MONITORING AND ADAPTING DEPLOYED DECISION-FOCUSED MODELS WITH SPARSE FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Decision-focused learning (DFL) trains predictive models to directly optimize downstream decision quality. Existing work addresses training and online updates with per-round feedback, but the deployment of a frozen DFL model under sparse feedback remains largely open. We study this setting: a trained predictor is deployed while the data distribution may drift, and the true cost vector is observed only through selective label queries subject to a finite budget. The operator must determine when decision quality has degraded and whether a candidate model deserves promotion. Yet regret is selectively observed, detection must remain valid at arbitrary stopping times, and an update triggered by an alarm can itself harm decisions. We propose E-REACT, a deployment framework whose detection rule provides finite-sample, anytime-valid false-alarm control and whose promotion gate controls erroneous deployments over the model's lifetime. We further characterize the optimal tradeoff between the label budget and detection delay, establish uniform sampling as the unique minimax allocation under volatility-reshaping shifts, and prove that scalar error monitors can be blind to regret drift. Across four combinatorial optimization problems and two real-world histories, E-REACT detects decision-relevant shifts missed by error monitoring. Its promotion gate blocks every spurious candidate under false alarms, and guarded updates reduce regret by up to 26.1% without harming any of 200 paired streams under genuine drift.

1 INTRODUCTION

Many operational systems use predictions as inputs to optimization models. Demand forecasts determine allocations, return forecasts determine portfolios, and travel-time forecasts determine routes. In such systems, predictive error is only indirectly relevant. An error can be harmless when it leaves the optimizer's solution unchanged, yet consequential when it moves the prediction across a decision boundary. Decision-focused learning (DFL) addresses this mismatch during training by fitting a predictor to the quality of the decisions it produces rather than to predictive accuracy alone (Donti et al., 2017; Elmachtoub & Grigas, 2022; Mandi et al., 2024).

We study the deployment problem in which a trained DFL predictor remains frozen while the joint distribution of features and costs may drift. This problem is common when model changes are discrete, expensive, or subject to approval. It also creates a feedback bottleneck. The feature and deployed decision are available immediately, but evaluating the decision requires the true cost vector, which is revealed only through a costly audit and an additional optimizer call. Deployment therefore depends on two statistical decisions: when the evidence warrants declaring that decision quality has degraded, and whether a candidate has earned promotion.

Recent work has extended DFL beyond static batch training. Robust training and worst-case evaluation address distribution shift ex ante, while online, on-policy, and continual methods update a model as feedback arrives (Schutte et al., 2024; Ren et al., 2024; Capitaine et al., 2026; Benslimane et al., 2026). In parallel, deployment monitoring methods provide time-uniform risk control, selective label acquisition, sequential upgrade gates, or risk-inspired intervention controllers (Podkopaev & Ramdas, 2022; Timans et al., 2025; Ginart et al., 2022; Zrnic & Candès, 2024; Xu et al., 2024; Csillag et al., 2025; Karampatziakis et al., 2021; Lamaakal et al., 2026). No existing method addresses the complete deployment problem for DFL: detecting decision degradation under a finite

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.

label budget and deciding whether to promote a candidate, with finite-sample guarantees for both decisions.

Existing approaches leave three gaps: prediction errors are poor proxies for regret drift (Elmachtoub & Grigas, 2022); detection with selective labels must remain valid under arbitrary stopping while conditioning on the realized sample (Zhou et al., 2026); and an alarm alone does not justify promoting a data-driven replacement (Karampatziakis et al., 2021).

We propose E-REACT, a deployment loop that uses sparse feedback to detect decision degradation and fresh shadow comparisons to decide whether to deploy an update.

Contributions. **(i) A statistically guarded DFL deployment loop.** Inverse-propensity-corrected clipped-regret betting under a finite cost-vector ledger provides finite-sample, anytime-valid false-alarm control under an independently calibrated i.i.d. source null. A signed e-process evaluates candidates on fresh, fully audited decision disagreements, with a predeclared α-spending schedule controlling erroneous promotions per candidate and across sequential tests.

(ii) A theory of budgeted regret detection. We prove that scalar error monitors can be blind to regret drift; establish a worst-onset lower bound of order $\Omega(\log(1/\alpha)/(B\Delta^2))$ for every outcome-oblivious audit policy and match its dependence on the audit budget and signal strength, up to logarithmic factors, with a restarted mixture under matched calibration; and show that uniform auditing is the unique minimax static allocation against feature-marginal-preserving volatility reshaping. These results identify when regret-specific detection is necessary and how scarce labels should be allocated.

(iii) End-to-end empirical evidence. Across four combinatorial decision architectures and two histories, E-REACT detects decision-relevant shifts missed by scalar error monitoring and guards adaptation. Under forced alarms, the gate rejects all 200 null-trained candidates, whereas ungated deployment harms 194/200 streams; under genuine drift, guarded updates reduce post-onset regret by up to 26.1% without increasing endpoint regret in any of 200 paired synthetic streams. The financial and bike-sharing histories further confirm detection and guarded adaptation under real temporal dependence.

2 DEPLOYMENT FRAMEWORK

A decision-focused model is trained on source data and then frozen. During deployment, features are observed in every round, whereas true costs are observed only in audited rounds. E-REACT uses the audited outcomes to determine when to raise an alarm and a separate shadow sample to determine whether to replace the incumbent model.

Decision and regret. At round $t = 1, 2, \ldots$, a feature vector x_t arrives, the frozen predictor returns $\widehat{c}_t = f(x_t)$, and the downstream solver produces $w_t = w^\star(\widehat{c}_t) \in \arg\min_{w\in\mathcal{W}} \widehat{c}_t^\top w$. Maximization problems are handled by changing signs. If the true cost vector c_t is observed, the regret of the deployed decision and the monitored regret are

$$\rho_t(f) = c_t^\top w^\star(\widehat{c}_t) - \min_{w\in\mathcal{W}} c_t^\top w, \qquad Y_t(f) = \min\{\rho_t(f)/R_q, 1\} \in [0,1]. \tag{1}$$

The scale $R_q > 0$ is fixed on the calibration split cal-A (Section 3.1), before cal-B is used and before deployment: $R_q = \widehat{r}_q$, the empirical $q = 0.99$ quantile of source decision regret on cal-A, if $\widehat{r}_q > 0$, and $R_q = r_0$ otherwise, where $r_0 > 0$ is a task constant declared before any calibration data are drawn. Thus, the monitor tests the mean of a fixed clipped-regret functional.

Audit policy and budget. The true cost c_t is hidden unless round t is audited. Before observing c_t, the system chooses an audit probability p_t from the current feature, the current prediction and decision, and past observations. An independent audit coin then determines whether the outcome is revealed. On active rounds, let A_t indicate whether round t is audited:

$$A_t = \mathbf{1}\{U_t \le p_t\}, \qquad U_t \sim \mathrm{Unif}[0,1], \qquad p_t = \mathrm{clip}(\widehat{\eta} s(x_t), p_{\min}, p_{\max}). \tag{2}$$

The score $s \ge 0$ and normalizer $\widehat{\eta}$ are fixed using source data. Setting $s \equiv 1$ yields uniform auditing. Each audit consumes one label and requires one additional solver call to evaluate decision regret.

Information order. The audit probability p_t may depend on x_t because it is chosen before c_t is observed. The monitoring bet may depend on past audited outcomes but not on the current x_t. This order permits context-dependent auditing while keeping the monitoring test valid for a marginal source-regret bound. Section 3.1 gives the formal result.

Alarm and adaptation. E-REACT makes two decisions. First, the monitoring rule raises an alarm when the audited data provide sufficient evidence that mean clipped regret exceeds its calibrated source level. Monitoring continues until either an alarm is raised or all allotted detection audits have been used. Second, an alarm starts one adaptation attempt. A head-only candidate is trained on a post-alarm sample and compared with the incumbent on a fresh, fully audited shadow sample restricted to rounds where their decisions disagree. The incumbent remains deployed during this comparison. The candidate is promoted only if a signed e-process rejects the null that it is no better on the shadow regime.

3 E-REACT: Monitoring and Adaptation

3.1 Audited-Regret Monitoring

Let $\mathcal{F}_{t-1}$ contain the calibration data and all deployment information through round $t-1$. Let $\mathcal{G}_t = \mathcal{F}_{t-1} \vee \sigma(x_t, \widehat{c}_t, w_t, p_t)$ denote the information available after the audit probability is chosen but before the true cost is observed. Define the active-ledger indicator $I_t = \mathbf{1}\{\sum_{s<t} A_s < K_{\mathrm{det}}\}$ and draw $A_t = I_t \mathbf{1}\{U_t \le p_t\}$, where $U_t \sim \mathrm{Unif}[0,1]$ is conditionally independent of (c_t, Y_t) given $\mathcal{G}_t$. The Horvitz–Thompson observation and its centered version are

$$Z_t = \frac{A_t Y_t}{p_t}, \qquad X_t = Z_t - I_t \overline{\mu}_0, \qquad \mathbb{E}[Z_t \mid \mathcal{G}_t, Y_t] = I_t Y_t. \tag{3}$$

The last equality motivates inverse-propensity weighting. The mean of the audited outcomes depends on the audit policy, whereas Z_t targets regret on the original deployment stream. The policy changes the variance of the evidence. The bound $p_t \ge p_{\min} > 0$ also gives $Z_t \le 1/p_{\min}$.

The null level $\overline{\mu}_0$ is computed before deployment. Calibration split cal-A fixes the normalization R_q and hence the monitored functional Y. On the independent split cal-B, a one-sided empirical-Bernstein bound at level α_{cal} satisfies

$$\mathcal{C}_{\mathrm{cal}} = \{\mu_0 \le \overline{\mu}_0\}, \qquad \mu_0 = \mathbb{E}_{P_{\mathrm{src}}}[Y], \qquad \mathbb{P}(\mathcal{C}_{\mathrm{cal}}) \ge 1 - \alpha_{\mathrm{cal}}. \tag{4}$$

Let λ_t be $\mathcal{F}_{t-1}$-measurable and satisfy $0 \le \lambda_t \le (2\overline{\mu}_0)^{-1}$. Thus, the bet may depend on previous audited outcomes but not on the current feature. The base audited-regret e-process and its alarm time are

$$\mathcal{E}_0 = 1, \qquad \mathcal{E}_t = \prod_{s=1}^{t} (1 + \lambda_s X_s), \qquad \tau_{\mathrm{base}} = \inf\{t \ge 1 : \mathcal{E}_t \ge \alpha_{\mathrm{mon}}^{-1}\}. \tag{5}$$

Each factor is nonnegative because $X_t \ge -\overline{\mu}_0$. Once the ledger is exhausted, $I_t = A_t = X_t = 0$, so the process remains constant.

The best fixed bet depends on the unknown post-change margin. We therefore use a finite grid $\Lambda = \{\lambda_1, \dots, \lambda_m\}$ with weights $\pi_j > 0$ and $\sum_j \pi_j = 1$. A process started long before a change may also have small wealth at the change time. To avoid this dependence on the onset, we start copies at predeclared times $r \in \mathcal{S}$ and assign weights $\omega_r > 0$ with $\sum_r \omega_r \le 1$:

$$\mathcal{E}^{\mathrm{mix}}_{r,t} = \sum_{j=1}^{m} \pi_j \prod_{s=r}^{t} (1 + \lambda_j X_s), \qquad \tau_{\mathrm{rmix}} = \inf\left\{ t : \max_{r \in \mathcal{S}: r \le t} \omega_r \mathcal{E}^{\mathrm{mix}}_{r,t} \ge \alpha_{\mathrm{mon}}^{-1} \right\}. \tag{6}$$

Theorem 1 (Anytime validity under audited feedback)**.** *Assume $0 < p_{\min} \le p_t \le p_{\max} \le 1$ on active rounds and $\overline{\mu}_0 > 0$. Under the conditional-mean null*

$$H_0(\overline{\mu}_0): \qquad \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] \le \overline{\mu}_0 \quad \textit{for every } t,$$

$(\mathcal{E}_t)_{t \ge 0}$ is a nonnegative supermartingale conditionally on $\mathcal{F}_0$, and $\mathbb{P}(\tau_{\mathrm{base}} < \infty \mid \mathcal{F}_0) \le \alpha_{\mathrm{mon}}$. The restarted mixture in equation 6 likewise satisfies $\mathbb{P}(\tau_{\mathrm{rmix}} < \infty \mid \mathcal{F}_0) \le \alpha_{\mathrm{mon}}$. If the deployment pairs (x_t, c_t) are i.i.d. from P_{src} and independent of the calibration data, then

$$\mathbb{P}_{H_{\mathrm{src}}}(\tau_{\mathrm{rmix}} < \infty) \le \alpha_{\mathrm{cal}} + \alpha_{\mathrm{mon}} =: \alpha_{\mathrm{total}}. \tag{7}$$

To prove the theorem, take conditional expectations in equation 3. Under $H_0(\overline{\mu}_0)$, $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leq 0$, so each factor in equation 5 has conditional mean at most one. Ville's inequality gives the base crossing bound. Convex mixing preserves the supermartingale property, and the restart spending rule gives $\sum_r \alpha_{\mathrm{mon}}\omega_r \leq \alpha_{\mathrm{mon}}$. Under the i.i.d. source null, $\mathcal{C}_{\mathrm{cal}}$ implies the conditional-mean null, while calibration failure contributes at most α_{cal}. Thus, i.i.d. sampling is used for calibration; the sequential argument itself only needs the conditional-mean null. The same result holds for any bounded functional $\phi(\rho) \in [0,1]$ fixed before deployment, including regret-exceedance indicators, with separate error spending.

3.2 Detection Delay under an Audit Budget

We next study how the audit budget affects detection. An audit policy is outcome-oblivious if $p_t = g_t(x_t)$ for maps fixed before deployment. This class includes uniform, score-guided, and front-loaded policies but excludes policies that change audit intensity after observing audited outcomes. Let $\mathcal{D}(\kappa, M)$ denote the class of product-form alternatives that keep the feature marginal fixed and change only the conditional cost kernel after an unknown onset. For every round and feature x, the post-change kernel of c_t has Kullback–Leibler divergence at most κ from the null kernel and a log-likelihood ratio bounded by M in absolute value. We require the average audit rate to be at most B. An alarm time τ is admissible if it depends only on revealed information, namely the features, the audit coins, and the full cost vectors of audited rounds, and satisfies $\mathbb{P}_0(\tau \leq T) \leq \alpha$. This class contains every audited-regret monitor.

Theorem 2 (Budget–delay characterization)**.** *The following statements hold.*

(i) Lower bound. *Let $d^\star = \lfloor \log(1/\alpha)/(16B\kappa) \rfloor$ and assume $1 \leq d^\star \leq T/2$. For every outcome-oblivious audit policy and every admissible alarm time, at least half of the possible onsets ν satisfy the following condition for every alternative in $\mathcal{D}(\kappa, M)$:*

$$\mathbb{E}_\nu[(\tau - \nu)_+] \geq d^\star(1 - \alpha - \sqrt{\alpha} - \alpha^\gamma), \qquad \gamma = \frac{3\kappa}{24M^2 + 16\kappa M}. \tag{8}$$

Hence, $\sup_\nu \mathbb{E}_\nu[(\tau-\nu)_+] = \Omega(\log(1/\alpha)/(B\kappa))$ whenever $\alpha + \sqrt{\alpha} + \alpha^\gamma \leq 1/2$. Among deterministic expected-intensity profiles with average rate B, the constant profile is the unique profile whose expected audit mass is at least dB in every window of length d, simultaneously for all $d \geq 1$.

(ii) Matching upper bound. *Suppose the drift is regret-sufficient, that is, the conditional law of c_t given (x_t, Y_t) is unchanged. Suppose also that the conditional regret kernels are Bernoulli, their means lie in $[\varsigma, 1-\varsigma]$, and the mean increases from μ_0 to $\mu_0 + \Delta$ after an onset $\nu \leq T$. Then the cost kernels satisfy $\kappa \leq \Delta^2/(\varsigma(1-\varsigma))$ and $M \leq \Delta/\varsigma$, so part (i) gives a lower bound of order $\log(1/\alpha)/(B\Delta^2)$. Conversely, suppose $p_t = B$, the ledger is unbounded, and $\delta = \Delta - (\overline{\mu}_0 - \mu_0) > 0$. If a fixed-ratio bet grid contains a point within a fixed factor of $B\delta$, restart copies are launched every L rounds, and monitoring continues past T, then L can be chosen such that*

$$\sup_{\nu \leq T} \mathbb{E}_\nu[(\tau_{\mathrm{rmix}} - \nu)_+ \mid \mathcal{F}_\nu] \leq C\frac{\log(T/\alpha_{\mathrm{mon}}) + \log m}{B\delta^2} =: D^\infty. \tag{9}$$

When $\delta \geq c\Delta$ for a fixed $c > 0$, the lower and upper bounds match up to constant factors and the logarithmic grid and restart terms.

(iii) Finite ledger. *Let $K_\nu = K_{\mathrm{det}} - \sum_{s \leq \nu} A_s$ denote the number of audits remaining at the onset, and let $\tau_{\mathrm{cap}} = \inf\{t : \sum_{s \leq t} A_s = K_{\mathrm{det}}\}$. Under the conditions of part (ii), if $K_\nu \geq 2B$, then*

$$\mathbb{P}_\nu(\tau_{\mathrm{rmix}} \leq \tau_{\mathrm{cap}} \mid \mathcal{F}_\nu) \geq 1 - \frac{4BD^\infty}{K_\nu} - \exp(-3K_\nu/16). \tag{10}$$

The lower bound shows that the factor $1/B$ cannot be removed within the stated policy class, even by detectors that use full audited cost vectors. Under regret-sufficient drift, the restarted audited-regret mixture attains this rate, so access to full costs cannot improve it beyond constant and logarithmic factors. The upper bound follows from a grid component with $\lambda \asymp B\delta$: its expected log growth is of order $B\delta^2$, so reaching a logarithmic threshold requires order $1/(B\delta^2)$ rounds. Mixtures accommodate the unknown margin at a cost of $\log m$, while restarts accommodate the unknown onset at a cost of at most $\log T$. Part (iii) distinguishes the audit rate from the remaining audit

stock. The rate B controls the delay when auditing can continue, whereas K_ν determines whether enough audits remain to realize that delay. For a target failure probability β, it is sufficient that $K_\nu \geq \max\{8BD^\infty/\beta, (16/3)\log(2/\beta), 2B\}$.

3.3 Why Regret-Specific Detection Is Necessary

We now establish that regret detection cannot be replaced by scalar error-magnitude monitoring and characterize how the label probability should depend on the current context.

A limit of prediction-error monitoring. Consider a two-action decision problem. Let d_t denote the predicted cost gap and let g_t denote its signed prediction error. A scalar error-magnitude monitor is any sequential procedure measurable with respect to

$$\mathcal{H}_t^{\mathrm{acc}} = \sigma\{d_s, |g_s| : s \leq t\}. \tag{11}$$

This class includes CUSUM and windowed-mean monitoring procedures based on scalar error magnitude.

Theorem 3 (Separation between error and regret monitoring)**.** *For any $\gamma > d > 0$, there exist pre- and post-change distributions P_0 and P_1 over two-action decision problems such that (i) for every change time, the entire observation path $(d_t, |g_t|)_{t\geq 1}$ has the same law under the all-P_0 process and the changed process; (ii) with P_0 as the source law, source regret vanishes, so $R_q = r_0$ and the mean clipped regret increases by*

$$\Delta_{01} = \min\{(\gamma - d)/R_q, 1\} > 0; \tag{12}$$

and (iii) every level-α sequential test measurable with respect to $(\mathcal{H}_t^{\mathrm{acc}})$ has detection probability at most α, that is, no better than chance. If $\Delta_{01} > \overline{\mu}_0$ and the upper-bound conditions of Theorem 2(ii) hold (uniform rate B, unbounded ledger, grid coverage, monitoring past T), the audited-regret monitor has the corresponding delay upper bound at margin $\Delta_{01} - \overline{\mu}_0$.

The construction is direct. Under P_0, set $g = -\gamma$, so the true gap is $d + \gamma > 0$ and the predicted action remains optimal. Under P_1, set $g = +\gamma$, so the true gap is $d - \gamma < 0$ and the decision boundary is crossed. The observed values d and $|g|$ are identical, but regret changes from zero to $\gamma - d$ before clipping.

Oracle allocation for a fixed deployment law. Let Q be a deployment distribution and define

$$\sigma_Q(x) = \sqrt{\mathbb{E}_Q[Y^2 \mid X = x]}, \qquad V(p; Q) = \mathbb{E}_Q\left[\frac{Y^2}{p(X)}\right]. \tag{13}$$

The quantity $V(p; Q)$ is the second moment of the Horvitz–Thompson observation and controls the growth rate of the monitor. For a positive score s, consider the proportional allocation $p_s(x) = Bs(x)/\mathbb{E}_Q[s(X)]$.

Proposition 1 (Oracle allocation and exact excess)**.** *Assume that $0 < B \leq 1$, $\sigma_Q > 0$ almost surely, $0 < \mathbb{E}_Q[\sigma_Q] < \infty$, and $s > 0$ with $0 < \mathbb{E}_Q[s] < \infty$ and $\mathbb{E}_Q[\sigma_Q^2/s] < \infty$. Suppose that $p_s \leq 1$ and $p_{\sigma_Q} \leq 1$ almost surely. Among feasible proportional allocations, $V(p; Q)$ is minimized by*

$$p^\star(x) = B\frac{\sigma_Q(x)}{\mathbb{E}_Q[\sigma_Q(X)]}, \qquad V^\star(Q) = \frac{\mathbb{E}_Q[\sigma_Q(X)]^2}{B}. \tag{14}$$

For any feasible score s, write $\sigma_Q(x) = s(x)(1 + \varepsilon(x))$ and define $d\mu_s = s\, dQ/\mathbb{E}_Q[s]$. Then

$$V(p_s; Q) - V^\star(Q) = \frac{\mathbb{E}_Q[s]^2}{B}\operatorname{Var}_{\mu_s}(\varepsilon), \qquad \frac{V(p_{\mathrm{unif}}; Q)}{V^\star(Q)} = 1 + \mathrm{CV}_Q(\sigma_Q)^2. \tag{15}$$

If the proportional oracle is infeasible, the constrained optimum is the capped rule $p^\star(x) = \min\{1, c\sigma_Q(x)\}$, where the constant c is chosen to meet the budget.

