31819	0.31805
iT-ETTh2 Spectral	0.31846	0.31787	0.31837	0.31837	0.31835	0.31834

Table 10: Quarter allowance, all temperature groups. Units and labels as in Table 1; all entries are means of three seeds.

Setting	Frozen	Candidate	Indep. ray	Exact ray	Indep. vec.	Near vec.
ETTh1 24, 0.0001	1.21653	1.19530	1.20222	1.20222	1.20035	1.20263
ETTh1 24, 0.001	1.21653	1.19878	1.19806	1.19717	1.19431	1.20048
ETTh1 24, 0.01	1.21653	1.42197	1.21316	1.21403	1.20829	1.23214
ETTh1 24, 0.1	1.21653	1.76881	1.21463	1.21924	1.21191	1.24057
ETTh1 96, 0.0001	1.83697	1.78975	1.80362	1.80362	1.80359	1.82200
ETTh1 96, 0.001	1.83697	1.96705	1.81308	1.81386	1.81571	1.83908
ETTh1 96, 0.01	1.83697	2.18909	1.82066	1.82719	1.82492	1.86386
ETTh1 96, 0.1	1.83697	2.30598	1.81789	1.83027	1.82646	1.86575
ETTh2 24	2.22769	2.21860	2.22080	2.22080	2.22054	2.22149
ETTh2 96	3.24163	3.27704	3.24381	3.24381	3.24458	3.24556
iT-ETTh2 ADAPT-Z	0.31846	0.31536	0.31720	0.31693	0.31723	0.31689
iT-ETTh2 Head OGD	0.31846	0.31704	0.31713	0.31709	0.31711	0.31708
iT-ETTh2 Spectral	0.31846	0.31787	0.31787	0.31787	0.31787	0.31787

Table 11: Fourfold allowance, all temperature groups. Units and labels as in Table 1; all entries are means of three seeds. No test-optimal allowance or candidate rate is selected from this table.

M Frozen new-domain Exchange Rate replication

We obtain the eight-series daily dataset from the LSTNet authors' repository (Lai et al., 2018), pinned to commit `7f402f1`; its compressed Git blob and decompressed SHA256 are verified. There are 7,588 rows and eight finite columns. Chronological 60/10/30% partitions and within-partition windows give 4,409/616/2,134 forecast origins. Scaling uses training-only column moments. Input length is 96, horizon 48, and seeds are $0, 1, 2$.

The protocol was frozen before obtaining forecast outcomes on this dataset. It transfers the same iTransformer width, layers, optimizer, epoch limit, validation-checkpoint rule, normalization convention, and adapter/head/spectral rates as Appendix E. The official ADAPT-Z class uses its existing `exchange` settings. Three validation warmup passes precede chronological test batches of 24; no rate, column, or seed is selected by test accuracy. Each adaptive candidate is guarded at all three allowance levels. The frozen warm adapter additionally separates initialization from online updates in the archived candidate results.

The harness verifies zero-feature perturbation equivalence, independent batch predictions, exact shared-target identities, and unchanged prediction prefixes after perturbing unobserved observations. Separate checks cover the official feature class and both head/spectral controls. This is a new-domain common-protocol replication, not a reproduction of published Exchange benchmark scores, a trading strategy, or evidence about investment returns.

Proposal, budget	Frozen	Candidate	Indep. ray	Exact ray	Indep. vec.	Near vec.
ADAPT-Z, 0.25	0.18177	0.18113	0.18169	0.18168	0.18172	0.18158
ADAPT-Z, 1	0.18177	0.18113	0.18148	0.18146	0.18152	0.18144
ADAPT-Z, 4	0.18177	0.18113	0.18105	0.18104	0.18108	0.18111
Head OGD, 0.25	0.18177	0.18150	0.18169	0.18169	0.18164	0.18160
Head OGD, 1	0.18177	0.18150	0.18150	0.18150	0.18147	0.18146
Head OGD, 4	0.18177	0.18150	0.18150	0.18150	0.18150	0.18150
Spectral, 0.25	0.18177	0.18171	0.18174	0.18174	0.18175	0.18174
Spectral, 1	0.18177	0.18171	0.18172	0.18172	0.18172	0.18172
Spectral, 4	0.18177	0.18171	0.18171	0.18171	0.18171	0.18171

Table 12: Exchange Rate: three-seed mean training-standardized MAE for every candidate and allowance factor. All four guard policies have zero checked interval-budget violation; unprotected candidates violate 19/27 seedwise candidate/allowance cells. Tiny mean differences are not significance claims.