The proposition quantifies the exact value of knowing the deployment-law conditional root second moment. It also separates score estimation from score transfer. For a source-fitted score $\widehat{\sigma} > 0$, let

$d\mu_{\widehat{\sigma}} = \widehat{\sigma}\, dQ/\mathbb{E}_Q[\widehat{\sigma}]$. Then

$$\begin{aligned} \varepsilon &= \frac{\sigma_Q - \widehat{\sigma}}{\widehat{\sigma}} = a + b, \\ a &= \frac{\sigma_{\mathrm{src}} - \widehat{\sigma}}{\widehat{\sigma}} \quad \text{(estimation)}, \qquad\qquad b = \frac{\sigma_Q - \sigma_{\mathrm{src}}}{\widehat{\sigma}} \quad \text{(transfer)}, \\ \mathrm{Var}_{\mu_{\widehat{\sigma}}}(\varepsilon) &= \mathrm{Var}_{\mu_{\widehat{\sigma}}}(a) + \mathrm{Var}_{\mu_{\widehat{\sigma}}}(b) + 2\,\mathrm{Cov}_{\mu_{\widehat{\sigma}}}(a, b), \\ &\le 2\,\mathrm{Var}_{\mu_{\widehat{\sigma}}}(a) + 2\,\mathrm{Var}_{\mu_{\widehat{\sigma}}}(b). \end{aligned} \tag{16}$$

Even if $\widehat{\sigma} = \sigma_{\mathrm{src}}$, the allocation has positive excess whenever $\sigma_Q/\sigma_{\mathrm{src}}$ is nonconstant. A uniform rescaling of the volatility profile causes no transfer error; a context-dependent reshaping does.

Minimax allocation under reshaping drift. The result assumes that Q is known. We now allow drift to change where regret volatility occurs. Let $(\mathcal{X}, \mu)$ be finite with equal weights or nonatomic standard, fix $\overline{\sigma} : \mathcal{X} \to [0, 1]$ with $\mathbb{E}_\mu[\overline{\sigma}^2] > 0$, and let $\mathcal{Q}(\overline{\sigma})$ be the deployment laws with feature marginal μ for which σ_Q has the same distribution as $\overline{\sigma}$. Define the class of all static budgeted samplers

$$\mathcal{P}(B) = \{p : \mathcal{X} \to (0, 1] \text{ measurable} : \mathbb{E}_\mu[p(X)] \le B\}. \tag{17}$$

This class contains proportional, capped, and clipped static samplers.

Theorem 4 (Uniform auditing is minimax under reshaping)**.** *For every $p \in \mathcal{P}(B)$,*

$$\sup_{Q \in \mathcal{Q}(\overline{\sigma})} V(p; Q) \ge \mathbb{E}_\mu[\overline{\sigma}^2]\mathbb{E}_\mu[1/p] \ge \frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B} = V(p_{\mathrm{unif}}; Q), \qquad Q \in \mathcal{Q}(\overline{\sigma}). \tag{18}$$

Equality throughout holds if and only if $p(x) = B$ for μ-almost every x. Therefore, uniform auditing is the unique minimax sampler in this drift class. Every nonuniform sampler has worst-case excess of at least

$$(B\mathbb{E}_\mu[1/p] - 1)\,\frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B}. \tag{19}$$

Proposition 1 and Theorem 4 answer different questions. If the deployment law is known, sampling in proportion to σ_Q is optimal. If only the source law is known and drift may rearrange the volatility profile, uniform auditing is uniquely minimax. The difference is whether the sampler knows where high-volatility contexts will occur after drift. This is why E-REACT uses uniform auditing as its default.

3.4 Guarded Adaptation

An alarm is evidence that the incumbent has degraded; it is not evidence that a particular update is better. E-REACT therefore separates detection from promotion. The total label ledger is divided as $K_{\mathrm{tot}} = K_{\mathrm{det}} + K_{\mathrm{train}} + K_{\mathrm{shadow}}$. After an alarm, the protocol (i) collects K_{train} audited post-alarm examples at a predictable rate; (ii) trains one head-only candidate f' while keeping the representation fixed; (iii) keeps the incumbent deployed and compares the two models on a fresh shadow sample; and (iv) promotes the candidate only if the shadow test crosses its threshold. The promotion result below does not depend on the candidate training method, provided that the candidate is fixed before shadow evaluation begins.

The shadow test uses only rounds on which the two models produce different decisions. Define the pre-outcome disagreement indicator

$$S_t = \mathbf{1}\{w^\star(\widehat{c}_t) \ne w^\star(\widehat{c}'_t)\}. \tag{20}$$

Because S_t is computed from x_t and the two predictions before c_t is observed, disagreement selection does not use the current outcome. Shadow rounds with $S_t = 1$ are fully audited. Let i index these audited disagreements and define

$$D_i = Y_i(f) - Y_i(f') \in [-1, 1], \qquad \mathcal{G}_m = \prod_{i=1}^{m}(1 + \lambda_i^{\mathrm{sh}} D_i), \qquad 0 \le \lambda_i^{\mathrm{sh}} \le 1. \tag{21}$$

Positive D_i favors the candidate. The bet λ_i^{sh} is measurable before D_i is observed. Candidate k is promoted if $\mathcal{G}_m \ge 1/\alpha_k$ before its shadow ledger is exhausted. Agreement rounds contribute $D = 0$ and provide no comparison evidence, so the disagreement restriction avoids spending shadow labels on zero increments.

Theorem 5 (Per-candidate and lifetime promotion control)**.** *Let $\mathcal{F}^{\mathrm{sh}}_{i-1}$ denote the selected-sequence history before the ith audited disagreement, including the fixed candidate and its training randomness. The following statements hold.*

(i) Per candidate. *Under the conditional null*

$$H^{\mathrm{sh}}_{0,k}: \qquad \mathbb{E}[D_i \mid \mathcal{F}^{\mathrm{sh}}_{i-1}] \le 0 \quad \textit{for every } i,$$

the probability of promoting candidate k is at most α_k.

(ii) Deployment lifetime. *Suppose candidates are generated sequentially and may depend on all data collected before their shadow phases. If the candidate-specific levels follow a predeclared spending rule $\sum_{k\ge 1}\alpha_k \le \alpha_{\mathrm{dep}}$, then the probability that any candidate satisfying its null $H^{\mathrm{sh}}_{0,k}$ is ever promoted is at most α_{dep}. The expected number of such promotions is also at most α_{dep}.*

For part (i), each factor in equation 21 is nonnegative and has conditional mean at most one, so $(\mathcal{G}_m)$ is a nonnegative supermartingale, and Ville's inequality yields the bound. Part (ii) follows by summing the conditional error probabilities over the spending sequence; adaptive candidate generation and a random number of alarms do not affect the result. Define the raw-time difference as $D^{\mathrm{raw}}_t = S_t(Y_t(f) - Y_t(f'))$, and let $\mathcal{G}^{\mathrm{sh}}_t$ denote the shadow information available before the cost is observed. The conditional null $\mathbb{E}[D^{\mathrm{raw}}_t \mid \mathcal{G}^{\mathrm{sh}}_t] \le 0$ implies the selected-sequence null because S_t is $\mathcal{G}^{\mathrm{sh}}_t$-measurable.

4 Experiments

We evaluate four questions suggested by the theory: whether regret monitoring detects changes that error monitoring misses, how detection delay scales with the audit budget, whether a source-fitted audit score transfers after drift, and whether the promotion gate prevents harmful updates. Additional experiments and protocol details are in Section A.

Setting. We use four predict-then-optimize problems with exact solvers: shortest path (SP), knapsack (KP), top-k selection (TK), and bipartite assignment (AS). Predictors are two-layer MLPs trained with SPO+ on $2{,}000$ source samples. Each synthetic stream has $T = 4{,}000$ rounds and changes at $t = 2{,}000$ under an abrupt or gradual concept or cost-structure drift. Unless stated otherwise, the audit budget is $B = 5\%$, the label ledger is $K = BT$, and the monitor is the restarted mixture in Section 3.1. Synthetic results use five independently generated problem instances, and confidence intervals resample instances. We also study FF30 industry selection and Citi Bike station allocation on fixed historical paths.

Benchmarks. We compare the regret e-process with CUSUM monitors of scalar absolute prediction error and of audited regret, whose thresholds are calibrated to a 5% false-alarm rate over the horizon; uniform auditing with oracle and source-fitted allocations at the same audit rate, and guarded adaptation with periodic, oracle-timed, and ungated updates. Paired comparisons use the same deployment streams and audit coins. A non-detection is censored at the horizon when computing restricted mean time to alarm (RMTA).

Results. With fixed calibration, crossing rates are 0–2% across 36 null cells, and redrawing calibration in every replication gives 0–4 crossings in 200 streams per cell. Repeated standardized-mean testing on the same audited streams crosses in 28% of null runs. These results agree with Theorem 1: the e-process remains valid under continuous monitoring, while repeated fixed-time tests do not. False-alarm control persists under real temporal dependence: on stationary block-bootstrap nulls built from calm FF30 returns, where the lag-1 autocorrelation of Y reaches 0.205, the largest crossing count across 18 cells is 6 of 600 against an allowance of 30, and the restarted mixture crosses at most 3 times (Section A.4).

Figure 1 also verifies the separation established in Theorem 3. On that construction, the scalar error process has the same law before and after the change, so error CUSUM remains at its null detection rate, whereas the regret monitor detects every changed stream. The same pattern occurs without a constructed symmetry on assignment: the regret monitor detects $50/50$ abrupt and $45/50$ gradual streams, compared with $1/50$ and $3/50$ for error CUSUM. The clipped functional preserves

this signal: in all 15 problem–drift cells, raw regret, the clipped mean Y, and the $q_{.90}$ exceedance rate rise together after the onset, even where 64% of post-onset audits sit at the clipping ceiling (Section A.6).

The regret signal carries this detection power. A regret CUSUM on the same audits, with its threshold bootstrapped from the calibration sample, detects 714 of 750 synthetic streams at $B = 5\%$, against 100 for error CUSUM and 586 for the e-process (Table 7). The two regret procedures trade delay against the scope of their guarantee: the CUSUM detects sooner (mean RMTA 587 against 1,051) under a false-alarm rate calibrated for the horizon T, whereas the e-process certifies its level at every horizon. Monitored to $4T$, the CUSUM crosses in 306 of 1,500 null streams and the e-process in 7; on the two event-free histories whose mean clipped regret stays below $\overline{\mu}_0$, the CUSUM alarms in up to 6 of 20 runs and the e-process in none (Table 8).

Predeclared restarts make detection robust to late onsets. With the onset moved to $\nu = 3{,}200$, the restarted mixture detects 217 of 300 streams (72.3%), whereas an unrestarted mixture, whose wealth has decayed over the null prefix, detects 34 (11.3%); across the fifteen cells of the primary grid, restarts lower RMTA in every cell (Section A.3).

Detection delay shortens with the audit budget close to the $1/B$ rate of Theorem 2. A log–log fit of median delay on budget and margin gives a budget exponent of -0.835, against the predicted -1; at the largest margin, raising B from 1% to 10% shortens the median delay from 1,124.5 to 272.5 rounds while detection rises from 16/50 to 50/50. The finite ledger separates audit rate from audit stock, as in part (iii): from $B = 1\%$ to 5%, the streams that exhaust the ledger before crossing fall from 29/150 to 0/150 under abrupt cost-structure drift and from 82/150 to 45/150 under gradual concept drift, whereas the share that exhausts the ledger at some point stays at 51.5% and 53.5% (Section A.5).

The allocation results match Proposition 1 and Theorem 4. When the conditional root second moment under the deployment law is known, the rate-matched oracle improves pooled RMTA by 139 steps. The gains are 157 steps on SP, 229 on KP, and 31 on TK. In contrast, the source-fitted score yields a pooled change of -4.1 steps relative to uniform auditing, with a confidence interval spanning zero. The oracle result shows that informative allocation can help, whereas the source result shows that the required score need not transfer after drift. We therefore use uniform auditing as the default and report the full intensity–allocation decomposition in Section A.

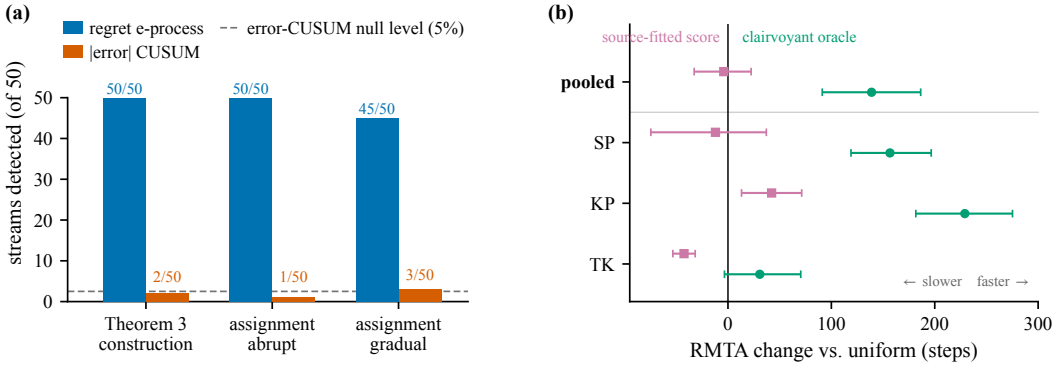

Figure 1: Monitoring and audit allocation at $B = 5\%$. Left: the regret e-process detects all 50 streams of the Theorem 3 construction and 50/50 abrupt and 45/50 gradual assignment streams, whereas error CUSUM detects 2/50, 1/50, and 3/50, respectively. Right: at the same audit rate, the clairvoyant oracle reduces pooled RMTA relative to uniform auditing by 139 steps (95% CI $[91, 186]$), while the source-fitted score changes pooled RMTA by -4.1 steps (95% CI $[-33, 22]$). Error bars are instance-level cluster-bootstrap confidence intervals.

Table 1 shows why an alarm should not directly trigger an update. Guarded adaptation reduces post-onset regret by 3.1–26.1% and records no harmful endpoints in any of 200 paired streams. The contrast under gradual drift is stark: every oracle-timed update harms the stream (50/50 on SP, 45/50 on KP), whereas guarded adaptation harms none. The gate also rejects all 200 candidates

Table 1: Guarded adaptation under a common 200-label ceiling, allocated as in Algorithm 1. Reduction is the all-stream decrease in post-onset regret relative to no adaptation. Harmful means that post-onset cumulative regret exceeds no adaptation. Periodic and oracle-timed updates do not use the promotion gate.

Problem	Drift	Alarms/ promotions	Guarded reduction	Guarded harmful	Periodic harmful	Oracle harmful
SP	Abrupt	$48/24$	$+\mathbf{18.5}\%$	$\mathbf{0/50}$	$0/50$	$0/50$
SP	Gradual	$28/11$	$+\mathbf{3.1}\%$	$\mathbf{0/50}$	$28/50$	$50/50$
KP	Abrupt	$45/33$	$+\mathbf{26.1}\%$	$\mathbf{0/50}$	$0/50$	$0/50$
KP	Gradual	$24/16$	$+\mathbf{8.1}\%$	$\mathbf{0/50}$	$22/50$	$45/50$

trained after forced pre-onset alarms; deploying the same candidates without the gate harms $194/200$ streams. The protection holds across repeated interventions, weak candidate operators, and the gate's free constants. Under up to three forced alarms per stream with the spending schedule $\alpha_k = \alpha_{\text{dep}} 2^{-k}$, the gate makes no null promotion across 200 streams, against 600 ungated. A decision-blind ridge head fitted to squared error is harmful on 50 of 200 streams when deployed ungated and on none behind the gate, and the harmful-endpoint count stays at 0 in all 48 settings of α_{dep} and the shadow cap. On bipartite assignment, with adaptation labels budgeted separately from the detection ledger, the gate promotes 41 of 50 abrupt-drift candidates without a harmful endpoint and reduces post-onset regret by 38.0% (Section A.8).

Real data. On FF30, the restarted mixture gives no alarm in 40 calm-window runs at $B = 5\%$. Detection on COVID rises from $1/20$ to $17/20$ and detection on the GFC rises from $1/20$ to $18/20$ as the budget increases from 5% to 20%. On Citi Bike, detection rises from $7/20$ to $15/20$ over the same budgets (Table 8). The choice of null matters on real data: on the calm 2014–2016 FF30 window, an online conformal exchangeability martingale on the same audit stream alarms in 7, 11, and 17 of 20 runs at $B = 5\%$, 10%, and 20%, against 0, 1, and 3 for the restarted mixture. On the eight Citi Bike candidates trained after plug-in e-process alarms (Section A.10), the gate promotes $6/8$ and reduces post-event regret by 7.3% on average; a strict level $\alpha_{\text{dep}} = 0.01$ gives $5/8$ at 7.2%. On FF30, the gate rejects all four post-alarm candidates, whereas ungated deployment of the same candidates produces one harmful endpoint. The historical results confirm that E-REACT detects real decision-relevant events and that guarded adaptation transfers to dependent, non-stationary histories.

5 Conclusion

We develop E-REACT for detecting regret drift and guarding adaptation in deployed decision-focused models. The framework separates evidence that the incumbent has degraded from evidence that a candidate improves on it. An audited-regret e-process controls false alarms, and an independent shadow test controls erroneous promotions. The error-regret separation proves that scalar error-magnitude monitoring cannot substitute for regret-specific detection, while the budget–delay characterization quantifies how label budgets translate into detection speed. The minimax allocation result explains why uniform sampling is optimal when drift can reshape the conditional regret-volatility profile.

AI use statement

Generative AI tools were used for language editing, figure polishing, and limited code assistance. All AI-assisted outputs were reviewed and verified by the authors. The authors take full responsibility for the scientific content, results, and conclusions of this work.

Reproducibility statement

We provide the code, configurations, and scripts needed to reproduce the experiments and reported results. All datasets used in this work are publicly available, and the supplementary material contains the implementation details and experimental settings required for reproduction.

REFERENCES

Wyame Benslimane, Tinghan Ye, Pascal Van Hentenryck, and Paul Grigas. Decision-Focused On-Policy Learning for Contextual Linear Optimization with Partial Feedback. *arXiv preprint arXiv:2606.01081*, 5 2026. URL https://arxiv.org/abs/2606.01081.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Aymeric Capitaine, Maxime Haddouche, Eric Moulines, Michael I. Jordan, Etienne Boursier, and Alain Durmus. Online Decision-Focused Learning. In *International Conference on Learning Representations*, 4 2026. URL https://arxiv.org/abs/2505.13564.

Daniel Csillag, Claudio José Struchiner, and Guilherme Tegoni Goedert. Prediction-Powered E-Values. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 11493–11514. PMLR, 7 2025. URL https://proceedings.mlr.press/v267/csillag25a.html.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based End-to-end Model Learning in Stochastic Optimization. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5484–5494. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3fc2c60b5782f641f76bcefc39fb2392-Paper.pdf.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 5 edition, 2019.

Adam N. Elmachtoub and Paul Grigas. Smart "Predict, then Optimize". *Management Science*, 68 (1):9–26, 1 2022. ISSN 15265501. doi: 10.1287/mnsc.2020.3922. URL https://arxiv.org/abs/1710.08005.

David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118, 1975.

Antonio A. Ginart, Martin Jinye Zhang, and James Zou. MLDemon: Deployment Monitoring for Machine Learning Systems. *Proceedings of Machine Learning Research*, 151:3962–3997, 3 2022. ISSN 26403498. URL https://arxiv.org/pdf/2104.13621.

G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 2 edition, 1952.

Nikos Karampatziakis, Paul Mineiro, and Aaditya Ramdas. Off-policy Confidence Sequences. *Proceedings of Machine Learning Research*, 139:5301–5310, 7 2021. ISSN 26403498. URL https://arxiv.org/pdf/2102.09540.

Ismail Lamaakal, Chaymae Yahyati, Khalid El Makkaoui, Ibrahim Ouahbi, and Yassine Maleh. Drift-to-Action Controllers: Budgeted Interventions with Online Risk Certificates. *arXiv preprint arXiv:2603.08578*, 3 2026. URL https://arxiv.org/abs/2603.08578.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-Focused Learning: Foundations, State of the Art, Benchmark and Future Opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 8 2024. doi: 10.1613/jair.1.15320. URL https://arxiv.org/abs/2307.13565.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample-variance penalization. In *Proceedings of the 22nd Conference on Learning Theory (COLT)*, 2009.

Aleksandr Podkopaev and Aaditya Ramdas. Tracking the risk of a deployed model and detecting harmful distribution shifts. *ICLR 2022 - 10th International Conference on Learning Representations*, 4 2022. URL https://arxiv.org/pdf/2110.06177.

Kevin Ren, Yewon Byun, and Bryan Wilder. Decision-Focused Evaluation of Worst-Case Distribution Shift. *Proceedings of Machine Learning Research*, 244:3076–3093, 7 2024. ISSN 26403498. URL https://arxiv.org/pdf/2407.03557.

John V. Ryff. Measure preserving transformations and rearrangements. *Journal of Mathematical Analysis and Applications*, 31(2):449–458, 1970.

Noah Schutte, Krzysztof Postek, and Neil Yorke-Smith. Robust Losses for Decision-Focused Learning. *IJCAI International Joint Conference on Artificial Intelligence*, pp. 4868–4875, 8 2024. doi: 10.24963/ijcai.2024/538. URL https://www.ijcai.org/proceedings/2024/538.

Alexander Timans, Rajeev Verma, Eric Nalisnick, and Christian A. Naesseth. On Continuous Monitoring of Risk Violations under Unknown Shift. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 4204–4226. PMLR, 7 2025. URL https://proceedings.mlr.press/v286/timans25a.html.

Jean Ville. *Étude critique de la notion de collectif*. Monographies des Probabilités. Gauthier-Villars, Paris, 1939.

Ziyu Xu, Nikos Karampatziakis, and Paul Mineiro. Active, anytime-valid risk controlling prediction sets. In *Advances in Neural Information Processing Systems*, volume 37, pp. 60110–60132. Curran Associates, Inc., 2024. doi: 10.52202/079017-1920. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/6eb05d8bc6bd7bb6868c64b5802125bd-Abstract-Conference.html.

Siyu Zhou, Patrick Vossler, Venkatesh Sivaraman, Yifan Mai, and Jean Feng. Adaptive auditing of AI systems with anytime-valid guarantees. *arXiv preprint arXiv:2605.07002*, 5 2026. URL https://arxiv.org/abs/2605.07002.

Tijana Zrnic and Emmanuel J. Candès. Active Statistical Inference. *Proceedings of Machine Learning Research*, 235:62993–63010, 7 2024. ISSN 26403498. URL https://arxiv.org/pdf/2403.03208.

A Experimental details and additional results

A.1 Setup and protocol

Decision architectures. We evaluate four predict-then-optimize problems with different feasible-set geometries and exact solvers. Reported regret is therefore an exact optimality gap. Shortest path (SP) routes on a 5×5 grid with 40 directed edges and a batched dynamic program over lattice paths. Knapsack (KP) is a $0/1$ maximization over 20 items with public integer weights drawn once and capacity 60, solved by a batched capacity DP. Top-k selection (TK) picks 8 of 30 assets, solved exactly by sorting. Bipartite assignment (AS) matches 6 workers to 6 tasks: the feasible set is the 720 permutation matrices over 36 cost coordinates and the solver is the Hungarian algorithm.

All four use the same synthetic generator: $P = 5$ standard normal features, a sparse random loading matrix B, a degree-2 polynomial link $c = ((XB^\top/\sqrt{P}) + 3)^2 + 1$, and coordinate-wise multiplicative noise drawn uniformly from $[0.75, 1.25]$. An instance is defined by one independent draw of the loading matrix, the drift perturbation, and the predictor seed.

Predictors and candidates. Each instance trains a two-layer MLP on $2{,}000$ source samples with the SPO+ surrogate for 30 epochs, after which the resulting predictor is frozen for deployment. A candidate produced after an alarm refits the output head for 150 epochs while holding the representation fixed, so candidate training requires one head fit and no representation training. Section A.8 additionally uses a deliberately decision-blind candidate operator: a closed-form ridge head with $\lambda = 0.01$ fitted to squared prediction error, which carries no information about the downstream solver.

Calibration splits. Calibration uses $n_{\mathrm{cal}} = 2{,}000$ source samples split in half. Split cal-A fixes the normalization R_q at $q = 0.99$ and hence the monitored functional Y in Equation (1). We use a predeclared task-specific fallback $r_0 > 0$ if the cal-A quantile is zero. For all reported synthetic and historical calibrations other than the construction of Theorem 3, we set $r_0 = 1$ in the task's cost units; the quantile is positive in every such calibration (smallest value 0.096), so the fallback is never

activated. For that construction, we predeclare $r_0 = \gamma = 2.5$; since source regret is identically zero, the fallback is activated and $R_q = r_0 = 2.5$, while $\overline{\mu}_0 = 0.0043$ is the one-sided empirical-Bernstein bound on 2,000 independent P_0 draws. Cal-A also fits the audit-score normalizer $\widehat{\eta}$ in Equation (2). The independent split cal-B provides the one-sided empirical-Bernstein upper confidence bound $\overline{\mu}_0$ at $\alpha_{\mathrm{cal}} = 0.05$, which is the event $\mathcal{C}_{\mathrm{cal}}$ in Equation (4). Because Y is fixed on cal-A and bounded on cal-B, the two splits together yield the union bound $\alpha_{\mathrm{total}} = \alpha_{\mathrm{cal}} + \alpha_{\mathrm{mon}} = 0.10$ in Theorem 1; the fixed-calibration tables isolate the α_{mon} term, while the recalibration study in Section A.2 exercises the full chain.

Audit policy and ledger. On active rounds the audit probability is $p_t = \mathrm{clip}(\widehat{\eta}\, s(x_t), p_{\min}, p_{\max})$ with $p_{\min} = B/5$ and $p_{\max} = 0.5$. Uniform auditing uses $s \equiv 1$ and is the default; the score-based arms use the decision margin, the source-fitted second-moment score $\widehat{\sigma}$ (the source-fitted arm in the main text), an online-refit score (reported as exploratory), and the clairvoyant per-phase oracle score, which is recalibrated at every distinct drift fraction in each cell and therefore presupposes knowledge of the onset and the post-change law. In the detection experiments the ledger contains $K_{\mathrm{det}} = \mathrm{round}(BT)$ audits; the update arms reserve part of BT for training and shadow labels (Algorithm 1). Once the detection ledger is exhausted, the active indicator I_t is zero and the e-process remains constant, yielding the stopped process to which Theorem 1 applies.

Monitors. Three arms consume the same audited outcomes on the same audit stream. The plug-in arm uses a GRAPA-style predictable bet computed from a sliding window of 300 past audited observations after a warm-up of 20 and clipped to $\lambda_t \in [0, (2\overline{\mu}_0)^{-1}]$. The mixture arm uses the fixed grid in Equation (6), with $J = 24$ logarithmically spaced bets spanning three decades below $(2\overline{\mu}_0)^{-1}$, and requires no tuning. The restarted mixture, which is the deployed monitor, launches a fresh mixture copy every $L = 1000$ rounds and compares each copy with the per-copy threshold $(R+1)/\alpha_{\mathrm{mon}}$, where $R = \lceil T/L \rceil$, so a union bound over the $R+1$ copies yields the nominal any-crossing level. The copies share audits and consume no additional labels. Two predeclared tail functionals, the exceedance indicators at $q_{.99}$ and $q_{.90}$, are monitored using the same procedure with separate error spending, as permitted by Theorem 1.

Promotion gate. An alarm starts one adaptation attempt: $K_{\mathrm{train}} = 50$ audited post-alarm examples collected at a predictable rate are used to train the candidate, which is then compared with the incumbent on a fresh shadow sample capped at $K_{\mathrm{shadow}} = 30$ labels and restricted to the disagreement rounds of Equation (20). The signed e-process in Equation (21) uses a predictable bet in $[0, 1]$ after a warm-up of five increments and promotes the candidate when it reaches $1/\alpha_k$. Single-intervention runs spend $\alpha_1 = \alpha_{\mathrm{dep}} = 0.05$; the multi-intervention runs in Section A.8 use the geometric spending rule $\alpha_k = \alpha_{\mathrm{dep}} 2^{-k}$, that is, $(0.025, 0.0125, 0.0063)$ for up to three cycles, which satisfies the predeclared budget in Theorem 5(ii).

Label accounting. Algorithm 1 gives the per-stream label accounting of the update arms in Tables 1, 12 and 18 and of the gate sweep. Every arm shares the ceiling $K = BT = 200$, and the guarded and ungated arms reserve it in advance as $K_{\mathrm{det}} = 120$, $K_{\mathrm{train}} = 50$, and $K_{\mathrm{shadow}} = 30$. Detection audits uniformly at the paced rate K_{det}/T and runs the restarted mixture; it stops at the first alarm, and detection labels left unused at that point are not reassigned. Because the three allotments sum to K, a late alarm never leaves training or the shadow sample short of labels; only the end of the stream truncates them. The forced-alarm arms set $\tau = 1{,}000$ and charge the 30 labels that detection spends in expectation by then. The scheduled and oracle-timed arms have no detection or shadow stage and are bound only by the ceiling. The multi-intervention runs of Table 13 ($T = 6{,}000$, $K = 300$) make no reservation: detection, training, and shadow draw on one running balance, and each cycle's detection uses the labels that remain. On the real histories and on bipartite assignment, detection keeps its full ledger $K_{\mathrm{det}} = \mathrm{round}(BT)$, and the adaptation labels form a separate budget of at most $K_{\mathrm{train}} + K_{\mathrm{shadow}} = 80$.

Streams, drift, and replication. Each synthetic stream has $T = 4{,}000$ rounds with onset $\nu = 2{,}000$, a gradual ramp of 1,500 rounds where applicable, and cost laws regenerated in chunks of 100 rounds. Concept drift rotates the feature-to-cost mapping, whereas cost-structure drift perturbs the loadings themselves. A late-onset cell sets $\nu = 3{,}200$ to evaluate monitors near the end of the horizon, and the multi-intervention runs in Section A.8 extend to $T = 6{,}000$ with onsets at 2,000 and

Algorithm 1 Per-stream label accounting of the update arms. Every audit increments the label count ℓ, which starts at 0 and never exceeds $K = BT = 200$.

Require: $K_{\mathrm{det}} = 120$, $K_{\mathrm{train}} = 50$, $K_{\mathrm{shadow}} = 30$, shadow window $W = 800$

Guarded and ungated arms

1: Audit each round with probability K_{det}/T and update the restarted mixture; stop at the first alarm τ or when $\ell = K_{\mathrm{det}}$. If no alarm occurs, keep f and stop.
2: From round $\tau + 1$, audit each round with probability $1/2$ and add it to the training set $\mathcal{I}$ until $|\mathcal{I}| = K_{\mathrm{train}}$, $\ell = K$, or the stream ends. If $|\mathcal{I}| < 10$, keep f and stop.
3: Fit the head-only candidate f' on $\mathcal{I}$ with inverse-propensity weights. Ungated arm: deploy f' from the next round and stop.
4: Guarded arm: from the round after training stops, scan at most W rounds, audit every round with $S_t = 1$, and update $\mathcal{G}$; stop at promotion, after K_{shadow} shadow labels, when $\ell = K$, or at the end of the window.
5: If f' is promoted, deploy it from the next round; otherwise keep f.

Scheduled arm

6: Audit each round with probability B while $\ell < K$. At $t \in \{T/4, T/2, 3T/4\}$, if at least 20 labels exist, refit the head on the latest 50 and deploy it without a gate.

Oracle-timed arm

7: From the true onset ν, audit each round with probability $1/2$ until 50 labels are collected; fit the head and deploy it without a gate.

4,000. Every cell uses five independent instances $\times$ ten stream seeds, yielding 50 drifted streams per cell, and forty null streams per instance, yielding 200 null streams per cell. All confidence intervals are obtained by resampling the five instances with replacement (5,000 cluster-bootstrap draws, percentile interval), so they quantify uncertainty across problem instances rather than across streams from a single instance.

Censoring convention. Restricted mean time to alarm (RMTA) is the mean over *all* streams of $\min\{\tau - \nu, H\}$ at horizon $H = T - \nu$; both a stream that never alarms and a stream whose alarm precedes the onset contribute the full horizon H. RMTA is the primary detection summary throughout because it prices non-detection explicitly. Medians conditional on post-onset detection are reported alongside as descriptive statistics, together with the detection counts underlying each median.

Real histories. The FF30 study uses the value-weighted daily section of the Ken French 30-industry portfolio file and deploys a top-8-of-30 selection. Its four windows are COVID (train 2010–2015, calibrate 2016–2017, deploy 2018-01-02 to 2021-12-31, $T = 1{,}008$, reference event 2020-02-20), GFC (train 1995–2003, calibrate 2004–2005, deploy 2006-01-03 to 2009-12-31, $T = 1{,}007$, reference event 2008-09-15), and two event-free windows, calm14 (deploy 2014-01-02 to 2016-12-30, $T = 756$) and calm93 (deploy 1993-01-04 to 1996-12-31, $T = 1{,}011$). The Citi Bike Jersey City study uses 26 name-bridged stations active in at least 72 of 76 months, predicts each station's daily share of departures, and deploys a top-7 allocation. Its windows are COVID (train 2016–2018, calibrate 2019, deploy 2020-01-01 to 2021-12-31, $T = 730$, reference event 2020-03-16) and a calm 2019 window ($T = 365$). The spatial share is the natural value scale here because seasonal volume fluctuations cancel, and top-k selection based on shares coincides with top-k selection based on counts. Each real cell is replicated using five predictor seeds $\times$ four audit seeds, giving 20 runs, and the adaptation labels ($K_{\mathrm{train}} + K_{\mathrm{shadow}}$) constitute a separate budget in addition to the monitoring ledger.

A.2 Anytime validity of the audited-regret monitor

The stopped regret e-process controls the probability of ever raising a false alarm at its nominal level under every audit-allocation rule we run, at both budgets, and on all three synthetic predict-then-optimize problems, whether the calibration is held fixed or redrawn in every replication. Figure 2 and Table 2 report the evidence.

Across the 36 fixed-calibration null cells, comprising six monitor and audit-allocation configurations $\times$ three problems $\times$ two budgets, each with 200 null streams, the largest any-crossing rate is 0.02 (4

of 200, for KP at $B = 1\%$ with the mixture monitor), compared with the nominal level $\alpha_{\text{mon}} = 0.05$; seventeen of the 36 cells have no crossings. The uniform arm, which is the deployed default, crosses at most once in 200 streams in every cell. Holding the level fixed while sampling audits by margin, by a source-fitted second-moment score, adaptively, or by a clairvoyant oracle score confirms the mechanism of Theorem 1: inverse-propensity weighting makes the conditional mean of the evidence independent of the audit policy, so the policy changes the variance of the evidence while leaving the null guarantee intact.

The end-to-end panel closes the calibration loop. Instead of fitting the calibration once per instance and reusing it, this study redraws the calibration sample in every replication for a frozen model and monitors a fresh null stream under it, which is the composition that Equation (7) bounds by α_{total}. Across all twelve recalibration cells, comprising three problems $\times$ two budgets $\times$ two monitors, with 200 streams each, the crossing counts range from 0 to 4 out of 200 under redrawn calibration and from 0 to 4 out of 200 under fixed calibration. The two protocols therefore agree, and both remain well below the allowance.

The naive repeated fixed-window test isolates the benefit of the e-process. Applied to the same audited streams, with the first 40 audits as a reference window and a two-sample test at level 0.05 on the most recent 20 audits repeated every 10 audits, the any-crossing rate on null streams is 0.275: repeated testing at a fixed level spends the error budget multiple times, whereas the e-process spends it once regardless of how often the operator checks it. The threshold of the error-CUSUM baseline is calibrated on its own null streams to attain a 5% any-crossing rate by construction; it therefore lies on the nominal line by design and serves as a detection-power baseline rather than a validity comparison.

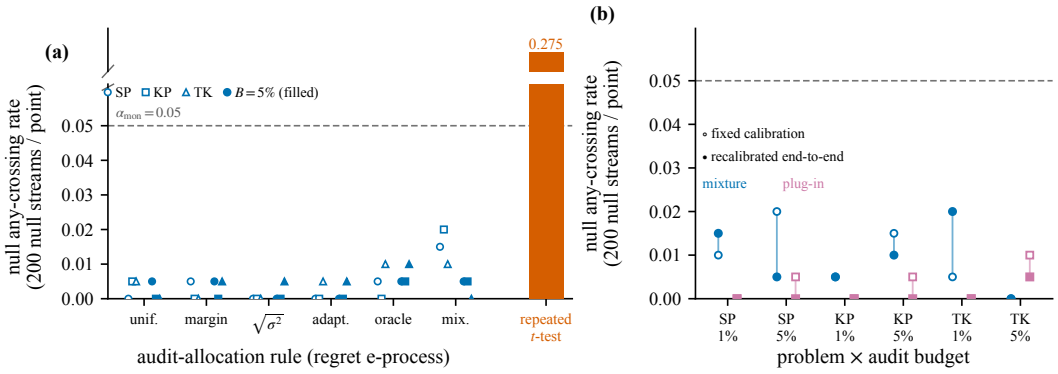

Figure 2: Anytime validity of the stopped regret e-process. **(a)** Null any-crossing rate for six audit-allocation and monitor configurations, one marker per problem (SP/KP/TK) and budget (open $B = 1\%$, filled $B = 5\%$), 200 null streams each. All 36 cells lie at or below 0.02 against the nominal $\alpha_{\text{mon}} = 0.05$ (dashed); the bar on the broken axis is the naive repeated fixed-window test on the same audited streams, at 0.275. **(b)** Fixed calibration (open) against calibration redrawn in every replication (filled), 200 null streams per point: 0 to 4 crossings in all twelve cells, so closing the calibration loop leaves the level intact. Points are exact binomial counts; cell-level values are given in Section A.2 and Table 2.

A.3 Which monitor to deploy

We compare the restarted mixture, unrestarted mixture, and plug-in monitor on shared audit streams. Figure 3 and Tables 3 and 4 report the comparison; Tables 5 and 6 give the full detection grids.

Figure 3(a) pools SP, KP, and TK into 300 streams per point at $B = 5\%$. The restarted mixture attains the lowest RMTA among the compared arms in both abrupt cells: 451.8 versus 514.9 for the plug-in and 932.4 for the unrestarted mixture under abrupt cost-structure drift, and 910.7 versus 970.3 and 1495.7, respectively, under abrupt concept drift. It also remains within 26 steps of the plug-in in the two gradual cells. Relative to the unrestarted mixture, its margin ranges from 153 to 585 steps across all four cells. The primary grid in Table 3 evaluates all fifteen cells with the

Table 2: Fixed-calibration empirical any-crossing of the stopped regret e-process (events/200 null streams per cell, pooled over 5 instances; nominal $\alpha_{\mathrm{mon}} = 0.05$; formal total $\alpha_{\mathrm{total}} \leq 0.10$ incl. calibration — this table holds calibrations fixed; the end-to-end recalibration nulls in Section A.2 exercise the full union bound). Last row: repeated standardized-mean test on the same audited stream (SP, B=5%; configuration in Section A.1).

Problem	Budget B	min	max (worst arm)	#arms
SP	1%	0/200	3/200 (mixture)	6
SP	5%	0/200	1/200 (uniform)	6
KP	1%	0/200	4/200 (mixture)	6
KP	5%	0/200	1/200 (oracle)	6
TK	1%	0/200	2/200 (oracle)	6
TK	5%	0/200	2/200 (oracle)	6
naive repeated t-test			0.28	—

restarted mixture as the primary monitor. The restarted mixture has a lower RMTA than the plug-in in nine of the fifteen cells and is within 65 steps in the remaining six, while outperforming the single mixture in all fifteen cells.

Figure 3(b) places the onset at $\nu = 3{,}200$ with 800 rounds of horizon left. The restarted mixture detects 217 of 300 streams (72.3%) at RMTA 518.7, while the unrestarted mixture detects 34 of 300 (11.3%) at RMTA 774.5: a process started at round one has spent wealth on the long null prefix, and predeclared restarts return that wealth. An independent replication in the primary grid, on the same cell with 150 streams, reproduces the ordering at 72.7% against 14.7%.

Figure 3(c) reports 8 crossings in 1,200 null streams for the restarted mixture (0.0067) against 9 for the unrestarted mixture and 3 for the plug-in, and 3 pre-onset alarms in 1,500 drifted streams against 10 for the unrestarted mixture; every arm is an order of magnitude below $\alpha_{\mathrm{mon}} = 0.05$. The two predeclared tail monitors also hold the level, at $1/1{,}200$ for $q_{.99}$ and $5/1{,}200$ for $q_{.90}$.

Table 4 also includes an online conformal exchangeability martingale evaluated on the identical audit stream. That arm tests exchangeability of the stream, a different null from the one-sided statement that mean clipped regret remains at its calibrated source level. The operational consequence of this distinction is visible in the real calm windows of Section A.10, where the exchangeability null alarms on genuine temporal dependence while the regret monitors alarm on at most 3 of 20 runs.

Table 5 jointly reports detection probabilities, exhaustion-before-detection counts, and RMTA for the audit-allocation arms on the detection grids, while Table 6 reports the conditional medians and pre-onset false-alarm counts for the same cells. Two patterns are consistent across the tables. Cost-structure drift is detected faster than concept drift at the same budget, and abrupt onsets are detected faster than gradual ramps, in every problem. Moreover, error CUSUM using the same audits detects only a small fraction of the streams detected by the regret monitor in the same cells, providing the empirical counterpart to Theorem 3.

The same pattern holds for the fourth predict-then-optimize problem. On 6×6 bipartite assignment at $B = 5\%$, the restarted mixture crosses in 2 of 200 null streams and detects 50 of 50 abrupt streams with an RMTA of 343.9 and 45 of 50 gradual streams with an RMTA of 1,413.0; error CUSUM using the same audits detects 1 of 50 and 3 of 50, respectively, at a null rate of 0.05 fixed by its own calibration. On the abrupt cell for this problem, the restarted mixture also outperforms the plug-in (445.2) and the unrestarted mixture (863.3).

Regret CUSUM on the same audits. Tables 7 and 8 separate the monitored statistic from the detection procedure. Every arm reads the primary-grid streams, audit coins, and ledger. With the procedure fixed at CUSUM, replacing absolute prediction error by audited regret raises detection at $B = 5\%$ from 100 to 714 of 750 streams, so the regret statistic carries the detection power. With the statistic fixed at regret, a CUSUM whose threshold is calibrated for the horizon detects sooner than the restarted mixture: mean RMTA 587 against 1,051 at $B = 5\%$, and 459 against 120 detected streams at $B = 1\%$. The source of the CUSUM threshold matters little: bootstrapping cal-B alone, simulating 1,000 source nulls per instance, and tuning the reference to the realized shift give 714, 714, and 724 detections. The two procedures differ in what they certify. The CUSUM threshold

Table 3: Primary detection grid with the **restarted mixture as the deployed monitor** (B=5%, uniform auditing, exact ledger, shared audit stream; 50 streams/cell over 5 instances). Each entry is **RMTA** (restricted mean time to alarm over *all* streams, non-detections censored at the horizon) with detections/ledger-exhaustions below it. The restarted mixture carries the full guarantee chain (validity, near-oracle delay, worst-onset rate, finite ledger) and has lower RMTA than the plug-in (validity only) in 9 of 15 cells and is within 65 steps in the remaining 6.

arm	prob.	co-abr	co-grad	cs-abr	cs-grad	cs-abr late
restarted mixture	SP	852 (46/1)	1824 (21/12)	407 (50/0)	1482 (46/2)	499 (35/9)
	KP	1263 (35/4)	1937 (10/22)	582 (48/0)	1504 (42/5)	516 (37/8)
	TK	784 (49/0)	1764 (31/11)	449 (50/0)	1363 (49/0)	540 (37/4)
plug-in	SP	882 (47/0)	1843 (21/13)	527 (49/1)	1448 (45/2)	487 (38/7)
	KP	1289 (35/4)	1945 (9/22)	639 (48/0)	1498 (40/7)	537 (38/7)
	TK	805 (48/1)	1699 (32/8)	566 (49/1)	1341 (47/1)	536 (38/4)
single mixture	SP	1496 (34/6)	2000 (0/28)	883 (50/0)	1860 (23/10)	763 (12/21)
	KP	1807 (16/9)	2000 (0/27)	1058 (44/2)	1877 (19/19)	781 (4/22)
	TK	1253 (46/1)	1954 (7/21)	864 (49/1)	1732 (32/4)	782 (6/14)

controls false alarms over the horizon T under a stationary null; monitored to $4T$, the three regret CUSUMs cross in 288–306 of 1,500 null streams, while the e-process crosses in 7. On the real histories, the block bootstrap does not carry this horizon-T control over: on FF30 calm93 and Citi Bike calm19, where the deployment mean of Y stays below $\overline{\mu}_0$, the block-bootstrap regret CUSUM alarms in 0–2 and 5–6 of 20 runs across budgets, and the restarted mixture in none. On calm14 the deployment mean of Y (0.441) exceeds $\overline{\mu}_0$ (0.430), so alarms there reflect a small level shift rather than a false alarm for either procedure; the same caution applies to pre-event alarms in the event windows.