N EXACT-SET VECTOR PROJECTION: MATCHED STATES, TRAJECTORIES AND COST

This appendix collects the studies behind the "cost and solver choice" paragraph of §6. Their common finding is negative for the exact geometry as a *forecasting* advantage: at identical states the exact intersection retains more of a candidate, but completing that projection more precisely does not consistently lower online error, and it costs one to three orders of magnitude more latency than independent projection. The complete numerical audits (witness enclosures, exact-rational feasibility checks, serial timing protocols) are retained in the artifact; only the results that the main text cites are reproduced here: the matched-state comparison (Table 13), the two complete trajectory cohorts (§N.1, §N.2) and guard latency (§N.3, Table 16).

Matched-state protocol. After inspecting the stronger independent controls, we froze a matched-objective study on all 48 candidates at their original allowance, with both $W = I$ and $w_h = 1/h$. Common states come from the corresponding independent policy at issuance indices $\lfloor q(T-1) \rfloor$, $q \in \{.25, .5, .75, 1\}$; the diagnostic exact actions do not enter the carrier policy's subsequent ledger. The resulting $48 \times 2 \times 4 = 384$ projections isolate the effect of constraint geometry at identical histories.

Classical solver and numerical certificate. Write $v = c - b$, $a_j = m_j - b$, and $C_j = \{z : \|z - a_j\|_1 \le r_j\}$; the reference 0 is feasible. The statewise optimum is $P^* = \min_{z \in \cap_j C_j} \frac{1}{2}\|z - v\|_W^2$. With support functions $\sigma_j(u) = a_j^\top u + r_j \|u\|_\infty$ and $S = \sum_j u_j$, weak duality gives

$$L(u) = v^\top S - \tfrac{1}{2} S^\top W^{-1} S - \sum_j \sigma_j(u_j) \le P^* \tag{13}$$

for arbitrary u_j, without assuming that a finite solver iterate is optimal. Weighted Dykstra (Tibshirani, 2017) cycles through the balls using the threshold rule of Appendix L; a finite sweep need not be feasible for every ball, so we separately repair it along the feasible reference ray and retain the best feasible primal point $\hat{z}$. With the largest lower bound $\max(0, L)$ and the best feasible objective $P(\hat{z})$, the gap G bounds suboptimality: $\|\hat{z} - z^*\|_W^2 \le 2G$ and $|\,\mathrm{MAE}(b + \hat{z}, y) - \mathrm{MAE}(b + z^*, y)| \le \sqrt{2G/(HN \min w)}$ for any target y. This is a current-state bound, not a bound on two complete online policies whose later histories differ. Calculations use float64 with a conservative roundoff allowance and a 10^{-7} feasibility audit. The common-state cap is 1,000 sweeps and the stopping threshold 10^{-6} in the displayed MAE bound; 369 of 384 projections reach it and all capped cases are reported. No procedure here is claimed as a new general solver.

Family	Weight	Conv./total	Improved	Gain bracket (%)	Max. MAE bound
ETTh1	1	90/96	95	[2.91, 2.92]	0.112
ETTh1	$1/h$	89/96	74	[0.99, 0.99]	0.0282
ETTh2	1	24/24	19	[1.57, 1.57]	1.38e-07
ETTh2	$1/h$	24/24	14	[4.17, 4.17]	2.64e-07
Exchange	1	35/36	16	[30.04, 30.04]	0.000118
Exchange	$1/h$	36/36	16	[24.22, 24.22]	9.52e-08
iT-ETTh2	1	35/36	29	[51.70, 51.70]	0.000356
iT-ETTh2	$1/h$	36/36	29	[43.57, 43.57]	1.02e-07

Table 13: Matched-history comparison. "Improved" counts feasible quadratic objective improvements exceeding 10^{-10}. The gain bracket bounds the reduction of the family's summed independent distortion, using feasible primal and dual values; it is not a confidence interval or an MAE gain. ETTh1 pools its fixed rates and horizons descriptively. Maximal numerical MAE bounds expose the 15 cases that missed the requested tolerance.