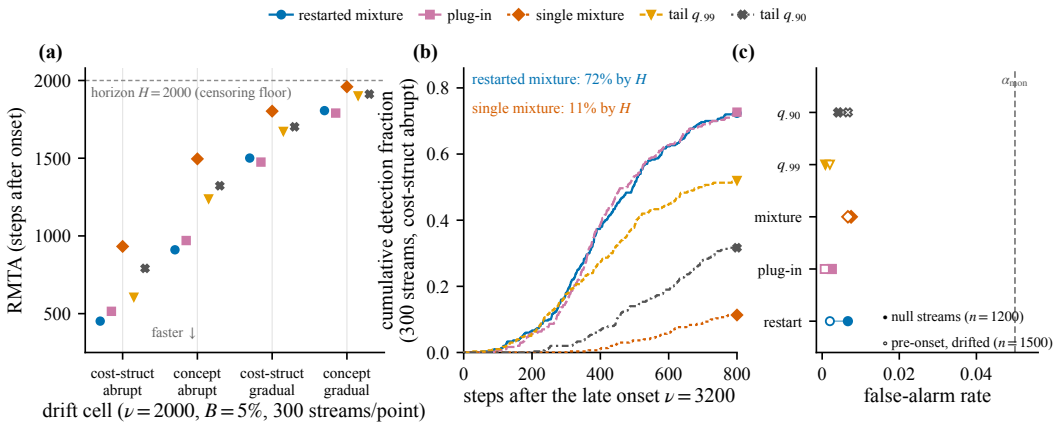

Figure 3: Monitor head-to-head on a shared audit stream ($B = 5\%$, uniform auditing, exact ledger $K = 200$; every arm sees identical audits and identical labels). **(a)** RMTA per drift cell at onset $\nu = 2{,}000$ and horizon $H = 2{,}000$, pooled over SP/KP/TK (300 streams per point); the restarted mixture leads on both abrupt cells and stays within 26 steps of the plug-in on both gradual cells. **(b)** Cumulative detection after a late onset $\nu = 3{,}200$ (300 streams): the restarted mixture reaches 72.3% against 11.3% for the unrestarted mixture. **(c)** Null any-crossings out of $n = 1{,}200$ (filled) and pre-onset alarms out of $n = 1{,}500$ (open) per arm, all far below $\alpha_{\text{mon}} = 0.05$. Points are exact means or exact binomial counts.

A.4 Validity under realistic temporal dependence

The monitor's false-alarm control is maintained under the autocorrelation and volatility clustering of a real daily financial series. Figure 4 and Table 9 report stationary block-bootstrap nulls constructed

Table 4: Head-to-head monitor comparison on a *shared* audit stream (uniform $B{=}5\%$, exact ledger $K{=}200$; 5 instances; null cells with $n{=}400$ and drifted cells with $n{=}100$). The null column reports any-crossings at $\alpha_{\mathrm{mon}}{=}.05/.001$. The detection columns report detected/median delay at the stated α_{mon} (cs = cost-structure, co = concept, abr/grad = abrupt/gradual; late = cost-structure abrupt with onset $\nu{=}3200$, otherwise $\nu{=}2000$). The single mixture incurs a change-time wealth penalty at late onsets; statistic-only restarts with thresholds $(R{+}1)/\alpha$ remove this penalty. The conformal martingale tests *exchangeability*, not one-sided regret (see Table 16).

arm	prob.	null $\frac{.05}{.001}$	cs-abr, .05	cs-abr, .001	cs-grad, .05	co-abr, .05	co-grad, .05	late, .05	late, .001
plug-in	SP	0/0	100/404	99/756	89/1448	89/735	42/1508	80/398	27/618
	KP	1/0	96/462	89/878	69/1445	80/1016	25/1743	62/378	22/518
	TK	2/0	99/436	99/809	94/1296	98/581	67/1476	76/384	32/670
mixture	SP	2/0	99/803	86/1141	48/1612	68/1246	9/1714	14/580	0/–
	KP	2/0	90/908	73/1187	29/1566	45/1594	1/1436	9/527	1/710
	TK	5/0	97/789	95/1097	78/1689	93/1112	20/1564	11/608	0/–
restarted mixture ($L{=}1000$)	SP	0/0	100/344	99/511	89/1446	87/693	39/1544	79/382	45/533
	KP	4/0	96/426	91/640	71/1429	86/969	20/1660	60/388	30/442
	TK	4/0	99/388	100/670	94/1369	99/579	66/1476	78/407	43/562
conformal martingale	SP	0/0	100/350	100/478	97/1440	99/711	60/1607	80/381	61/506
	KP	2/0	98/426	98/600	79/1435	96/928	38/1705	57/427	35/451
	TK	6/0	98/183	100/267	98/1058	100/270	97/1305	95/224	92/313
tail $q_{.99}$	SP	0/0	99/578	95/826	58/1680	67/1152	1/1735	36/518	13/649
	KP	1/0	92/683	82/932	41/1665	33/1257	1/1956	29/462	11/662
	TK	0/0	100/257	100/358	97/1308	100/416	67/1587	91/287	85/346
tail $q_{.90}$	SP	1/0	98/812	87/1077	38/1628	63/1341	7/1732	10/647	3/619
	KP	2/0	92/828	81/1049	35/1666	48/1448	1/1814	12/611	2/728
	TK	2/0	98/468	100/629	98/1368	100/656	64/1614	73/519	41/646

Table 5: Joint operating characteristics on the detection grids (B=5%, 50 streams/cell over 5 instances): detection probability / exhaustion-before-detection count / restricted mean time-to-alarm (RMTA, non-detections censored at horizon 2000).

		concept-abr		concept-grad		cs-abr		cs-grad	
arm	prob.	det/exh	RMTA	det/exh	RMTA	det/exh	RMTA	det/exh	RMTA
uniform	SP	45/2	968	24/13	1821	50/0	510	43/2	1545
	KP	36/2	1309	10/23	1906	47/0	678	41/6	1593
	TK	48/1	680	35/8	1623	49/1	484	47/1	1283
source score	SP	44/3	1006	18/11	1880	50/0	532	47/1	1441
	KP	39/6	1133	14/24	1870	46/1	708	40/6	1539
	TK	50/0	838	39/5	1634	50/0	509	48/1	1424
clairvoyant	SP	48/1	734	30/15	1746	50/0	456	47/2	1423
	KP	45/1	990	23/13	1814	50/0	428	43/5	1506
	TK	50/0	689	39/6	1575	50/0	459	50/0	1312
mixture	SP	33/4	1495	2/23	1967	50/0	804	27/14	1798
	KP	24/3	1688	0/24	2000	46/0	1014	15/17	1896
	TK	45/0	1245	6/26	1946	49/0	856	35/3	1742

from a calm FF30 source pool. Resampling in blocks preserves the empirical marginal distribution, autocorrelation, and volatility clustering of the real series while removing any regime change, so each stream is a null stream with realistic dependence.

Each of the 18 cells, comprising two budgets × three mean block lengths × three monitors, contains 600 null streams (3 predictor seeds × 200 resamples) of $T = 756$ deployment days. Against a nominal allowance of $\alpha_{\text{mon}} n = 30$ crossings per cell, the largest observed count is 6 of 600, and the largest count for the deployed restarted mixture is 3 of 600; the exact Clopper–Pearson 95% interval for the worst cell reaches 12.98 crossings, still well within the allowance. The dependence is real and measurable: the realized lag-1 autocorrelation of absolute returns rises from -0.002 for the i.i.d. control (mean block 1) to 0.227 at mean block 5 and 0.266 at mean block 21, while the autocorrelation of the monitored functional Y rises from -0.002 to 0.171 and 0.205. The mean of Y remains between 0.267 and 0.268 across all six stream configurations, so the block length changes the dependence structure while holding the quantity under test fixed.

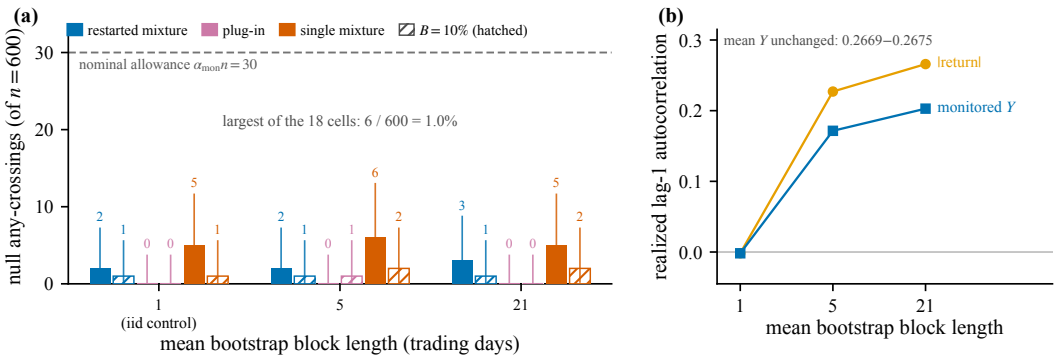

Figure 4: Anytime validity under the temporal dependence of a real daily series (FF30). Null streams are stationary block bootstraps of a calm source pool, preserving the marginal, the autocorrelation and the volatility clustering of the real data while removing any regime change. **(a)** Null any-crossings out of $n = 600$ streams per cell, for three monitors at two budgets and three mean block lengths. The nominal allowance at $\alpha_{\text{mon}} = 0.05$ is 30 crossings; the largest of the 18 cells records 6 and the deployed restarted mixture at most 3. Bars are exact Clopper–Pearson 95% intervals. **(b)** Realized lag-1 autocorrelation of $|r|$ and of the monitored functional Y against mean block length, with the mean of Y held at 0.267–0.268 throughout.

Table 6: Detection grid at $B = 5\%$ (50 streams per cell, 10 per instance over 5 instances). Entries are detections$/n$, pre-onset false alarms in brackets when nonzero, and in parentheses the median delay conditional on post-onset detection. Arms are compared unpaired within each cell; the paired comparison of allocations is in Table 11. The online-refit arm is exploratory.

Problem	Drift	uniform	source-fitted	online-refit	oracle (RM)	mixture	|err|-CUSUM
SP	concept-abrupt	45/50 (738)	44/50 [1] (818)	41/50 (902)	48/50 (604)	33/50 [1] (1260)	1/50 [1] (564)
SP	concept-gradual	24/50 (1636)	18/50 (1766)	23/50 (1686)	30/50 (1603)	2/50 (1170)	1/50 [1] (732)
SP	cost-structure-abrupt	50/50 (448)	50/50 (445)	50/50 (594)	50/50 (377)	50/50 (727)	0/50 [1]
SP	cost-structure-gradual	43/50 (1485)	47/50 (1431)	41/50 (1405)	47/50 (1398)	27/50 (1700)	0/50 [2]
KP	concept-abrupt	36/50 (960)	39/50 (869)	37/50 (1154)	45/50 (715)	24/50 (1319)	13/50 (602)
KP	concept-gradual	10/50 (1492)	14/50 (1518)	14/50 (1563)	23/50 (1671)	0/50 [1]	6/50 [3] (913)
KP	cost-structure-abrupt	47/50 (529)	46/50 [1] (508)	49/50 (766)	50/50 (385)	46/50 (843)	30/50 (425)
KP	cost-structure-gradual	41/50 (1554)	40/50 (1424)	36/50 (1537)	43/50 (1447)	15/50 (1690)	21/50 [1] (1596)
TK	concept-abrupt	48/50 [1] (576)	50/50 (719)	50/50 (825)	50/50 (627)	45/50 (1182)	0/50 [2]
TK	concept-gradual	35/50 (1493)	39/50 (1567)	29/50 [1] (1522)	39/50 (1475)	6/50 (1545)	1/50 [2] (1092)
TK	cost-structure-abrupt	49/50 [1] (420)	50/50 (466)	48/50 [2] (553)	50/50 (416)	49/50 (840)	5/50 [2] (1231)
TK	cost-structure-gradual	47/50 (1228)	48/50 (1394)	44/50 [1] (1375)	50/50 (1354)	35/50 (1633)	5/50 [4] (1349)

Table 7: Regret CUSUM against the restarted mixture on the primary detection streams (SP, KP, TK pooled; five instances each). Every arm reads the same deployment streams, audit coins, uniform rate B and ledger $K = BT$. CUSUM thresholds are the 95% quantile of the maximal statistic over the horizon T: the bootstrap arm resamples the cal-B regrets only, the simulation arms use $1{,}000$ simulated source null streams per instance, and the oracle arm uses reference $k = \Delta/2$ with Δ the realized post-onset shift of each cell (its null columns report the maximum over cells). Detection: detected streams and mean RMTA over cells. Nulls: any-crossings in the held-out null streams by T, and in fresh null streams monitored to $4T$ at $B = 5\%$ with every threshold kept at its horizon-T value.

	$B = 1\%$		$B = 5\%$		null crossings		
arm	detected	RMTA	detected	RMTA	by T, 1%	by T, 5%	by $4T$, 5%
error CUSUM (simulation)	40/600	1935	100/750	1643	35/600	26/600	297/1500
regret CUSUM (bootstrap)	459/600	1152	714/750	587	31/600	32/600	306/1500
regret CUSUM (simulation)	459/600	1142	714/750	583	29/600	28/600	288/1500
regret CUSUM (oracle k)	458/600	1142	724/750	564	29/600	36/600	304/1500
restarted mixture e-process	120/600	1814	586/750	1051	3/600	6/600	7/1500

Table 8: Regret CUSUM on the real deployment histories (20 runs per entry: 5 predictor $\times$ 4 audit seeds; shared audit stream and ledger). Entries are alarms out of 20, with alarms before the reference event in parentheses for the event windows. CUSUM thresholds are the 95% quantile of the maximal statistic over deployment-length bootstrap sequences drawn from cal-B, either i.i.d. or in circular blocks of 21 days. The last two rows give the deployment mean of Y and $\overline{\mu}_0$, averaged over predictor seeds: on calm93 and calm19 the mean stays below $\overline{\mu}_0$, whereas on calm14 it exceeds $\overline{\mu}_0$.

B	arm	FF30 COVID	FF30 GFC	FF30 calm14	FF30 calm93	Bike COVID	Bike calm19
5%	restarted mixture	1 (0)	1 (1)	0	0	7 (0)	0
	error CUSUM, block	19 (14)	20 (14)	8	2	20 (0)	1
	regret CUSUM, i.i.d.	20 (10)	20 (17)	16	1	20 (0)	6
	regret CUSUM, block	20 (10)	20 (17)	16	1	20 (0)	5
10%	restarted mixture	7 (0)	8 (1)	1	0	14 (0)	0
	error CUSUM, block	20 (19)	20 (14)	10	5	20 (0)	1
	regret CUSUM, i.i.d.	20 (13)	20 (20)	20	1	20 (0)	6
	regret CUSUM, block	20 (14)	20 (20)	19	0	20 (0)	6
20%	restarted mixture	17 (0)	18 (2)	3	0	15 (0)	0
	error CUSUM, block	20 (20)	20 (19)	10	4	20 (0)	1
	regret CUSUM, i.i.d.	20 (17)	20 (20)	20	3	20 (1)	6
	regret CUSUM, block	20 (17)	20 (20)	19	2	20 (0)	5
	mean of Y	0.376	0.562	0.441	0.420	0.521	0.192
	$\overline{\mu}_0$	0.300	0.478	0.430	0.474	0.343	0.225

Table 9: Anytime validity under *realistic temporal dependence* (FF30). Null streams are stationary block bootstraps of a calm source pool: the resample preserves the empirical marginal, the autocorrelation and the volatility clustering of the real series while destroying any regime change, so each stream is a genuine null with real dependence. Mean block length 1 is the iid control. Entries: any-crossing count out of 600 (3 predictor seeds $\times$ 200 resamples); α_{mon}=0.05. Realized dependence is reported as the lag-1 autocorrelation of |returns| and of the monitored functional Y.

B	mean block	$\mathrm{ac}_1(\lvert r\rvert)$	$\mathrm{ac}_1(Y)$	restart	plug-in	mixture
0.05	1	-0.002	-0.002	2/600	0/600	5/600
	5	0.228	0.171	2/600	0/600	6/600
	21	0.266	0.205	3/600	0/600	5/600
0.10	1	-0.001	-0.002	1/600	0/600	1/600
	5	0.226	0.172	1/600	1/600	2/600
	21	0.266	0.201	1/600	0/600	2/600

A.5 Budget, delay, and the finite ledger

Detection delay shortens with the audit budget close to the $1/B$ rate predicted by Theorem 2, and the finite ledger behaves as part (iii) of that theorem describes: the audit *rate* determines the delay, whereas the audit *stock* determines whether that delay can be realized. Figure 5 reports both.

The scaling sweep considers four drift magnitudes at four budgets, with 50 streams per point on the uniform arm. Fitting $\log(\text{median delay})$ on $\log B$ and $\log \delta_{\text{eff}}$ over the cells that detect at least half of their streams yields a budget exponent of -0.835, compared with the -1 predicted by Equation (9), and a margin exponent of -1.004. The lower bound in Theorem 2(i) isolates the budget dependence, which the fit recovers: at the largest margin, the median delay falls from $1{,}124.5$ rounds at $B = 1\%$ to 417.5 at $B = 5\%$ and 272.5 at $B = 10\%$, while the detection count rises from $16/50$ to $50/50$. Cells that detect fewer than half of their streams (open markers) are excluded from the fit, because their conditional medians summarize only the easiest detected streams.

The ledger panel separates audit rate from audit stock. The fraction of streams that exhaust the entire ledger without ever crossing falls sharply with the budget in all four drift cells: from $29/150$ to $0/150$ on cost-structure abrupt, $59/150$ to $5/150$ on concept abrupt, $65/150$ to $7/150$ on cost-structure gradual, and $82/150$ to $45/150$ on concept gradual, as B increases from 1% to 5%. Meanwhile, the fraction of streams that exhaust the ledger at some point, whether or not they have already alarmed, remains essentially flat: $309/600$ (51.5%) at $B = 1\%$ and $321/600$ (53.5%) at $B = 5\%$. The contrast concerns whether exhaustion occurs before detection, as in Equation (10).

A.6 What the monitored functional tracks

The clipped functional used by the monitor moves with the decision quality it is intended to certify, as does the predeclared tail functional, in every cell of the grid. Figure 6 and Table 10 report these three functionals on the same audited stream before and after the onset for all fifteen problem $\times$ drift cells at $B = 5\%$, with each entry averaged over 50 streams.

Raw decision regret rises in 15 of 15 cells, by factors ranging from $1.726\times$ (KP, concept gradual: $1.42 \to 2.45$) to $5.253\times$ (SP, cost-structure abrupt: $1.45 \to 7.46$). The monitored clipped mean Y rises in the same 15 of 15 cells, including from 0.156 to 0.576 on SP cost-structure abrupt and from 0.327 to 0.715 on TK concept abrupt. The predeclared $q_{.90}$ tail-exceedance functional also rises in the same 15 of 15 cells, from a pre-onset level near its calibrated 0.10 to between 0.228 and 0.856 post-onset; its movement is at least as large as that of the clipped mean in every cell. The post-onset share of audited rounds at the clipping ceiling ranges from 0.064 to 0.641 across the grid, and the clipped mean separates pre-onset from post-onset values across this entire range.

Clipping provides bounded observations for sequential testing while preserving the deterioration seen in raw regret across this grid. The tail-exceedance rate shows that deterioration also takes the form of more frequent high-regret decisions.

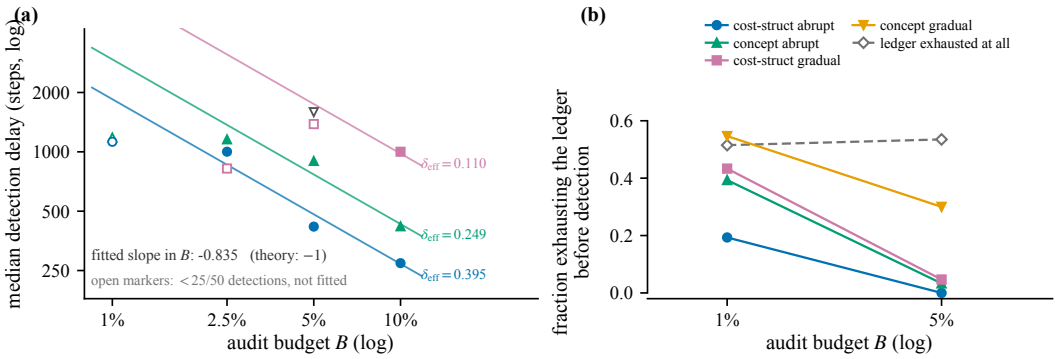

Figure 5: Budget–delay scaling and the finite ledger. **(a)** Median detection delay against audit budget on log–log axes, one series per drift magnitude, uniform auditing, 50 streams per point. Lines show the log-log fit over the cells with at least $25/50$ detections; the fitted budget exponent is -0.835 against the -1 predicted by Equation (9). Open markers are shown but excluded from the fit. **(b)** Fraction of streams exhausting the detection ledger before crossing (150 streams per point), falling from 19.3–54.7% at $B = 1\%$ to 0–30% at $B = 5\%$. Points are exact medians or exact counts; per-cell values are in Section A.5.

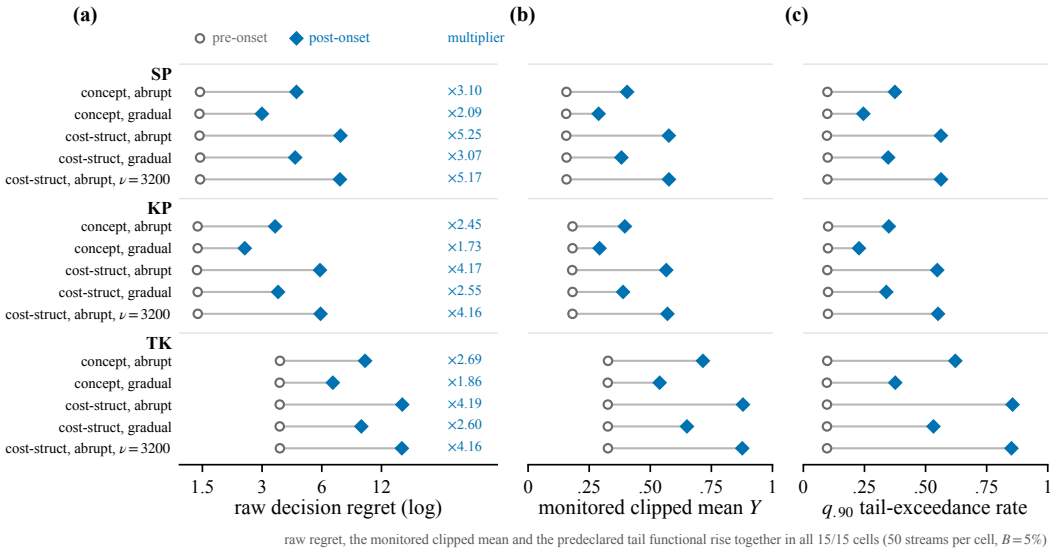

Figure 6: Three functionals of the same audited regret stream, pre-onset (open) against post-onset (filled), for all 15 problem × drift cells at $B = 5\%$; each value is a cell-level mean over 50 streams (5 instances × 10 seeds). **Left:** raw decision regret on a log axis, with the post/pre multiplier printed per row, ranging from $1.73\times$ to $5.25\times$. **Centre:** the monitored clipped mean Y of Equation (1), the quantity the e-process bets on. **Right:** the predeclared $q_{.90}$ tail-exceedance rate, rising from its calibrated pre-onset level near 0.10 to between 0.228 and 0.856. All three functionals rise in 15 of 15 cells; each value is the exact cell mean.

Table 10: The clipped certificate tracks decision quality. Per drifted cell (B=5%, means over 50 streams × 5 instances): mean *raw* regret pre→post (and its ratio), the monitored *clipped* mean Y, the $q_{.90}$ *tail exceedance* rate, and the clipping-saturation fraction $\mathbb{P}(Y{=}1)$ post-onset. All three functionals move in the same direction on every cell; the clipped mean separates pre- from post-onset values even at post-onset saturation up to 0.641, and the predeclared tail functional (Theorem 1) moves at least as much.

prob.	cell	raw regret pre→post (×)	clipped Y pre→post	tail $q_{.90}$ pre→post	sat. post
SP	co-abr	1.46→4.47 (3.102×)	0.157→0.405	0.098→0.375	0.163
SP	co-grad	1.45→2.99 (2.088×)	0.156→0.289	0.098→0.246	0.084
SP	cs-abr	1.45→7.46 (5.253×)	0.156→0.576	0.096→0.563	0.325
SP	cs-grad	1.46→4.41 (3.067×)	0.157→0.382	0.100→0.347	0.159
SP	cs-abr late	1.46→7.41 (5.175×)	0.157→0.576	0.099→0.563	0.326
KP	co-abr	1.42→3.49 (2.452×)	0.182→0.396	0.101→0.350	0.126
KP	co-grad	1.42→2.45 (1.726×)	0.182→0.292	0.101→0.228	0.064
KP	cs-abr	1.41→5.88 (4.175×)	0.180→0.565	0.098→0.548	0.296
KP	cs-grad	1.42→3.61 (2.553×)	0.182→0.389	0.101→0.339	0.140
KP	cs-abr late	1.42→5.92 (4.162×)	0.182→0.570	0.100→0.551	0.300
TK	co-abr	3.69→9.92 (2.69×)	0.327→0.715	0.098→0.622	0.341
TK	co-grad	3.68→6.82 (1.857×)	0.326→0.538	0.096→0.376	0.162
TK	cs-abr	3.68→15.28 (4.189×)	0.326→0.879	0.097→0.856	0.641
TK	cs-grad	3.68→9.50 (2.599×)	0.326→0.650	0.097→0.532	0.314
TK	cs-abr late	3.69→15.21 (4.161×)	0.327→0.876	0.097→0.852	0.638

A.7 AUDIT ALLOCATION: THE EXACT EXCESS AND WHAT THE ORACLE IS MADE OF

The exact-excess identity in Proposition 1 holds numerically to machine precision. Decomposing the apparent advantage of the clairvoyant oracle shows that most of it comes from scheduling audit *intensity* around the onset, which presupposes knowledge of when the change occurs. Figure 7 and Table 11 report both results.

Figure 7(a) evaluates both sides of Equation (15) on the same $N = 4{,}000$ evaluation sample for five instances × three drift magnitudes. Measured $V(p_s; Q) - V^\star(Q)$ and the closed form $\mathbb{E}_Q[s]^2 \operatorname{Var}_{\mu_s}(\varepsilon)/B$ agree on all 15 checks with a maximum absolute relative error of 0.0; the checks span excesses from 165.96 to 2,241.60. The penalty for auditing by a score s instead of σ_Q is the μ_s-variance of the relative discrepancy $\varepsilon = \sigma_Q/s - 1$, scaled by $\mathbb{E}_Q[s]^2/B$. It vanishes exactly when the score is proportional to the deployment volatility profile and grows with the spread of the ratio, which is the mechanism behind the transfer decomposition of Equation (16) and the reason a source-fitted score gains nothing once drift reshapes that profile.

Figure 7(b) splits the clairvoyant comparison into its two ingredients on paired streams, 100 per cell over the six abrupt cells, with instance-level cluster-bootstrap 95% intervals. Pure intensity scheduling — uniform allocation run at the oracle's post-onset audit intensity, with no context information whatsoever — is worth a median 302.6 steps of RMTA across the six cells and is positive with an interval excluding zero in every one of them (+346, +226, +500, +273, +332 and +229 steps). Allocation at a common 5% rate — the clairvoyant score with the intensity advantage removed — is worth a median 82.25 steps. The unmatched clairvoyant oracle, which is the comparison that mixes both, is worth a median 339.0 steps and thus sits close to the intensity arm alone. Reading the same rows in Table 11, the unmatched oracle schedules a post-onset propensity of 9.2% to 13.1% against the nominal 5% and exhausts its ledger in 100% of streams, whereas the rate-matched arms schedule 5.0%.

The unmatched oracle benefits mainly from concentrating audits after the change. The rate-matched comparison isolates the smaller gain from choosing contexts at a common audit rate. This distinction explains why an oracle advantage need not transfer to a policy fixed before deployment, and reinforces the minimax motivation for uniform auditing under volatility reshaping in Theorem 4.

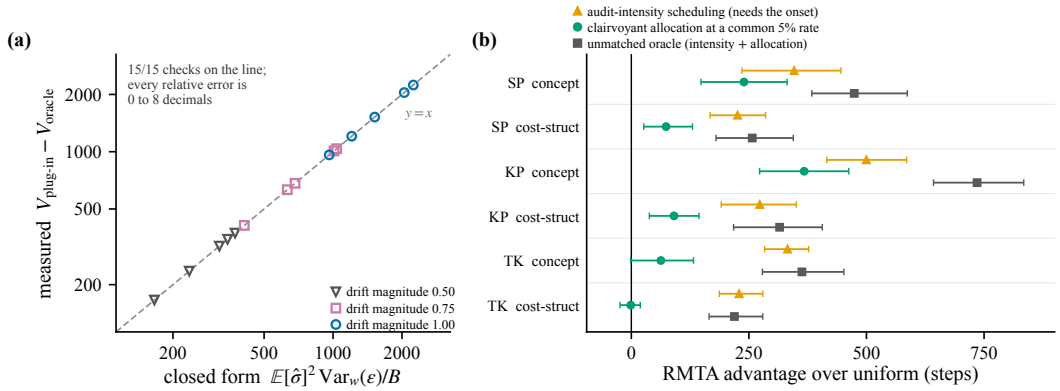

Figure 7: Audit allocation. **(a)** Numerical check of the exact-excess identity Equation (15): measured $V_{\text{plug-in}} - V_{\text{oracle}}$ against its closed form, one point per instance $\times$ drift magnitude (15 checks, $B = 5\%$). All 15 lie on the identity line, with maximum absolute relative error 0.0 over excesses from 165.96 to 2,241.60. **(b)** RMTA advantage over uniform auditing at horizon $H = 2{,}000$ on paired abrupt streams, 100 pairs per cell, with instance-level cluster-bootstrap 95% intervals. Intensity scheduling alone — uniform allocation run at the oracle's post-onset intensity, using no context information — is worth a median 302.6 steps, and allocation at a common 5% rate a median 82.25 steps: the clairvoyant advantage is dominated by audit-intensity scheduling, which presupposes knowing the onset.

A.8 Guarded adaptation

Figure 8 organizes the evidence around harmful endpoints, while Tables 12 and 13 report the corresponding regret reductions and label accounting.

Genuine drift. Figure 8(a) pools SP and KP into 100 paired streams per bar under a common 200-label ceiling. Guarded adaptation records $0/100$ harmful endpoints on abrupt drift and $0/100$ on gradual drift. Updating on a fixed schedule records $0/100$ on abrupt drift and $50/100$ on gradual drift; updating at the true onset, handed for free, records $0/100$ and $95/100$. Table 12 gives the matching all-stream regret reductions for the guarded arm — $+18.5\%$ and $+3.1\%$ on SP abrupt and gradual, $+26.1\%$ and $+8.1\%$ on KP — together with its alarm and promotion counts ($48/24$, $28/11$, $45/33$, $24/16$) and its mean label usage (142 to 151 of the 200-label ceiling).

A decision-blind candidate operator. Figure 8(b) hands the loop a closed-form ridge head fitted to squared prediction error, an operator that carries no information about the solver, and runs $T = 6{,}000$ streams with two onsets and up to three alarm–train–gate cycles under the spending rule $\alpha_k = \alpha_{\text{dep}} 2^{-k}$. Deployed without the gate, this candidate is harmful on $6/50$, $5/50$, $23/50$ and $16/50$ streams of the four cells, 50 of 200 pooled, promoting itself 1.88 to 2.72 times per stream. Behind the gate it is harmful on $0/50$ in every cell, 0 of 200 pooled, and reaches promotion at most 0.10 times per stream. The decision-aware SPO+ head runs through the same protocol with 0 harmful endpoints gated and ungated, so across both operators the gate holds harmful endpoints at 0 of 400 while ungated deployment records 50 of 400. Table 13 reports the reductions over the whole post-first-onset horizon, which prices any harm the second regime change inflicts on an already-promoted model: the gated SPO+ arm reduces regret by $+18.4\%$, $+11.8\%$, $+30.1\%$ and $+28.6\%$ on the four cells.

Forced false alarms. Figure 8(c) forces an alarm at round 1,000, before the true onset at 2,000, so every candidate is trained on null data and satisfies the shadow null by construction. Across 200 streams the gate promotes 0 and deploys 0, and therefore records 0 harmful endpoints; deploying the same 200 candidates without the gate harms 194 of 200 streams. Repeating the forced alarm up to three times per stream under the spending schedule $(0.025, 0.0125, 0.0063)$ leaves the gated count

Table 11: Intensity/allocation decomposition of the oracle advantage (abrupt drifts, B=5%, 100 paired streams per cell over 5 instances; W/L/T among both-alarm pairs; RMTA saving at horizon H=2000 with instance-level cluster-bootstrap 95% CIs; the last three columns give the comparison arm's scheduled post-onset audit propensity $\bar{p}$, its mean post-onset audit count, and the fraction of streams that exhaust the ledger).

Prob.	Drift	Comparison	W/L/T	Med. save	RMTA save [CI]	sched. $\bar{p}$	audits	exh.
SP	concept	allocation at common 5%	74/19/3	138	+240 [+148, +331]	5.0%	95	58%
SP	concept	allocation at elevated rate	65/27/3	42	+128 [+84, +165]	9.8%	102	100%
SP	concept	intensity scheduling only	86/5/3	249	+346 [+235, +446]	9.8%	102	100%
SP	concept	unmatched oracle (confounded)	93/3/0	358	+474 [+384, +587]	9.8%	102	100%
SP	cost-struct	allocation at common 5%	56/33/10	17	+74 [+26, +130]	5.0%	94	52%
SP	cost-struct	allocation at elevated rate	65/27/8	9	+31 [+12, +59]	13.1%	101	100%
SP	cost-struct	intensity scheduling only	93/4/2	170	+226 [+168, +285]	13.0%	103	100%
SP	cost-struct	unmatched oracle (confounded)	97/0/2	201	+257 [+180, +344]	13.1%	101	100%
KP	concept	allocation at common 5%	57/15/1	152	+368 [+273, +462]	5.0%	94	51%
KP	concept	allocation at elevated rate	60/22/1	55	+235 [+166, +305]	9.2%	98	100%
KP	concept	intensity scheduling only	61/4/2	342	+500 [+416, +585]	9.2%	100	100%
KP	concept	unmatched oracle (confounded)	68/4/0	552	+735 [+642, +834]	9.2%	98	100%
KP	cost-struct	allocation at common 5%	64/25/10	47	+91 [+38, +144]	5.0%	96	54%
KP	cost-struct	allocation at elevated rate	75/19/5	31	+42 [+26, +57]	12.3%	100	100%
KP	cost-struct	intensity scheduling only	92/4/2	222	+273 [+191, +351]	12.3%	101	100%
KP	cost-struct	unmatched oracle (confounded)	94/3/2	229	+315 [+217, +406]	12.3%	100	100%
TK	concept	allocation at common 5%	57/30/11	22	+63 [-2, +132]	5.0%	92	52%
TK	concept	allocation at elevated rate	49/41/9	0	+31 [-16, +78]	10.0%	98	100%
TK	concept	intensity scheduling only	87/10/2	260	+332 [+283, +377]	10.0%	100	100%
TK	concept	unmatched oracle (confounded)	89/8/2	244	+363 [+279, +452]	10.0%	98	100%
TK	cost-struct	allocation at common 5%	50/35/14	1	-1 [-24, +19]	5.0%	94	43%
TK	cost-struct	allocation at elevated rate	52/36/11	1	-10 [-49, +11]	11.9%	101	100%
TK	cost-struct	intensity scheduling only	94/6/0	186	+229 [+187, +280]	11.9%	102	100%
TK	cost-struct	unmatched oracle (confounded)	94/4/1	188	+219 [+165, +279]	11.9%	101	100%

at 0 null promotions across 200 streams and all three cycles, against 600 ungated null promotions and at least one on every stream.

Gate constants. Figure 8(d) sweeps the two free constants of the gate, $\alpha_{\mathrm{dep}} \in \{0.01, 0.05, 0.1, 0.2\}$ and the shadow-window cap in $\{15, 30, 60\}$ labels, over two problems and two drift timings with 25 streams per setting. Across all 48 settings the harmful-endpoint count is 0. Promotion counts move as the constants imply, from 1 to 21 of 25, and the regret reduction moves with them, from 0.0% up to +32.4% across the sweep; the safety property is invariant to both. The ceiling $K = BT = 200$ holds in every configuration: with the cap of 60 labels the allotments exceed K, and the shadow scan stops at $\ell = K$ (Algorithm 1).

The fourth architecture, end to end. On 6×6 bipartite assignment the complete loop runs at $B = 5\%$, with adaptation labels budgeted separately from the detection ledger: all 50 abrupt streams alarm, 41 candidates are promoted, 0 streams are harmful at the endpoint, and post-onset regret falls by 38.0%; in the gradual cell, 46 streams alarm, 33 candidates are promoted, 0 are harmful at the endpoint, and regret falls by 16.8%. Decision disagreements are dense on this architecture (88% to 91% of scanned shadow rounds), so nearly every scanned round contributes comparison evidence to the shadow test.

A.9 Calibration size, normalization, and hyperparameter sensitivity

The monitor's behaviour is governed by quantities the operator controls and understands, and it is stable in the free constants. Tables 14 and 15 report the two predeclared studies.

Calibration size acts through the null level. Growing n_{cal} from 500 to 2,000 to 8,000 tightens the one-sided empirical-Bernstein bound $\overline{\mu}_0$ from 0.240 to 0.181 to 0.165, and the detection improves accordingly, from $48/50$ at median delay 764 to $50/50$ at 522 and $50/50$ at 477 on SP. A tighter calibrated null is a larger post-change margin δ in Equation (9), so the delay shortens as the theory

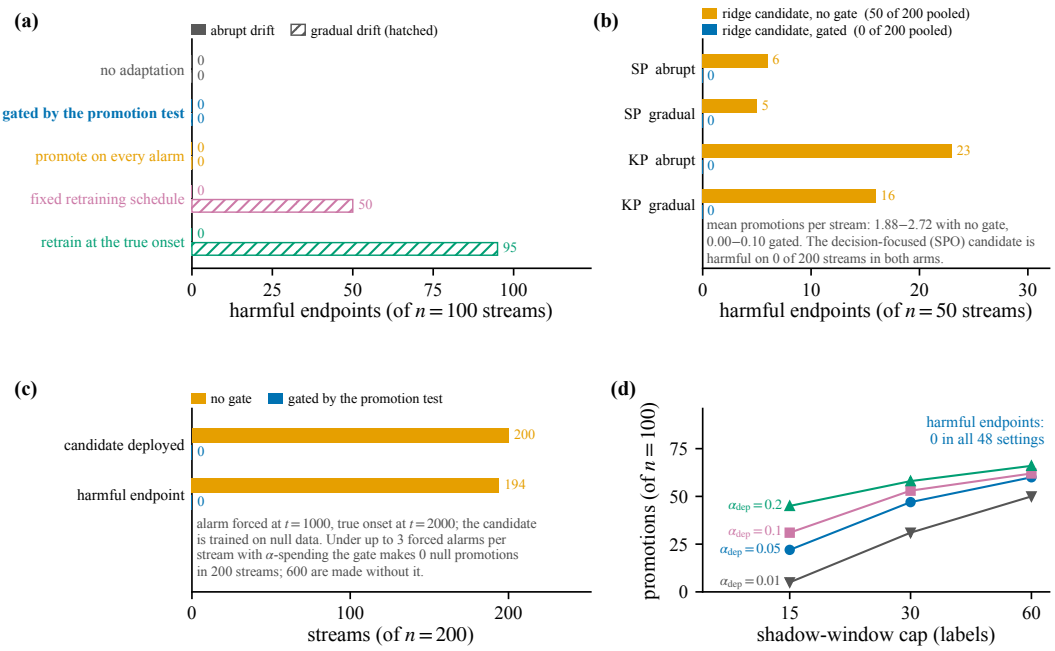

Figure 8: Guarded adaptation, organized around harmful endpoints. A stream is harmful when its post-onset cumulative decision regret exceeds the no-adaptation reference on the same stream. **(a)** Harmful endpoints by update policy under a common 200-label ceiling, pooled over SP and KP, $n = 100$ streams per bar: guarded $0/100$ on both drift shapes, scheduled updating $50/100$ and updating at the true onset $95/100$ on gradual drift. **(b)** A deliberately decision-blind ridge candidate operator ($T = 6{,}000$, onsets at $2{,}000$ and $4{,}000$, up to three gated cycles, $n = 50$ per bar): harmful on 50 of 200 pooled streams ungated and on 0 of 200 gated. **(c)** Forced pre-onset alarms at round $1{,}000$ with the true onset at $2{,}000$, $n = 200$: the gate deploys 0, while ungated deployment of the identical candidates harms 194. **(d)** Promotions against the shadow-window cap for $\alpha_{\text{dep}} \in \{0.01, 0.05, 0.1, 0.2\}$; the harmful-endpoint count is 0 in all 48 settings. Bars are exact counts.

Table 12: Paired update arms under a common 200-label ceiling (50 streams/cell = 10 per instance $\times$ 5 instances). All-stream mean post-onset regret reduction vs. no-adapt; harmful = post-onset cumulative regret exceeding no-adapt; guarded alarm/promotion denominators. Label usage and solver calls differ across arms (Table 18). The oracle-timed arm is given the true onset.

Cell	Arm	reduction	harmful	alarms/promotions	mean labels
SP-abrupt	guarded	+18.5%	0/50	48/24	147
	ungated	+36.5%	0/50	—	125
	periodic	+20.5%	0/50	—	194
	oracle-timed	+52.3%	0/50	—	50
SP-gradual	guarded	+3.1%	0/50	28/11	151
	ungated	+10.6%	0/50	—	137
	periodic	-2.8%	28/50	—	194
	oracle-timed	-21.3%	50/50	—	50
KP-abrupt	guarded	+26.1%	0/50	45/33	146
	ungated	+36.6%	0/50	—	129
	periodic	+22.9%	0/50	—	197
	oracle-timed	+55.2%	0/50	—	50
KP-gradual	guarded	+8.1%	0/50	24/16	142
	ungated	+9.7%	0/50	—	133
	periodic	+4.8%	22/50	—	195
	oracle-timed	-13.1%	45/50	—	50

Table 13: Adaptation under *continued* drift with *multiple* interventions and *two* candidate operators (T=6000; drift onsets at 2000 and 4000; up to three alarm→train→gate cycles per run under the α-spending schedule of Theorem 5, $\alpha_k = \alpha_{\text{dep}} 2^{-k}$; detector = restarted mixture; 50 streams/cell over 5 instances). Regret reduction is measured over the whole post-first-onset horizon, so it prices any harm the second regime change inflicts on an already-promoted model. The ridge head is deliberately *decision-blind* (MSE, closed form): ungated it is harmful on 5–23 of 50 streams, gated it is harmful on none — the gate filters a weak operator instead of deploying it. The gate gives up part of the SPO+ gain in exchange for protection that holds for either operator.

arm	prob./cell	reduction	harmful	promotions/run	labels
no-adapt	SP abrupt	+0.0%	0/50	0.00	0
guarded (SPO+)	SP abrupt	+18.4%	0/50	0.70	300
guarded (ridge)	SP abrupt	+1.9%	0/50	0.10	300
ungated (SPO+)	SP abrupt	+41.0%	0/50	2.04	300
ungated (ridge)	SP abrupt	+16.4%	6/50	2.68	299
no-adapt	SP gradual	+0.0%	0/50	0.00	0
guarded (SPO+)	SP gradual	+11.8%	0/50	0.42	300
guarded (ridge)	SP gradual	+0.8%	0/50	0.06	300
ungated (SPO+)	SP gradual	+31.2%	0/50	1.68	300
ungated (ridge)	SP gradual	+15.1%	5/50	2.02	300
no-adapt	KP abrupt	+0.0%	0/50	0.00	0
guarded (SPO+)	KP abrupt	+30.1%	0/50	0.84	300
guarded (ridge)	KP abrupt	+1.1%	0/50	0.04	300
ungated (SPO+)	KP abrupt	+51.2%	0/50	1.86	300
ungated (ridge)	KP abrupt	+3.2%	23/50	2.72	297
no-adapt	KP gradual	+0.0%	0/50	0.00	0
guarded (SPO+)	KP gradual	+28.6%	0/50	0.80	300
guarded (ridge)	KP gradual	+0.0%	0/50	0.00	300
ungated (SPO+)	KP gradual	+37.0%	0/50	1.10	300
ungated (ridge)	KP gradual	+5.9%	16/50	1.88	300

says it should; the operator can convert calibration data into detection speed at a known exchange rate. Table 14 also records the fixed-constant normalization variant: replacing the calibrated $q_{.99}$ scale R_q with an a-priori constant R_{fix} shrinks the monitored functional so far that a strong drift

Table 14: Calibration-size ablation (SP, source-fitted arm, five instances, 50 streams per row; $\bar{\mu}_0$ = mean one-sided empirical-Bernstein UCB across instances), and the fixed-constant R_{fix} normalization variant (zero power at strong drift; not an a-priori bound — costs are unbounded under Gaussian features; single instance).

n_{cal}	$\bar{\mu}_0$	detect	median delay
500	0.240	48/50	764
2000	0.181	50/50	522
8000	0.165	50/50	477
R_{fix} variant: detect 0/10 (vs 5/10 clipped $q_{.99}$ scale)			

Table 15: Predeclared one-at-a-time hyperparameter sensitivity (SP cost-structure abrupt, B=5%, source-fitted sampler, five instances; per setting: 100 null / 40 drifted streams). Base setting: q=0.99, window 300, warm-up 20.