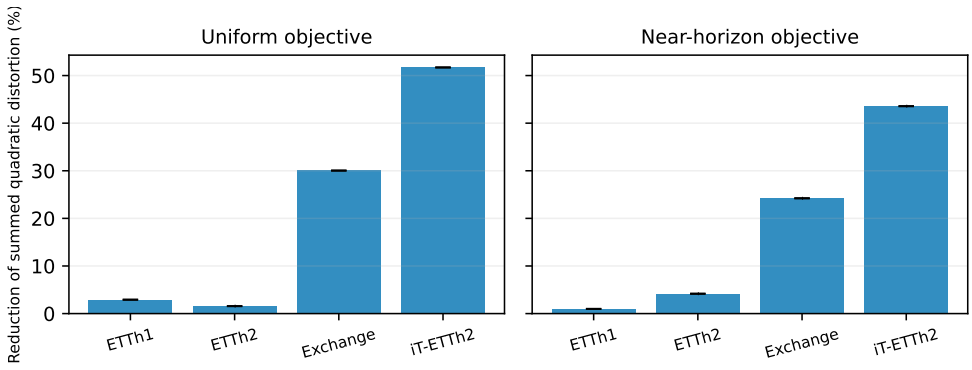


Figure 9: The exact intersection can retain more of a candidate at identical states and objectives. Bars are feasible lower bounds on the reduction of summed quadratic distortion; whiskers end at dual-derived upper bounds, not sampling uncertainty. This does not establish lower online MAE.

N.1 Complete finite-iteration vector trajectories

The frozen follow-up runs both matching quadratic objectives on all 48 candidate records at the original allowance, producing 96 full causal trajectories, each exact-set vector policy acting on its own history. Weighted Dykstra is capped at 100 sweeps per decision with the same 10^{-6} statewise tolerance and a separately feasible output; no test-selected rate, allowance, seed or candidate is omitted. An independent audit reloads every forecast and recomputes all-entry MAE and every-interval excess: all 96 trajectories satisfy their interval budgets at 10^{-7} tolerance (maximum violation 6.60×10^{-14}). Across 299,166 issued decisions, 72.88% meet the numerical accuracy criterion; the largest statewise MAE bound is 3.3015 in original temperature units.

Both fixed objectives give lower seed-mean MAE than their matching independent policies in 14 of 16 settings: all 13 temperature settings and Exchange head OGD. Exchange ADAPT-Z and spectral calibration have higher means, the latter differences below 3×10^{-7} in standardized MAE. This describes the implemented feasible approximation, not an exact optimizer's online performance or a significance claim; the 16 settings share data and backbone fits, and some individual seeds reverse the mean comparison (Table 14).

Setting	Indep.	Exact approx.	Conv.%	Near indep.	Near exact approx.	Conv.%
ETTh1 24, 0.0001	1.20946	1.20755	86.2	1.21057	1.21030	97.2
ETTh1 24, 0.001	1.20748	1.19911	63.3	1.20838	1.20520	87.3
ETTh1 24, 0.01	1.21112	1.20354	47.8	1.21889	1.21731	83.1
ETTh1 24, 0.1	1.21253	1.20636	44.0	1.22140	1.22060	73.6
ETTh1 96, 0.0001	1.82630	1.82526	62.5	1.83613	1.83610	94.6
ETTh1 96, 0.001	1.82974	1.82262	32.2	1.83949	1.83834	72.0
ETTh1 96, 0.01	1.83057	1.82503	34.7	1.84385	1.84277	63.1
ETTh1 96, 0.1	1.83090	1.82797	38.3	1.84479	1.84424	61.5
ETTh2 24	2.22403	2.22380	93.5	2.22193	2.22193	97.2
ETTh2 96	3.24256	3.24239	83.0	3.24345	3.24333	93.2
Exchange ADAPT-Z	0.18152	0.18155	57.5	0.18144	0.18147	79.0
Exchange Head OGD	0.18147	0.18146	92.8	0.18146	0.18145	97.0
Exchange Spectral	0.18172	0.18172	99.7	0.18172	0.18172	99.8
iT-ETTh2 ADAPT-Z	0.31800	0.31768	29.8	0.31764	0.31749	59.8
iT-ETTh2 Head OGD	0.31776	0.31747	68.5	0.31763	0.31743	76.9
iT-ETTh2 Spectral	0.31811	0.31809	93.8	0.31810	0.31809	97.8