Setting	$\bar{\mu}_0$	null crossings	detect	median delay
q=0.95	0.247	0/100	40/40	486
base	0.181	0/100	40/40	388
q=0.995	0.159	0/100	40/40	373
window 150	0.181	0/100	40/40	362
window 600	0.181	0/100	40/40	422
warm-up 10	0.181	0/100	40/40	388
warm-up 40	0.181	0/100	40/40	386

produces $0/10$ detections against $5/10$ for the clipped-quantile normalization, which is why E-REACT normalizes on cal-A.

The one-at-a-time sensitivity sweep varies the clipping quantile $q \in \{0.95, 0.99, 0.995\}$, the betting window in $\{150, 300, 600\}$, and the warm-up in $\{10, 20, 40\}$ around the base configuration, with 100 null and 40 drifted streams per configuration on SP cost-structure abrupt. Every configuration records $0/100$ null crossings and $40/40$ detections. The median delay ranges from 362 to 486 rounds, while the two purely computational constants, window and warm-up, change it by at most 34 rounds relative to the base value of 388. The quantile q changes $\overline{\mu}_0$ (0.247, 0.181, 0.159), and the delay changes in the direction predicted by the margin argument.

A.10 REAL DEPLOYMENT HISTORIES

We evaluate monitoring on FF30 and Citi Bike histories and run the adaptation loop on post-alarm candidates. Figure 9(a), (b), and (d) and the adaptation results use the plug-in e-process of Section A.1; Figure 9(c) and Tables 8 and 16 report the restarted mixture on the same audit streams. The restarted mixture alarms on 1, 7, and 17 of 20 FF30 COVID runs, 1, 8, and 18 GFC runs, and 7, 14, and 15 Citi Bike COVID runs at $B = 5\%$, 10%, and 20%, and on none of the calm-window runs at $B = 5\%$. Tables 16 and 17 compare monitored functionals and calibration choices.

Detection scales with the budget; calm windows stay quiet. Figure 9(a) reports alarm counts out of 20 runs (5 predictor $\times$ 4 audit seeds) for each budget. On FF30 the COVID window goes from $2/20$ at $B = 5\%$ to $10/20$ at 10% and $18/20$ at 20%, and the GFC window goes from $2/20$ to $9/20$ to $19/20$. On Citi Bike, the COVID window goes from $8/20$ to $15/20$ to $16/20$. The three event-free windows record $0/20$, $0/20$, and $0/20$ alarms at $B = 5\%$; across all three budgets, they record a single alarm in 180 runs (calm14 at $B = 10\%$), while the FF30 1993–1996 and Citi Bike 2019 windows record 0 alarms across all 120 runs. Figure 9(b) shows that the delay shortens as the budget increases in every event window: 257 $\rightarrow$ 210 days on FF30 COVID, 179 $\rightarrow$ 109.5 days on FF30 GFC, and 194 $\rightarrow$ 95 days on Citi Bike, measured from the reference event date among the runs that alarm.

Which null is being tested. Figure 9(c) runs an online conformal exchangeability martingale and the regret e-process on a shared FF30 audit stream under one ledger. On the calm14 window, the

Table 16: FF30 case study, multi-functional monitors on a shared audit stream (α_{mon}=.05; 20 runs/cell = 5 predictor × 4 audit seeds; one realized history). Event windows: detected/median delay in days (pre-event alarms in parentheses); calm windows: alarm count out of 20. Exceedance rates (one predictor seed): GFC $q_{.90}$ .18 → .54, $q_{.99}$ .06 → .30; COVID $q_{.90}$ .04 → .22 — the predeclared tail functionals see the crises clearly. The conformal martingale's exchangeability null false-alarms on real calm dependence (calm14 at B=20%: 17/20), while the one-sided regret monitors alarm on at most 3 of 20 runs.

	GFC		COVID		calm14		calm93	
arm	B=5%	20%	5%	20%	5%	20%	5%	20%
mean mixture	0/–	7/148	0/–	11/252(1)	0	0	0	1
restarted mixture (L=252)	0/–(1)	16/128(2)	1/72	17/250	0	3	0	0
tail $q_{.99}$	3/146	18/122	0/–	0/–	0	0	0	0
tail $q_{.90}$	9/143	16/82(3)	2/390	13/193	0	0	0	0
conformal martingale	13/50(6)	3/17(17)	14/70(5)	3/20(17)	7	17	0	2

Table 17: Calibration hygiene on a stressed window (FF30; calibrate on 2018–19 incl. Volmageddon and Q4-2018, deploy 2020–21, COVID event; 20 runs/cell; α_{mon}=.05). *split*: the deployed calibration protocol; *block-mean EB*: 21-day block-mean empirical-Bernstein bound, more conservative under dependence; *calm screen*: predeclared calm-regime screen (trailing volatility at most the training-period $q_{.75}$), under which the certificate refers to the screened null.

	B=5%			B=10%		
variant	$\bar{\mu}_0$	det (mix/plug)	med delay	$\bar{\mu}_0$	det (mix/plug)	med delay
split	0.446	1/1 of 20	212d	0.446	8/7 of 20	143d
block-mean EB	1.000	0/0 of 20	–	1.000	0/0 of 20	–
calm screen	0.480	1/1 of 20	216d	0.480	8/6 of 20	176d

exchangeability null alarms on 7, 11, and 17 of 20 runs at $B = 5\%$, 10%, and 20%, while the restarted mixture alarms on 0, 1, and 3; tightening the level to $\alpha = 0.001$ leaves the conformal counts at 0, 5, and 12 and the regret counts at 0, 1, and 0. Table 16 extends the comparison to the predeclared tail functionals, which see both crises clearly: the $q_{.90}$ exceedance rate increases from 0.18 to 0.54 across the GFC and from 0.04 to 0.22 across COVID, while the $q_{.99}$ rate increases from 0.06 to 0.30 across the GFC.

The loop closes on the bike history. Figure 9(d) reports the gated loop on the eight post-alarm Citi Bike candidates. The gate promotes 6 of 8 and reduces post-event regret by 7.3% on average, using a mean of 20.75 shadow labels per comparison. This result is stable across gate variants. Enlarging the shadow cap to 60 labels leaves the promotion count and regret reduction unchanged at $6/8$ and 7.3%; the strict variant at $\alpha_{\text{dep}} = 0.01$ promotes $5/8$ at 7.2%; the combined variant promotes $6/8$ at 6.9%; and adding a 180-day reverse-gate rollback leaves the base result unchanged. On FF30 the same loop is offered four post-alarm candidates and promotes none, recording 0 harmful endpoints, while deploying the same candidates without the gate produces one harmful endpoint, with the gate acting in the direction guaranteed by Theorem 5 on a real history.

Calibration hygiene. Table 17 calibrates deliberately on a stressed window (FF30 2018–2019, which contains the February 2018 volatility spike and the Q4-2018 drawdown) and deploys into 2020–2021. The deployed protocol yields $\overline{\mu}_0 = 0.446$, and its mixture arm detects 1 of 20 runs at $B = 5\%$ and 8 of 20 at $B = 10\%$; a predeclared calm-regime screen that restricts the null to rounds whose trailing volatility is at or below the training-period $q_{.75}$ yields $\overline{\mu}_0 = 0.480$ and detects 1 of 20 and 8 of 20.

A.11 Computational cost of adaptation

Table 18 reports the per-stream computational cost of the update arms, where one solve denotes one call to the exact optimizer. Each label requires one true-cost solve, candidate training requires $n_{\text{train}} \times (1 + 150)$ solves, and the shadow scan requires two decision solves per scanned round. The

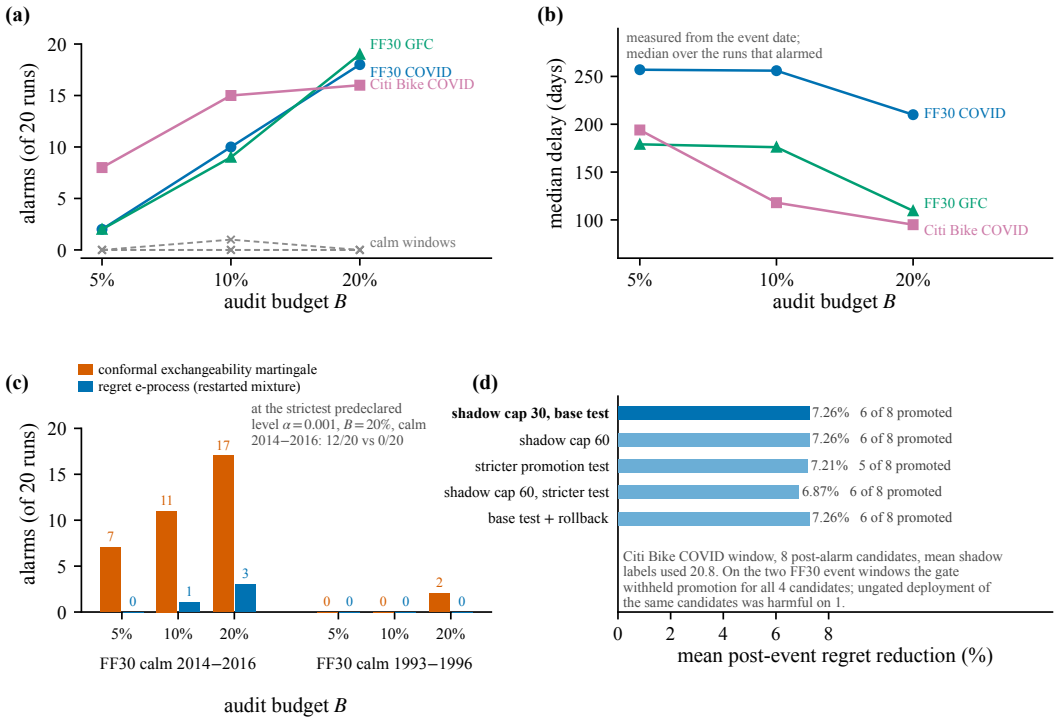

Figure 9: Both real deployment histories. Every cell is 20 runs (5 predictor seeds $\times$ 4 audit seeds) on one realized history, at $\alpha_{\text{cal}} = \alpha_{\text{mon}} = \alpha_{\text{dep}} = 0.05$. **(a)** Alarms out of 20 against the audit budget: FF30 COVID $2 \to 10 \to 18$, FF30 GFC $2 \to 9 \to 19$ and Citi Bike COVID $8 \to 15 \to 16$ at $B = 5\%, 10\%, 20\%$, against a single alarm across all 180 calm-window runs. **(b)** Median detection delay in days from the reference event, over the runs that alarm. **(c)** Two nulls on one shared FF30 audit stream and one ledger: on the calm 2014–2016 window the conformal exchangeability martingale alarms on 7, 11 and 17 of 20 runs, while the regret e-process (restarted mixture) alarms on 0, 1 and 3. **(d)** Mean post-event regret reduction and promotion counts for the eight post-alarm Citi Bike candidates under five gate variants; the deployed setting promotes $6/8$ at 7.3%. Points are exact counts, medians or means.

Table 18: Per-arm operational resource accounting for the update arms (label accounting of Algorithm 1; per-stream means, cost-structure drift, n=50/cell over 5 instances). "solves" are individual problem-solves: labels (one true-cost solve each), candidate training ($n_{\text{train}} \times (1{+}150)$), shadow scan (2 decision solves per scanned step). Deployment decisions (4,000/stream) are common to all arms. Arms share a 200-label ceiling; solver calls are not equalized across arms.

arm	prob./cell	labels	train solves	shadow solves	total extra	retrains
no-adapt	SP abrupt	0	0	0	0	0.0
guarded	SP abrupt	147	7248	52	7447	1.0
ungated	SP abrupt	125	7248	0	7373	1.0
oracle-timed	SP abrupt	50	7550	0	7600	1.0
scheduled (4×)	SP abrupt	194	22070	0	22264	3.0
scheduled (1×, cost-matched)	SP abrupt	194	7550	0	7744	1.0
no-adapt	SP gradual	0	0	0	0	0.0
guarded	SP gradual	151	4213	34	4398	0.6
ungated	SP gradual	137	4213	0	4350	0.6
oracle-timed	SP gradual	50	7550	0	7600	1.0
scheduled (4×)	SP gradual	194	22230	0	22424	3.0
scheduled (1×, cost-matched)	SP gradual	194	7550	0	7744	1.0
no-adapt	KP abrupt	0	0	0	0	0.0
guarded	KP abrupt	146	6795	42	6983	0.9
ungated	KP abrupt	129	6795	0	6924	0.9
oracle-timed	KP abrupt	50	7550	0	7600	1.0
scheduled (4×)	KP abrupt	197	22218	0	22415	3.0
scheduled (1×, cost-matched)	KP abrupt	197	7550	0	7747	1.0
no-adapt	KP gradual	0	0	0	0	0.0
guarded	KP gradual	142	3367	20	3529	0.5
ungated	KP gradual	133	3367	0	3501	0.5
oracle-timed	KP gradual	50	7550	0	7600	1.0
scheduled (4×)	KP gradual	195	22254	0	22449	3.0
scheduled (1×, cost-matched)	KP gradual	195	7550	0	7745	1.0

shadow comparison is inexpensive by this measure: 52 of the 7,447 extra solves on SP abrupt, 34 of 4,398 on SP gradual, 42 of 6,983 on KP abrupt, and 20 of 3,529 on KP gradual, because it scans decision disagreements rather than the entire stream. The 4,000 deployment decisions per stream are common to every arm and are excluded from the extra counts.

B PROOFS

Standing notation. Throughout, $(\Omega, \mathcal{A}, \mathbb{P})$ carries the calibration data and the deployment stream. $\mathcal{F}_0$ contains all calibration information (both splits cal-A and cal-B, hence R_q, $\widehat{\eta}$, and $\overline{\mu}_0$); $\mathcal{F}_{t-1}$ contains $\mathcal{F}_0$ and all deployment information through round $t-1$; and $\mathcal{G}_t = \mathcal{F}_{t-1} \vee \sigma(x_t, \widehat{c}_t, w_t, p_t)$ is the pre-outcome field for round t, as in Section 3.1. We write $I_t = \mathbf{1}\{\sum_{s<t} A_s < K_{\text{det}}\}$ for the active-ledger indicator, $Z_t = A_t Y_t / p_t$, and $X_t = Z_t - I_t \overline{\mu}_0$ as in equation 3, and $a_+ = \max\{a, 0\}$. All filtrations are discrete-time, and all conditional expectations are understood as fixed versions. When a statement concerns a property of a conditional law, we assume, as is standard, that the underlying observation spaces are standard Borel, so regular conditional distributions exist and such properties define measurable events.

Standing assumptions. Three assumptions are stated here for reference and used at the points indicated.

(A1) *Calibration ordering.* In Theorem 2(ii), the null level is deterministic and satisfies $\mu_0 \leq \overline{\mu}_0 < \mu_0 + \Delta$, so the margin $\delta = \Delta - (\overline{\mu}_0 - \mu_0)$ lies in $(0, \Delta]$. The left inequality places the pre-onset stream under the null of Theorem 1, whereas the right inequality supplies the post-onset drift. For a calibrated random $\overline{\mu}_0$, validity holds unconditionally at level $\alpha_{\text{cal}} + \alpha_{\text{mon}}$ by Theorem 1, and the delay bound holds on the event $\{\mu_0 \leq \overline{\mu}_0 < \mu_0 + \Delta\}$ when the grid covers the realized margin.

(A2) *Ledger-free companion.* In Theorem 2(iii), the hypotheses of part (ii) hold with the ledger finite at cap K_{det}, D^{∞} denotes the part-(ii) delay bound for the ledger-free companion process, and the audit coins $(U_t)_{t>\nu}$ are conditionally i.i.d. $\mathrm{Unif}[0,1]$ given $\mathcal{F}_\nu$.

(A3) *Population budget.* Membership in the sampler class $\mathcal{P}(B)$ of equation 17 requires the population condition $\mathbb{E}_\mu[p] \le B$. A deployed sampler, including the clipped, $\widehat{\eta}$-calibrated rule of equation 2, is required to satisfy this condition, and $\widehat{\eta}$ is calibrated to that target.

B.1 Auxiliary results

We record the standard results used below, each in the exact form in which it is invoked. Facts B.1–B.8 are quoted from the literature; Lemmas B.9–B.11 are proved here.

Fact B.1 (Ville's inequality, conditional form). *Let $\mathcal{H} \subseteq \mathcal{F}_{t_0}$ be a sub-σ-field, and let $(W_t)_{t\ge t_0}$ be a nonnegative supermartingale with respect to $(\mathcal{F}_t)_{t\ge t_0}$, i.e. $W_t \ge 0$ and $\mathbb{E}[W_t \mid \mathcal{F}_{t-1}] \le W_{t-1}$ for $t > t_0$. Then, for every $u > 0$,*

$$\mathbb{P}\Big(\sup_{t\ge t_0} W_t \ge u \,\Big|\, \mathcal{H}\Big) \;\le\; \frac{\mathbb{E}[W_{t_0} \mid \mathcal{H}]}{u}.$$

Reference: Ville (1939). Applying the unconditional statement to (W_t) under the regular conditional law given $\mathcal{H}$ yields the displayed form. In all applications below, W_{t_0} is $\mathcal{H}$-measurable, so the right-hand side is W_{t_0}/u.

Fact B.2 (optional stopping for submartingales at bounded times). *Let $(M_t)_{t\ge t_0}$ be a submartingale with respect to $(\mathcal{F}_t)_{t\ge t_0}$, and let S be a stopping time satisfying $t_0 \le S \le t_0 + n$ almost surely for a finite n. Then M_S is integrable and $\mathbb{E}[M_S \mid \mathcal{F}_{t_0}] \ge M_{t_0}$. Reference:* Durrett (2019, Thm. 4.8.4).

Fact B.3 (Freedman's maximal inequality). *Let $(\xi_t)_{t=1}^n$ be a martingale difference sequence with respect to a filtration $(\mathcal{O}_t)$, with $|\xi_t| \le b$ almost surely and $\sum_{t=1}^n \mathbb{E}[\xi_t^2 \mid \mathcal{O}_{t-1}] \le \overline{V}$ almost surely for a deterministic $\overline{V}$. Then, for every $a > 0$,*

$$\mathbb{P}\Big(\max_{k\le n} \sum_{t=1}^{k} \xi_t \;\ge\; a\Big) \;\le\; \exp\Big(-\frac{a^2}{2(\overline{V} + ba/3)}\Big).$$

Reference: Freedman (1975).

Fact B.4 (Bernstein's inequality for bounded independent summands). *Let $\zeta_1, \dots, \zeta_n$ be independent, with $\zeta_t \in [0,1]$, $S = \sum_t \zeta_t$, $\mathbb{E}S = m$, and $\mathrm{Var}(S) \le V$. Then, for every $s > 0$, $\mathbb{P}(S - m \ge s) \le \exp\big(-s^2/(2(V + s/3))\big)$. Reference:* Boucheron et al. (2013, Thm. 2.10, Cor. 2.11).

Fact B.5 (one-sided empirical-Bernstein upper confidence bound). *Let $Y_1, \dots, Y_n$ be i.i.d. with values in $[0,1]$, sample mean $\overline{Y}$, and sample variance $\widehat{v} = \frac{1}{n(n-1)} \sum_{i<j} (Y_i - Y_j)^2$. Then, with probability at least $1 - \alpha$,*

$$\mathbb{E}[Y] \;\le\; \overline{Y} + \sqrt{\frac{2\widehat{v}\log(2/\alpha)}{n}} + \frac{7\log(2/\alpha)}{3(n-1)}.$$

Reference: Maurer & Pontil (2009, Thm. 4).

Fact B.6 (Ryff's representation of a function by its rearrangement). *Let $f \in L^1_+(0,1)$, and let f^* denote its nonincreasing rearrangement on $(0,1)$. Then there exists a measure-preserving map $\sigma_0 : (0,1) \to (0,1)$ such that $f = f^* \circ \sigma_0$ Lebesgue-almost everywhere. Reference:* Ryff (1970).

Fact B.7 (Chebyshev's correlation inequality). *(a) Discrete form: if $u_1 \ge \cdots \ge u_n$ and $v_1 \ge \cdots \ge v_n$ are similarly ordered real sequences, then $\frac{1}{n}\sum_i u_i v_i \ge \big(\frac{1}{n}\sum_i u_i\big)\big(\frac{1}{n}\sum_i v_i\big)$. (b) Integral form: if $u, v : (0,1) \to [0,\infty)$ are both nonincreasing and $uv \in L^1(0,1)$, then $\int_0^1 uv \ge \big(\int_0^1 u\big)\big(\int_0^1 v\big)$. Reference:* Hardy et al. (1952, Thms. 43 and 236).

Fact B.8 (Jensen's inequality with its equality case). *Let μ be a probability measure, let φ be strictly convex on an interval J, and let W be a J-valued integrable random variable. Then $\mathbb{E}_\mu[\varphi(W)] \geq \varphi(\mathbb{E}_\mu[W])$, with equality if and only if W is μ-almost surely constant.* We use $\varphi(u) = 1/u$ on $J = (0, 1]$, which is strictly convex because $\varphi''(u) = 2u^{-3} > 0$.

Lemma B.9 (elementary logarithm bound). *For every $u \geq -1/2$, $\log(1 + u) \geq u - u^2$.*

Proof. Let $h(u) = \log(1 + u) - u + u^2$ on $[-1/2, \infty)$. Then $h(0) = 0$ and

$$h'(u) = \frac{1}{1+u} - 1 + 2u = \frac{1 - (1+u) + 2u(1+u)}{1+u} = \frac{u(1+2u)}{1+u}.$$

For $u \in (-1/2, 0)$, the numerator $u(1 + 2u)$ is negative because its first factor is negative and its second is positive, while $1 + u > 0$. Thus, $h' < 0$ and h is strictly decreasing on this interval, which implies $h(u) > h(0) = 0$ on $(-1/2, 0)$. For $u > 0$, both factors are positive, so $h' > 0$ and $h(u) > h(0) = 0$. Finally, $h(-1/2) = \log(1/2) + 1/2 + 1/4 = 0.0568\ldots > 0$. Therefore, $h \geq 0$ on $[-1/2, \infty)$, which proves the claim. □

Lemma B.10 (expected hitting time of a drifting log-wealth process). *Fix an integer $t_0 \geq 0$ and let $(W_t)_{t \geq t_0}$ satisfy $W_{t_0} > 0$ and $W_t = W_{t-1}(1 + u_t)$ for $t > t_0$, where u_t is $\mathcal{F}_t$-measurable and, almost surely,*

$$u_t \geq -\tfrac{1}{2}, \qquad u_t \leq \overline{u} < \infty, \qquad \mathbb{E}[\log(1 + u_t) \mid \mathcal{F}_{t-1}] \geq g > 0.$$

Let $b \in \mathbb{R}$ and $N = \inf\{t \geq t_0 : W_t \geq e^b\}$. Then $N < \infty$ almost surely and

$$\mathbb{E}\big[N - t_0 \,\big|\, \mathcal{F}_{t_0}\big] \;\leq\; \frac{\big(b - \log W_{t_0}\big)_+ + \log(1 + \overline{u})}{g}.$$

Proof. Step 1 (a submartingale). For $t \geq t_0$, set $M_t = \log W_t - \log W_{t_0} - (t - t_0)g$, so that $M_{t_0} = 0$ and $M_t - M_{t-1} = \log(1 + u_t) - g$. By hypothesis, $\mathbb{E}[M_t - M_{t-1} \mid \mathcal{F}_{t-1}] \geq 0$, so $(M_t)_{t \geq t_0}$ is a submartingale. Its increments are bounded in absolute value by $\overline{C} := \max\{\log(1+\overline{u}), \log 2\} + g < \infty$, because $\log(1+u_t) \leq \log(1+\overline{u})$ and $\log(1+u_t) \geq \log(1/2) = -\log 2$; hence, $|M_{t_0+n}| \leq n\overline{C}$ is integrable for each n.

Step 2 (optional stopping). If $W_{t_0} \geq e^b$, then $N = t_0$ and the claim is immediate because the right-hand side is nonnegative. Assume $W_{t_0} < e^b$. For each $n \geq 1$, $N \wedge (t_0 + n)$ is a stopping time bounded by $t_0 + n$, so Fact B.2 gives $\mathbb{E}[M_{N \wedge (t_0+n)} \mid \mathcal{F}_{t_0}] \geq M_{t_0} = 0$, that is,

$$g\,\mathbb{E}\big[(N \wedge (t_0 + n)) - t_0 \,\big|\, \mathcal{F}_{t_0}\big] \;\leq\; \mathbb{E}\big[\log W_{N \wedge (t_0+n)} - \log W_{t_0} \,\big|\, \mathcal{F}_{t_0}\big]. \tag{22}$$

Step 3 (pathwise overshoot control). We bound the right-hand side of equation 22 pathwise. On $\{N \leq t_0 + n\}$, the value at time $N-1$ is below e^b by the definition of N (and $N > t_0$ because $W_{t_0} < e^b$), and the next step multiplies the wealth by $1 + u_N \leq 1 + \overline{u}$; hence, $\log W_N \leq b + \log(1 + \overline{u})$. On $\{N > t_0 + n\}$, we have $\log W_{t_0+n} < b$. In both cases, $\log W_{N \wedge (t_0+n)} \leq b + \log(1 + \overline{u})$. Excursions of $\log W$ below its starting value only decrease the left-hand side of equation 22 and therefore do not affect the bound. Substituting gives

$$g\,\mathbb{E}\big[(N \wedge (t_0 + n)) - t_0 \,\big|\, \mathcal{F}_{t_0}\big] \;\leq\; \big(b - \log W_{t_0}\big)_+ + \log(1 + \overline{u}) \qquad \text{for every } n \geq 1.$$

Step 4 (monotone convergence). $N \wedge (t_0 + n) \uparrow N$ pointwise as $n \to \infty$, so conditional monotone convergence applied to the nonnegative random variables $(N \wedge (t_0 + n)) - t_0$ yields the stated bound for $\mathbb{E}[N - t_0 \mid \mathcal{F}_{t_0}]$. The finiteness of this conditional expectation implies that $N < \infty$ almost surely. □

Lemma B.11 (Markov's inequality, conditional form). *If $V \geq 0$ and $\mathcal{H}$ is a sub-σ-field, then $\mathbb{P}(V > x \mid \mathcal{H}) \leq \mathbb{E}[V \mid \mathcal{H}]/x$ for every $x > 0$.* This result follows by applying Markov's inequality under a regular conditional law.

B.2 Proof of Theorem 1 (anytime validity under audited feedback)

Restatement. Theorem 1 (restated). *Assume* $0 < p_{\min} \le p_t \le p_{\max} \le 1$ *on active rounds and* $\overline{\mu}_0 > 0$*. Under the conditional-mean null*

$$H_0(\overline{\mu}_0): \qquad \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] \le \overline{\mu}_0 \quad \text{for every } t,$$

$(\mathcal{E}_t)_{t\ge 0}$ *of equation 5 is a nonnegative supermartingale conditionally on* $\mathcal{F}_0$*, and* $\mathbb{P}(\tau_{\text{base}} < \infty \mid \mathcal{F}_0) \le \alpha_{\text{mon}}$*. The restarted mixture in equation 6 also satisfies* $\mathbb{P}(\tau_{\text{rmix}} < \infty \mid \mathcal{F}_0) \le \alpha_{\text{mon}}$*. If the deployment pairs* (x_t, c_t) *are i.i.d. from* P_{src} *and independent of calibration, then* $\mathbb{P}_{H_{\text{src}}}(\tau_{\text{rmix}} < \infty) \le \alpha_{\text{cal}} + \alpha_{\text{mon}} =: \alpha_{\text{total}}$*, as in equation 7. The same result holds, with separate error spending, for any bounded functional* $\phi(\rho) \in [0,1]$ *fixed before deployment.*

Proof. Step 1 (the information order, and why it is exactly what is needed). Three measurability facts drive the whole argument, and each is used once.

(a) p_t *is* $\mathcal{G}_t$*-measurable.* By construction, $p_t = \text{clip}(\widehat{\eta} s(x_t), p_{\min}, p_{\max})$ in equation 2 is a fixed function of x_t and the $\mathcal{F}_0$-measurable constant $\widehat{\eta}$, while $\mathcal{G}_t$ contains $\sigma(x_t)$ by definition. This measurability allows p_t to be pulled out of the conditional expectation in Step 2 before the audit coin is drawn.

(b) I_t *is* $\mathcal{F}_{t-1}$*-measurable.* $I_t = \mathbf{1}\{\sum_{s<t} A_s < K_{\text{det}}\}$ depends only on audit indicators strictly preceding t. Consequently, I_t may be treated as a constant under $\mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$, and the ledger truncation introduces no look-ahead.

(c) λ_t *is* $\mathcal{F}_{t-1}$*-measurable.* The bet is computed from past Horvitz–Thompson terms only and never sees x_t. This is the step that makes a *marginal* null level sufficient. Indeed, suppose one allowed λ_t to be $\mathcal{G}_t$-measurable. Then

$$\mathbb{E}[\lambda_t X_t \mid \mathcal{F}_{t-1}] = \mathbb{E}\big[\lambda_t \, \mathbb{E}[X_t \mid \mathcal{G}_t] \,\big|\, \mathcal{F}_{t-1}\big] = \mathbb{E}\big[\lambda_t I_t \big(\mathbb{E}[Y_t \mid \mathcal{G}_t] - \overline{\mu}_0\big) \,\big|\, \mathcal{F}_{t-1}\big],$$

and $\mathbb{E}[Y_t \mid \mathcal{G}_t] - \overline{\mu}_0$ is a function of x_t whose $\mathcal{F}_{t-1}$-average is nonpositive under $H_0(\overline{\mu}_0)$ but which is positive on the set of contexts with above-average conditional regret. A bet that loads on exactly that set would then have strictly positive conditional mean, and no supermartingale property would hold at the stated null. Requiring $\lambda_t \in \mathcal{F}_{t-1}$ removes this correlation, so only the average matters. The audit *probability* p_t, in contrast, may depend on x_t: it enters through the exact inverse-propensity identity of Step 2 rather than through the bet.

Step 2 (the inverse-propensity identity). Fix t and condition on $\mathcal{G}_t \vee \sigma(Y_t)$. The audit coin satisfies $U_t \sim \text{Unif}[0,1]$ conditionally independent of (c_t, Y_t) given $\mathcal{G}_t$, and $A_t = I_t \mathbf{1}\{U_t \le p_t\}$ with I_t and p_t both measurable with respect to $\mathcal{G}_t$ by Step 1(a)–(b). Hence

$$\mathbb{E}[A_t \mid \mathcal{G}_t, Y_t] = I_t \, \mathbb{P}(U_t \le p_t \mid \mathcal{G}_t, Y_t) = I_t p_t,$$

and therefore, dividing by the $\mathcal{G}_t$-measurable and strictly positive p_t,

$$\mathbb{E}[Z_t \mid \mathcal{G}_t, Y_t] = \mathbb{E}\Big[\frac{A_t Y_t}{p_t} \,\Big|\, \mathcal{G}_t, Y_t\Big] = \frac{Y_t}{p_t}\, \mathbb{E}[A_t \mid \mathcal{G}_t, Y_t] = I_t Y_t, \tag{23}$$

which is the third identity in equation 3. Integrability follows because, on active rounds, $0 \le Z_t \le Y_t/p_t \le 1/p_{\min} < \infty$, whereas on inactive rounds, $Z_t = 0$; hence $|X_t| \le 1/p_{\min} + \overline{\mu}_0$ is bounded. Note that equation 23 is an identity for the *deployment* stream: the audit policy changes the conditional variance of Z_t (it multiplies the second moment by $1/p_t$; see Section 3.3) but not its conditional mean.

Step 3 (the centered increment has nonpositive conditional mean). By the tower property, equation 23, and the $\mathcal{F}_{t-1}$-measurability of I_t,

$$\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = \mathbb{E}\big[\mathbb{E}[Z_t \mid \mathcal{G}_t, Y_t] \,\big|\, \mathcal{F}_{t-1}\big] = \mathbb{E}[I_t Y_t \mid \mathcal{F}_{t-1}] = I_t \, \mathbb{E}[Y_t \mid \mathcal{F}_{t-1}].$$

Since $X_t = Z_t - I_t \overline{\mu}_0$ and $I_t \in \{0,1\}$ is $\mathcal{F}_{t-1}$-measurable,

$$\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = I_t\big(\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] - \overline{\mu}_0\big) \;\le\; 0 \qquad \text{under } H_0(\overline{\mu}_0), \tag{24}$$

because the bracket is nonpositive under the null and $I_t \ge 0$.

Step 4 (each factor is nonnegative and has conditional mean at most one). For nonnegativity, $Z_t \geq 0$ gives $X_t \geq -I_t\overline{\mu}_0 \geq -\overline{\mu}_0$, and $0 \leq \lambda_t \leq (2\overline{\mu}_0)^{-1}$ gives

$$1 + \lambda_t X_t \;\geq\; 1 - \lambda_t\overline{\mu}_0 \;\geq\; 1 - \tfrac{1}{2} = \tfrac{1}{2} \;>\; 0. \tag{25}$$

For the conditional mean, since λ_t is $\mathcal{F}_{t-1}$-measurable and bounded, equation 24 gives $\mathbb{E}[1+\lambda_t X_t \mid \mathcal{F}_{t-1}] = 1+\lambda_t\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leq 1$. This is the precise sense in which the ordering "p_t chosen from x_t, then the bet from the past only, then the coin, then the outcome" produces a conditional-mean-one increment.

Step 5 (supermartingale, and the ledger). $\mathcal{E}_t = \mathcal{E}_{t-1}(1+\lambda_t X_t)$ with $\mathcal{E}_{t-1} \geq 0$ and $\mathcal{F}_{t-1}$-measurable, so by Step 4

$$\mathbb{E}[\mathcal{E}_t \mid \mathcal{F}_{t-1}] = \mathcal{E}_{t-1}\,\mathbb{E}[1 + \lambda_t X_t \mid \mathcal{F}_{t-1}] \leq \mathcal{E}_{t-1},$$

and $\mathcal{E}_t \geq 0$ by equation 25. Thus $(\mathcal{E}_t)_{t\geq 0}$ is a nonnegative supermartingale with $\mathcal{E}_0 = 1$, which is $\mathcal{F}_0$-measurable. When the ledger is exhausted, $I_t = 0$, hence $A_t = 0$, $Z_t = 0$ and $X_t = 0$, so the factor equals 1 and the process is constant from that round on; no separate stopping argument is needed, and in particular the supermartingale property is not disturbed by the cap.

Step 6 (Ville, base process). Applying Fact B.1 with $\mathcal{H} = \mathcal{F}_0$, $t_0 = 0$, and $u = \alpha_{\mathrm{mon}}^{-1}$ gives

$$\mathbb{P}(\tau_{\mathrm{base}} < \infty \mid \mathcal{F}_0) = \mathbb{P}\Big(\sup_{t\geq 0} \mathcal{E}_t \geq \alpha_{\mathrm{mon}}^{-1} \;\Big|\; \mathcal{F}_0\Big) \leq \alpha_{\mathrm{mon}}\,\mathcal{E}_0 = \alpha_{\mathrm{mon}}.$$

Step 7 (the restarted mixture). Fix a restart time $r \in \mathcal{S}$. The set $\mathcal{S}$ is predeclared and therefore deterministic, so $\mathcal{F}_{r-1}$ is well defined, and the union in Step 7(c) is over a fixed countable index set.

(a) *Each component.* For a fixed grid point $\lambda_j \in \Lambda$, the process $t \mapsto \prod_{s=r}^{t}(1 + \lambda_j X_s)$, started at value 1 at time $r-1$, is a nonnegative supermartingale by Steps 4–5: constants are $\mathcal{F}_{s-1}$-measurable, and $\lambda_j \leq (2\overline{\mu}_0)^{-1}$ is the same admissibility condition.

(b) *Mixing.* $\mathcal{E}^{\mathrm{mix}}_{r,t} = \sum_j \pi_j \prod_{s=r}^{t}(1 + \lambda_j X_s)$ is a convex combination ($\pi_j > 0$, $\sum_j \pi_j = 1$) of nonnegative supermartingales with respect to the *same* filtration and is therefore itself a nonnegative supermartingale, with $\mathcal{E}^{\mathrm{mix}}_{r,r-1} = \sum_j \pi_j = 1$. This conclusion uses only the linearity of conditional expectation.

(c) *Restart spending.* Applying Fact B.1 conditionally on $\mathcal{F}_{r-1}$ at level $u = (\alpha_{\mathrm{mon}}\omega_r)^{-1}$ gives

$$\mathbb{P}\Big(\exists t \geq r :\; \omega_r\mathcal{E}^{\mathrm{mix}}_{r,t} \geq \alpha_{\mathrm{mon}}^{-1} \;\Big|\; \mathcal{F}_{r-1}\Big) \;\leq\; \alpha_{\mathrm{mon}}\omega_r.$$

Taking $\mathbb{E}[\cdot \mid \mathcal{F}_0]$ and summing over $r \in \mathcal{S}$ by countable subadditivity yields

$$\mathbb{P}(\tau_{\mathrm{rmix}} < \infty \mid \mathcal{F}_0) \leq \sum_{r\in\mathcal{S}} \alpha_{\mathrm{mon}}\omega_r = \alpha_{\mathrm{mon}} \sum_{r\in\mathcal{S}} \omega_r \leq \alpha_{\mathrm{mon}},$$

where the last inequality uses $\sum_r \omega_r \leq 1$. This is the restart spending rule. The copies use the same audited stream and consume no additional labels because X_s is shared.

Step 8 (from the i.i.d. source null to $H_0(\overline{\mu}_0)$: what calibration buys). Assume that $(x_t, c_t)_{t\geq 1}$ are i.i.d. from P_{src} and independent of the calibration data. Two points require care.

(a) *The monitored functional is calibration-measurable.* $Y_t = \min\{\rho_t/R_q, 1\}$ depends on the cal-A quantity R_q, so $\mu_0 = \mathbb{E}_{P_{\mathrm{src}}}[Y]$ is itself a random variable that is measurable with respect to cal-A and hence with respect to $\mathcal{F}_0$. Conditionally on $\mathcal{F}_0$, the variables $(Y_t)_{t\geq 1}$ are i.i.d. with mean μ_0. Since $\mathcal{F}_{t-1}$ is generated by $\mathcal{F}_0$ and $(x_s, c_s, U_s)_{s<t}$, and (x_t, c_t) is independent of the latter given $\mathcal{F}_0$,

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = \mu_0 \qquad \text{for every } t. \tag{26}$$

(b) *The calibration event.* On the event $\mathcal{C}_{\mathrm{cal}} = \{\mu_0 \leq \overline{\mu}_0\}$ in equation 4, equation 26 yields $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] \leq \overline{\mu}_0$, which is exactly $H_0(\overline{\mu}_0)$. The probability bound $\mathbb{P}(\mathcal{C}_{\mathrm{cal}}) \geq 1 - \alpha_{\mathrm{cal}}$ follows from Fact B.5 applied *conditionally on cal-A*. Given cal-A, the functional $y \mapsto \min\{y/R_q, 1\}$ is fixed, and the cal-B values are i.i.d. draws of that fixed $[0,1]$-valued functional. The one-sided empirical-Bernstein bound thus holds conditionally at level α_{cal}, and the tower property removes the conditioning. This is precisely why two independent splits are needed: cal-B must bound the mean of a functional selected on cal-A.

(c) *Union bound.* $\mathcal{C}_{\mathrm{cal}}$ is $\mathcal{F}_0$-measurable, so

$$\mathbb{P}(\tau_{\mathrm{rmix}} < \infty) \le \mathbb{P}(\mathcal{C}_{\mathrm{cal}}^c) + \mathbb{E}\big[\mathbf{1}_{\mathcal{C}_{\mathrm{cal}}}\,\mathbb{P}(\tau_{\mathrm{rmix}} < \infty \mid \mathcal{F}_0)\big] \le \alpha_{\mathrm{cal}} + \alpha_{\mathrm{mon}} = \alpha_{\mathrm{total}},$$

which is equation 7. The conditioning carries mathematical content: the supermartingale property is asserted on $\mathcal{C}_{\mathrm{cal}}$, where the calibrated level dominates the source mean, and the complement is charged at α_{cal} by the union bound. This two-term accounting is exactly what makes the guarantee finite-sample and distribution-free.

Step 9 (arbitrary predeclared bounded functionals). Let $\phi : [0,\infty) \to [0,1]$ be fixed before deployment and replace Y_t by $\phi(\rho_t)$ throughout, with null level $\overline{\mu}_{0,\phi}$ the cal-B one-sided empirical-Bernstein bound for $\mathbb{E}_{P_{\mathrm{src}}}[\phi(\rho)]$. Inspecting Steps 1–8, the only properties of Y_t that were used are: it takes values in $[0,1]$ (used for the bound $Z_t \le 1/p_{\min}$ and for Fact B.5); it is not observed before the audit coin (used in equation 23); and its conditional mean is bounded by the calibrated level (Step 3, Step 8). All three hold verbatim for $\phi(\rho_t)$. For the exceedance functional $\phi_q(\rho) = \mathbf{1}\{\rho > R_q\}$ the threshold R_q is cal-A-measurable, so ϕ_q is predeclared conditionally on cal-A, exactly as for the clipped mean, and Step 8(b) goes through unchanged. Finally, if k functionals are monitored simultaneously with monitoring levels $\alpha_{\mathrm{mon}}^{(1)}, \dots, \alpha_{\mathrm{mon}}^{(k)}$ summing to α_{mon} and calibration levels summing to α_{cal}, a union bound over the k monitors and the k calibration events preserves $\alpha_{\mathrm{total}} \le \alpha_{\mathrm{cal}} + \alpha_{\mathrm{mon}}$. □

Remark B.1 (the reach of the sequential argument). Steps 1–7 use exactly one property of the deployment stream: the conditional-mean null $H_0(\overline{\mu}_0)$. Theorem 1 therefore holds verbatim for *any* stream, dependent or not, satisfying $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] \le \overline{\mu}_0$ almost surely — the supermartingale, the mixture, the restart spending rule and Ville's inequality are all insensitive to temporal dependence. The i.i.d. hypothesis is used in Step 8 alone, where it converts the *marginal* source bound delivered by the cal-B empirical-Bernstein bound into the conditional-mean null. Independent calibration is thus the single interface through which the sampling model enters the guarantee.

B.3 Proof of Theorem 2 (budget–delay characterization)

Restatement. Theorem 2 (restated). *(i) Lower bound. Let $d^\star = \lfloor \log(1/\alpha)/(16B\kappa) \rfloor$ and assume $1 \le d^\star \le T/2$. For every outcome-oblivious audit policy and every admissible alarm time, at least half of the possible onsets ν satisfy, for every alternative in $\mathcal{D}(\kappa, M)$,*

$$\mathbb{E}_\nu[(\tau - \nu)_+] \ge d^\star(1 - \alpha - \sqrt{\alpha} - \alpha^\gamma), \qquad \gamma = \frac{3\kappa}{24M^2 + 16\kappa M},$$

as stated in equation 8. Hence, $\sup_\nu \mathbb{E}_\nu[(\tau-\nu)_+] = \Omega(\log(1/\alpha)/(B\kappa))$ whenever $\alpha + \sqrt{\alpha} + \alpha^\gamma \le 1/2$. Among deterministic expected-intensity profiles with average rate B, the constant profile is the unique profile whose expected audit mass is at least dB in every window of length d, simultaneously for all $d \ge 1$.

(ii) Matching upper bound. Suppose the drift is regret-sufficient, that is, the conditional law of c_t given (x_t, Y_t) is unchanged. Suppose also that the conditional regret kernels are Bernoulli, their means lie in $[\varsigma, 1-\varsigma]$, and the mean increases from μ_0 to $\mu_0 + \Delta$ after an onset $\nu \le T$. Then the cost kernels satisfy $\kappa \le \Delta^2/(\varsigma(1-\varsigma))$ and $M \le \Delta/\varsigma$, so part (i) gives a lower bound of order $\log(1/\alpha)/(B\Delta^2)$. Conversely, suppose $p_t = B$, the ledger is unbounded, and $\delta = \Delta - (\overline{\mu}_0 - \mu_0) > 0$. If a fixed-ratio bet grid contains a point within a fixed factor of $B\delta$, restart copies are launched every L rounds, and monitoring continues past T, then L can be chosen such that equation 9 holds,

$$\sup_{\nu \le T} \mathbb{E}_\nu[(\tau_{\mathrm{rmix}} - \nu)_+ \mid \mathcal{F}_\nu] \le C\frac{\log(T/\alpha_{\mathrm{mon}}) + \log m}{B\delta^2} =: D^\infty.$$

(iii) Finite ledger. Let $K_\nu = K_{\mathrm{det}} - \sum_{s \le \nu} A_s$ and $\tau_{\mathrm{cap}} = \inf\{t : \sum_{s \le t} A_s = K_{\mathrm{det}}\}$. Under the conditions of part (ii), if $K_\nu \ge 2B$, then equation 10 holds,

$$\mathbb{P}_\nu(\tau_{\mathrm{rmix}} \le \tau_{\mathrm{cap}} \mid \mathcal{F}_\nu) \ge 1 - \frac{4BD^\infty}{K_\nu} - \exp(-3K_\nu/16).$$

FORMAL SETTING FOR PART (I)

Fix $B \in (0, 1]$, $\kappa > 0$, $M > 0$, and a horizon T.