Table 14: All 96 finite-iteration exact-set vector trajectories. MAEs are means of three seeds in each row's original metric; "Conv." is the fraction of decisions meeting the 10^{-6} numerical MAE-bound criterion. Temperature DLinear rows use original units, iTransformer rows standardized units. Every guard passes the interval audit, including capped decisions.

N.2 COMPLETE CONDITIONAL-HYBRID VECTOR TRAJECTORIES

A second frozen solver completes the same 96 dense trajectories using up to 500 iterations of a classical scalar-multiplier dual (L-BFGS-B) and, when the stored accuracy criterion remains unresolved, up to 1,000 weighted Dykstra sweeps. Each policy acts on its own causal history; there is no test-selected cap or omitted seed. All 96 trajectories pass the interval audit at 10^{-7} (largest violation 4.11×10^{-14}). Across 299,166 decisions, 96.47% meet the numerical MAE-bound criterion, against 72.88% for Dykstra100; fallback is invoked on 11.83% of decisions and the largest remaining bound is 0.6189 in original ETTh1 units. A high aggregate completion fraction does not establish accuracy at every state.

As with Dykstra100, each objective improves seed-mean MAE over its matching independent vector policy in 14/16 settings. Against Dykstra100, however, hybrid MAE is lower in only 9/16 uniform-weight and 7/16 near-weight settings (Table 15). Differences are small and individual seeds disagree: more numerical work does not consistently improve this online prediction objective, because a per-state optimization bound does not bound the divergence of two policy histories.

	Uniform				Near-horizon			
Setting	Indep.	Dykstra100	Hybrid	Conv.%	Indep.	Dykstra100	Hybrid	Conv.%
ETTh1 24, 0.0001	1.20946	1.20755	1.20755	100.0	1.21057	1.21030	1.21030	100.0
ETTh1 24, 0.001	1.20748	1.19911	1.19910	97.0	1.20838	1.20520	1.20520	99.6
ETTh1 24, 0.01	1.21112	1.20354	1.20352	93.9	1.21889	1.21731	1.21731	97.9
ETTh1 24, 0.1	1.21253	1.20636	1.20633	91.9	1.22140	1.22060	1.22061	93.3
ETTh1 96, 0.0001	1.82630	1.82526	1.82526	99.5	1.83613	1.83610	1.83610	100.0
ETTh1 96, 0.001	1.82974	1.82262	1.82261	90.0	1.83949	1.83834	1.83835	94.5
ETTh1 96, 0.01	1.83057	1.82503	1.82501	83.7	1.84385	1.84277	1.84277	82.7
ETTh1 96, 0.1	1.83090	1.82797	1.82799	84.5	1.84479	1.84424	1.84427	78.6
ETTh2 24	2.22403	2.22380	2.22380	99.9	2.22193	2.22193	2.22193	100.0
ETTh2 96	3.24256	3.24239	3.24239	99.5	3.24345	3.24333	3.24333	99.8
Exchange ADAPT-Z	0.18152	0.18155	0.18155	98.4	0.18144	0.18147	0.18147	99.6
Exchange Head OGD	0.18147	0.18146	0.18146	100.0	0.18146	0.18145	0.18145	100.0
Exchange Spectral	0.18172	0.18172	0.18172	100.0	0.18172	0.18172	0.18172	100.0
iT-ETTh2 ADAPT-Z	0.31800	0.31768	0.31768	96.9	0.31764	0.31749	0.31749	98.0
iT-ETTh2 Head OGD	0.31776	0.31747	0.31747	99.8	0.31763	0.31743	0.31743	99.5
iT-ETTh2 Spectral	0.31811	0.31809	0.31809	100.0	0.31810	0.31809	0.31809	100.0

Table 15: All 96 completed hybrid trajectories and matched controls. Each MAE is a three-seed mean. DLinear temperature rows use original units; iTransformer rows use training-standardized units. "Conv." is the decision-weighted tolerance-attainment percentage. Values rounded to five decimals may conceal small differences; the seedwise paired changes are archived.

N.3 GUARD LATENCY ON FIXED HISTORIES

A frozen timing protocol replays all 640 pre-feedback histories (384 dense common states plus 256 states from four archived multivariate transit streams, two seeds, four candidates and two objectives) in one Linux process pinned to a single Xeon core with one numerical-library thread, three repetitions and deterministic interleaving: 7,680 timed calls. The timer covers feedback settlement, geometry, initialization, projection, feasibility repair and ledger insertion; backbone/candidate computation, state restoration and I/O are outside it. It compares the independent vector control, the conditional hybrid above, and a separately versioned variant of the same solvers that accepts only exactly verified feasible points ("verified" rows). An independent breakpoint/tail audit checks all 2,560 saved decisions (maximum violation 2.45×10^{-15} under the 10^{-7} tolerance).

Histories	Method	Mean	Median	P95	Maximum
Dense (384)	Independent vector	0.39	0.37	0.60	0.64
	Legacy hybrid	48.79	8.55	68.94	1930.42
	Verified scalar	22.61	15.17	45.58	314.27
	Verified hybrid	58.69	15.15	95.34	5892.06
Traffic (256)	Independent vector	0.27	0.26	0.38	0.42
	Legacy hybrid	278.42	5.77	813.04	976.63
	Verified scalar	36.05	32.18	66.83	223.06
	Verified hybrid	140.58	32.34	73.44	7180.50

Table 16: Full-guard wall latency in milliseconds. Each state contributes its median of three repetitions; columns summarize all such state medians. Maximum retains rare difficult states that P95 can conceal. CPU times and all individual repetitions are archived. "Traffic" rows are the transit histories.

Independent projection settles in a fraction of a millisecond on both history sets, whereas the exact-set solvers cost tens of milliseconds on dense histories and hundreds on transit histories. Verification is not uniformly faster: mean transit latency falls from 278.42 to 140.58 ms while the median rises from 5.77 to 32.34 ms and the maximum from 976.63 to 7,180.50 ms; on dense histories the mean rises from 48.79 to 58.69 ms. These inspected independent-policy histories measure guard overhead, not each policy's own state distribution, whole-stream throughput or forecast MAE.

O IRREGULAR ISSUANCE WITH COMPLETE PENDING FEEDBACK

Proposition 5. *Let physical issuance times $\tau_0 < \tau_1 < \cdots$ be distinct integers. Every issued forecast predicts $\tau_i + 1, \ldots, \tau_i + H$, and all required targets through the current physical time are observed before issuing. For an interval $[a, b]$ of* issued forecast indices*, the allowance is $B + \rho(b - a + 1)$. Then the same completed statistics and at most H active ball constraints suffice.*

Proof. Common physical horizon H means forecasts mature in issuance order, even if there are gaps. All completed starts have the same pending suffix, so their dominant completed contribution is still $Z_m - \min_{k \le m} Z_k$, with Z centered by ρ once per issued forecast. Before issuing at t, pending origins belong to $\{t - H + 1, \ldots, t - 1\}$; distinct integer issuance times imply at most $H - 1$ such origins. Every unresolved old target lies in the new horizon. The suffix geometry and shared-target liability argument therefore apply unchanged. Intermediate feedback can only decrease the liability certificates. Idle physical ticks create no allowance credit. □

Here the allowance is precisely $B + \rho(b - a + 1)$ for issued indices $[a, b]$; it is not an allowance proportional to elapsed calendar time. Multiple issuances per tick, missing required labels and variable horizons with out-of-order completion are outside this implementation.

The separate physical-clock implementation rejects skipped pending observations and retains completed statistics across empty gaps. Dense issuance outputs are bit-identical to the historical ray and independent vector implementations. Random irregular schedules test arbitrary-target interval budgets and future perturbations. An empty-gap test verifies that neither the allowance nor the completed ledger changes merely because time passes. The transit replays (blocks separated by idle ticks) run on this clock; no irregular-interval accuracy experiment is claimed.