Audit policies. An *outcome-oblivious* audit policy is a sequence of deterministic measurable maps $g_t : \mathcal{X} \to [0, 1]$ fixed before deployment, with $p_t = g_t(x_t)$ and average budget

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_\mu[g_t(X)] \;\le\; B. \tag{27}$$

Audits are $A_t = \mathbf{1}\{U_t \le p_t\}$. This class contains uniform ($g_t \equiv B$), score-guided ($g_t = g$ a fixed function of x), and every front-loaded schedule; it excludes intensity that reacts to audited outcomes.

Alternatives. For an onset ν and cost-kernel sequences $(P_{C,t})$, $(Q_{C,t})$ define P^0 and P^ν as *product laws*: under either, $((x_t, U_t, W_t))_{t\ge1}$ are i.i.d. with $x_t \sim \mu$, $U_t \sim \mathrm{Unif}[0, 1]$, and W_t an exogenous monitor-randomization variable of fixed law, mutually independent; conditionally on x_t, and independently of everything else, $c_t \sim P_{C,t}(\cdot \mid x_t)$ under P^0 for every t, while under P^ν the kernel is $Q_{C,t}(\cdot \mid x_t)$ from $t = \nu$ on. The class $\mathcal{D}(\kappa, M)$ requires, for all t and all x,

$$\mathrm{KL}\big(Q_{C,t}(\cdot \mid x) \,\|\, P_{C,t}(\cdot \mid x)\big) \le \kappa, \qquad \sup_{c'} \big|\log r_t(x, c')\big| \le M, \tag{28}$$

for a jointly measurable version r_t of $dQ_{C,t}/dP_{C,t}$. The product form is not a technicality: without it, the one-step marginal conditions equation 28 do not pin down the joint observation law. (For instance, with $\mathcal{X}$ a single point and scalar costs, letting W_t, c_t be independent Bernoulli variables before the change and setting $c_t = W_t$ afterwards leaves both kernels and hence κ and M unchanged, yet the change is perfectly detectable from $(W_t, A_t c_t)$.) The feature marginal μ is the same under P^0 and P^ν, so the drift changes only the conditional cost kernel. The monitored regret is a measurable function $Y_t = h_Y(x_t, c_t)$ of the feature and the cost vector, because the prediction and decision are functions of x_t.

Observations and alarms. The monitor observes $\mathcal{O}_t = \sigma\big((x_s, U_s, W_s, A_s, A_s c_s) : s \le t\big)$, which includes the full cost vector of every audited round. Predictions $\widehat{c}_s$ and decisions w_s are deterministic functions of x_s, and the audited regret $A_s Y_s = A_s h_Y(x_s, A_s c_s)$ is a measurable function of $(x_s, A_s, A_s c_s)$; all are therefore $\mathcal{O}_s$-measurable, so every audited-regret monitor is covered. An *admissible alarm time* is a stopping time τ of $(\mathcal{O}_t)$ satisfying $\mathbb{P}^0(\tau \le T) \le \alpha$. The "possible onsets" referred to in the statement are $\nu \in \{1, \dots, T - d^\star + 1\}$, i.e. those for which a full window of length $d^\star$ fits within the horizon.

PROOF OF PART (I)

Proof. Step 0 (likelihood structure). Fix $u \ge \nu$ and compute the Radon–Nikodym derivative of P^ν with respect to P^0, restricted to $\mathcal{O}_u$. Under the product form, the coordinates (x_t, U_t, W_t) have the same law under both measures and are independent of the cost vectors conditional on the features. Conditional on (x_t, U_t, W_t), the audit indicator $A_t = \mathbf{1}\{U_t \le g_t(x_t)\}$ is deterministic and therefore contributes no likelihood weight. There are two cases in each round. If $A_t = 0$, the cost vector c_t is unobserved and integrates to one under both kernels, contributing a factor 1. If $A_t = 1$, the observed vector c_t contributes the factor $r_t(x_t, c_t) = e^{\ell_t}$, where $\ell_t := \log r_t(x_t, c_t)$ and $|\ell_t| \le M$ by equation 28. Conditional independence across t implies that these contributions multiply, so

$$\log \frac{dP^\nu}{dP^0}\bigg|_{\mathcal{O}_u} = \sum_{t=\nu}^{u} A_t \ell_t. \tag{29}$$

The same computation for $u = \nu - 1$ gives derivative 1, because every coordinate through round $\nu - 1$ has the same law under both measures. Thus,

$$P^\nu = P^0 \quad \text{on } \mathcal{O}_{\nu-1}. \tag{30}$$

Step 1 (budget pigeonhole: most onsets are audit-starved). Write $d = d^\star$ and, for $\nu \in \{1, \dots, T - d + 1\}$, let

$$S_\nu := \sum_{t=\nu}^{\nu+d-1} \mathbb{E}_\mu[g_t(X)],$$

a deterministic quantity because the policy is outcome-oblivious. Each round t belongs to at most d of these windows, so by equation 27

$$\sum_{\nu=1}^{T-d+1} S_\nu \;\le\; d\sum_{t=1}^{T}\mathbb{E}_\mu[g_t] \;\le\; dBT.$$

There are $T-d+1$ windows, and $d \le T/2$ gives $T-d+1 \ge T/2+1 > T/2$. Drawing ν uniformly from $\{1,\dots,T-d+1\}$ and applying Markov's inequality,

$$\mathbb{P}\big(S_\nu > 4dB\big) \;\le\; \frac{1}{4dB}\cdot\frac{dBT}{T-d+1} \;=\; \frac{T}{4(T-d+1)} \;<\; \frac{T}{4\cdot T/2} \;=\; \frac{1}{2}.$$

Hence the set $\mathcal{N} = \{\nu \le T-d+1 : S_\nu \le 4dB\}$ satisfies $|\mathcal{N}| \ge (T-d+1)/2$: at least half of the possible onsets are followed by a window of length d carrying expected audit mass at most $4dB$.

Step 2 (change of measure on the detection window). Fix $\nu \in \mathcal{N}$ and set $\mathrm{LR} := \sum_{t=\nu}^{\nu+d-1} A_t\ell_t$, the log-likelihood ratio on $\mathcal{O}_{\nu+d-1}$ given by equation 29. The event $E := \{\nu \le \tau < \nu+d\}$ is $\mathcal{O}_{\nu+d-1}$-measurable. Since $\nu+d-1 \le T$, we have $\mathbb{P}^0(E) \le \mathbb{P}^0(\tau \le T) \le \alpha$. For any threshold $c \in \mathbb{R}$, split according to $\{\mathrm{LR} \le c\}$,

$$\begin{aligned}\mathbb{P}^\nu(E) &= \mathbb{E}^0\big[e^{\mathrm{LR}}\mathbf{1}_E\big] = \mathbb{E}^0\big[e^{\mathrm{LR}}\mathbf{1}_E\mathbf{1}\{\mathrm{LR}\le c\}\big] + \mathbb{P}^\nu\big(E\cap\{\mathrm{LR}>c\}\big)\\ &\le e^c\,\mathbb{P}^0(E) + \mathbb{P}^\nu(\mathrm{LR}>c) \;\le\; e^c\alpha + \mathbb{P}^\nu(\mathrm{LR}>c).\end{aligned} \tag{31}$$

Step 3 (concentration of the log-likelihood ratio under P^ν). For $t \ge \nu$, under P^ν, the cost vector c_t is drawn from $Q_{C,t}(\cdot \mid x_t)$, so

$$\mathbb{E}^\nu[A_t\ell_t \mid \mathcal{O}_{t-1}] = \mathbb{E}_{x\sim\mu}\Big[g_t(x)\,\mathbb{E}_{c'\sim Q_{C,t}(\cdot|x)}\big[\log r_t(x,c')\big]\Big] = \mathbb{E}_\mu\big[g_t(X)\,\mathrm{KL}_t(X)\big] =: D_t,$$

where $\mathrm{KL}_t(x) = \mathrm{KL}(Q_{C,t}(\cdot \mid x)\|P_{C,t}(\cdot \mid x)) \in [0,\kappa]$. The outer expectation is over $x_t \sim \mu$, which is independent of $\mathcal{O}_{t-1}$, so D_t is deterministic. Set $\xi_t := A_t\ell_t - D_t$, which is a martingale difference sequence with respect to $(\mathcal{O}_t)$. We use the following two bounds:

(a) $|\xi_t| \le |A_t\ell_t| + |D_t| \le M + M = 2M$, using $|\ell_t| \le M$ and $D_t = \mathbb{E}_\mu[g_t\mathrm{KL}_t] \le \mathbb{E}_\mu[g_t]\cdot\sup_x \mathrm{KL}_t(x) \le M$. The last inequality follows because $\mathrm{KL}_t(x) = \mathbb{E}_Q[\log r_t] \le M$ by equation 28 and $g_t \le 1$.

(b) $\mathbb{E}^\nu[\xi_t^2 \mid \mathcal{O}_{t-1}] \le \mathbb{E}^\nu[(A_t\ell_t)^2 \mid \mathcal{O}_{t-1}] = \mathbb{E}_\mu\big[g_t(X)\,\mathbb{E}_Q[\ell_t^2 \mid X]\big] \le M^2\,\mathbb{E}_\mu[g_t]$, using $A_t^2 = A_t$, $\mathbb{E}[A_t \mid x_t] = g_t(x_t)$, and $|\ell_t| \le M$.

Summing (b) over the window and using $\nu \in \mathcal{N}$ yields

$$\overline{V} := \sum_{t=\nu}^{\nu+d-1}\mathbb{E}^\nu[\xi_t^2 \mid \mathcal{O}_{t-1}] \le M^2 S_\nu \le 4dBM^2,$$

which is a deterministic bound. Likewise, the drift term satisfies

$$\sum_{t=\nu}^{\nu+d-1} D_t \;\le\; \kappa\sum_{t=\nu}^{\nu+d-1}\mathbb{E}_\mu[g_t] = \kappa S_\nu \;\le\; 4dB\kappa. \tag{32}$$

Fact B.3 with $b = 2M$ then gives, for every $a > 0$,

$$\mathbb{P}^\nu\Big(\max_{k\le d}\sum_{t=\nu}^{\nu+k-1}\xi_t \ge a\Big) \le \exp\Big(-\frac{a^2}{2(\overline{V}+2Ma/3)}\Big). \tag{33}$$

Step 4 (choosing the constants; where $d^\star$ and γ come from). Write $L_0 = \log(1/\alpha)$ and take

$$c = \frac{L_0}{2}, \qquad a = \frac{L_0}{4}.$$

Because $d = d^\star = \lfloor L_0/(16B\kappa) \rfloor \le L_0/(16B\kappa)$, the drift sum in equation 32 satisfies

$$\sum_{t=\nu}^{\nu+d-1} D_t \;\le\; 4dB\kappa \;\le\; 4B\kappa \cdot \frac{L_0}{16B\kappa} = \frac{L_0}{4}.$$

This is precisely the role of the constant 16 in $d^\star$: it ensures that the expected evidence accumulated over the window is at most one-quarter of the budget L_0. Consequently,

$$\mathrm{LR} = \sum_{t=\nu}^{\nu+d-1} \xi_t + \sum_{t=\nu}^{\nu+d-1} D_t \le \sum_{t=\nu}^{\nu+d-1} \xi_t + \frac{L_0}{4}, \qquad \text{so} \qquad \{\mathrm{LR} > c\} \subseteq \Big\{ \max_{k \le d} \sum_{t=\nu}^{\nu+k-1} \xi_t > \frac{L_0}{4} \Big\}.$$

Similarly, $\overline{V} \le 4dBM^2 \le 4BM^2 \cdot L_0/(16B\kappa) = L_0 M^2/(4\kappa)$. Substituting $a = L_0/4$ and these two bounds into equation 33 gives

$$\begin{aligned} \mathbb{P}^\nu(\mathrm{LR} > c) &\le \exp\left(-\frac{L_0^2/16}{2\big(\frac{L_0 M^2}{4\kappa} + \frac{2M}{3} \cdot \frac{L_0}{4}\big)} \right) = \exp\left(-\frac{L_0/16}{2\big(\frac{M^2}{4\kappa} + \frac{M}{6}\big)} \right) \\ &= \exp\left(-\frac{L_0}{\frac{8M^2}{\kappa} + \frac{16M}{3}} \right), \end{aligned}$$

where the last step multiplies numerator and denominator by 16: $16 \cdot 2(M^2/(4\kappa) + M/6) = 8M^2/\kappa + 16M/3$. Clearing the fraction,

$$\frac{1}{\frac{8M^2}{\kappa} + \frac{16M}{3}} = \frac{3\kappa}{24M^2 + 16\kappa M} = \gamma, \qquad \text{so} \qquad \mathbb{P}^\nu(\mathrm{LR} > c) \le e^{-\gamma L_0} = \alpha^\gamma.$$

Also $e^c \alpha = e^{L_0/2}\alpha = \alpha^{-1/2}\alpha = \sqrt{\alpha}$. Plugging both into equation 31,

$$\mathbb{P}^\nu(\nu \le \tau < \nu + d^\star) \;\le\; \sqrt{\alpha} + \alpha^\gamma. \tag{34}$$

This yields both constants in equation 8: $d^\star$ follows from matching the expected evidence over the window to $L_0/4$, and γ follows from Freedman's denominator evaluated at $a = L_0/4$.

Step 5 (pre-onset alarms are equally rare under P^ν). The event $\{\tau < \nu\} = \{\tau \le \nu - 1\}$ lies in $\mathcal{O}_{\nu-1}$ because τ is a stopping time with respect to $(\mathcal{O}_t)$. By equation 30, $\mathbb{P}^\nu(\tau < \nu) = \mathbb{P}^0(\tau < \nu) \le \mathbb{P}^0(\tau \le T) \le \alpha$, since $\nu \le T$.

Step 6 (from a tail bound to the delay). Combining Steps 4 and 5,

$$\mathbb{P}^\nu(\tau \ge \nu + d^\star) = 1 - \mathbb{P}^\nu(\tau < \nu) - \mathbb{P}^\nu(\nu \le \tau < \nu + d^\star) \ge 1 - \alpha - \sqrt{\alpha} - \alpha^\gamma,$$

and since $(\tau - \nu)_+ \ge d^\star \mathbf{1}\{\tau \ge \nu + d^\star\}$ pointwise,

$$\mathbb{E}_\nu[(\tau - \nu)_+] \;\ge\; d^\star\big(1 - \alpha - \sqrt{\alpha} - \alpha^\gamma\big),$$

which is equation 8. This bound holds for every $\nu \in \mathcal{N}$, that is, for at least half of the possible onsets by Step 1, and for every alternative in $\mathcal{D}(\kappa, M)$ because Steps 0–5 use only equation 28. If $\alpha + \sqrt{\alpha} + \alpha^\gamma \le 1/2$, the bracket is at least $1/2$ and $\mathcal{N} \ne \emptyset$, so

$$\sup_\nu \mathbb{E}_\nu[(\tau - \nu)_+] \;\ge\; \frac{d^\star}{2} \;\ge\; \frac{1}{2}\Big(\frac{\log(1/\alpha)}{16B\kappa} - 1\Big) = \Omega\Big(\frac{\log(1/\alpha)}{B\kappa}\Big),$$

where we use $\lfloor z \rfloor \ge z - 1$ and $d^\star \ge 1$.

Step 7 (uniqueness of the constant intensity profile). Let $\beta_t := \mathbb{E}_\mu[g_t(X)]$ denote the expected audit intensity profile, with $\frac{1}{T}\sum_t \beta_t \le B$.

(a) *No profile beats dB uniformly.* Fix d dividing T and consider the T/d *disjoint* windows starting at $1, d+1, \ldots, T-d+1$. Their masses S_ν sum to $\sum_{t=1}^T \beta_t \le BT$, so the smallest of these masses, and therefore the minimum over all sliding windows of length d, is at most $BT/(T/d) = dB$. Thus, no profile can have window mass strictly greater than dB in every window of length d, whereas the constant profile $\beta_t \equiv B$ attains exactly dB in every window of every length.

(b) *Uniqueness.* Suppose a profile satisfies $S_\nu \ge dB$ for every window of every length d; taking $d = 1$ gives $\beta_t \ge B$ for every t. Combined with $\frac{1}{T}\sum_t \beta_t \le B$ this forces $\beta_t = B$ for every t. Hence the constant profile is the unique such profile: any front-loading (some $\beta_t > B$) must be compensated by $\beta_{t'} < B$ elsewhere, and the window starting at t' is then strictly starved. The condition is imposed for all d simultaneously; a single length $d \ge 2$ does not determine the profile. For example, with $d = 2$, even T, and small $\varepsilon > 0$, intensities alternating between $B + \varepsilon$ and $B - \varepsilon$ give mass exactly $2B$ in every window of length 2. □

PROOF OF PART (II)

Throughout this part, the audit rate is constant, $p_t \equiv B$ (so $p_{\min} = p_{\max} = B$), the ledger is unbounded ($I_t \equiv 1$, hence $X_t = Z_t - \overline{\mu}_0$), and monitoring is allowed to continue past T. By (A1), $\overline{\mu}_0$ is deterministic and satisfies $\mu_0 \le \overline{\mu}_0 < \mu_0 + \Delta$, so that $\delta = \Delta - (\overline{\mu}_0 - \mu_0) \in (0, \Delta]$ and $\overline{\mu}_0 \le \mu_0 + \Delta \le 1$. The post-change kernel governs rounds $t > \nu$. Fix the restart set $\mathcal{S} = \{1, L+1, 2L+1, \dots, RL+1\}$ with $R = \lceil T/L \rceil$, so $|\mathcal{S}| = R+1$, and take uniform weights $\omega_r = 1/(R+1)$ and $\pi_j = 1/m$.

Proof. Step 1 (the class constants for Bernoulli kernels). Let $P_{Y,t}$ and $Q_{Y,t}$ denote the regret kernels induced by $P_{C,t}$ and $Q_{C,t}$ through h_Y. Under regret-sufficient drift, the conditional law of c_t given (x_t, Y_t) is the same before and after the onset, so the cost likelihood ratio factors through the regret, $r_t(x, c') = (dQ_{Y,t}/dP_{Y,t})(x, h_Y(x, c'))$. The chain rule for Kullback–Leibler divergence then gives $\mathrm{KL}(Q_{C,t}(\cdot \mid x) \| P_{C,t}(\cdot \mid x)) = \mathrm{KL}(Q_{Y,t}(\cdot \mid x) \| P_{Y,t}(\cdot \mid x))$, and $\sup_{c'} |\log r_t(x, c')| \le \sup_y |\log(dQ_{Y,t}/dP_{Y,t})(x, y)|$. It therefore suffices to bound the constants of the regret kernels. Let $P_Y = \mathrm{Bern}(\mu_0)$ and $Q_Y = \mathrm{Bern}(\mu_0 + \Delta)$, with both means in $[\varsigma, 1-\varsigma]$. For Bernoulli laws, the Kullback–Leibler divergence is bounded above by the χ^2 divergence,

$$\mathrm{KL}\big(\mathrm{Bern}(q) \,\|\, \mathrm{Bern}(p)\big) \;\le\; \chi^2\big(\mathrm{Bern}(q) \,\|\, \mathrm{Bern}(p)\big) = \frac{(q-p)^2}{p(1-p)} \;\le\; \frac{\Delta^2}{\varsigma(1-\varsigma)} =: \kappa,$$

using $\mathrm{KL} \le \log(1+\chi^2) \le \chi^2$ and $p \in [\varsigma, 1-\varsigma]$. For the log-likelihood ratio,

$$\left|\log \frac{q(y)}{p(y)}\right| \le \max\left\{\log \frac{\mu_0 + \Delta}{\mu_0},\, \log \frac{1-\mu_0}{1-\mu_0-\Delta}\right\} \le \max\left\{\frac{\Delta}{\mu_0},\, \frac{\Delta}{1-\mu_0-\Delta}\right\} \le \frac{\Delta}{\varsigma} =: M,$$

using $\log(1+u) \le u$ twice. Substituting these bounds into $d^\star = \lfloor \log(1/\alpha)/(16B\kappa) \rfloor$ and applying Step 6 of part (i) gives

$$\sup_\nu \mathbb{E}_\nu[(\tau - \nu)_+] \;\ge\; \frac{d^\star}{2} \;\ge\; \frac{\varsigma(1-\varsigma)}{32} \cdot \frac{\log(1/\alpha)}{B\Delta^2} - \frac{1}{2},$$

in the joint regime $1 \le d^\star \le T/2$ and $\alpha + \sqrt{\alpha} + \alpha^\gamma \le 1/2$, which yields the claimed lower bound of order $\log(1/\alpha)/(B\Delta^2)$.

Step 2 (post-change drift and second moment of the evidence). For $t > \nu$, the audited outcome satisfies $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = \mu_0 + \Delta$. Thus, by Step 3 of the proof of Theorem 1 (with $I_t = 1$),

$$\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = \mu_0 + \Delta - \overline{\mu}_0 = \delta > 0. \tag{35}$$

For the second moment, $\mathbb{E}[Z_t^2 \mid \mathcal{G}_t, Y_t] = Y_t^2/B$ because $A_t^2 = A_t$ and $\mathbb{E}[A_t \mid \mathcal{G}_t, Y_t] = B$. Hence, $\mathbb{E}[Z_t^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}[Y_t^2 \mid \mathcal{F}_{t-1}]/B \le 1/B$ since $Y_t \in [0,1]$. Moreover,

$$\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}[Z_t^2 \mid \mathcal{F}_{t-1}] - \overline{\mu}_0\big(2\,\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] - \overline{\mu}_0\big) \;\le\; \mathbb{E}[Z_t^2 \mid \mathcal{F}_{t-1}] \;\le\; \frac{1}{B}, \tag{36}$$

where the middle inequality holds because $\overline{\mu}_0 \le \mu_0 + \Delta = \mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] \le 2\,\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}]$, which makes the subtracted term nonnegative.

Step 3 (an admissible grid point and its log-growth). Suppose the fixed-ratio geometric grid Λ has ratio $r_{\mathrm{g}} \ge 1$ and its range covers $B\delta/2$. Then Λ contains a point $\lambda_j \in [B\delta/(2r_{\mathrm{g}}),\, B\delta/2]$, i.e. $\lambda_j = c_j B\delta$ with $c_j \in [1/(2r_{\mathrm{g}}), 1/2]$. This is the precise meaning of being within a fixed factor of $B\delta$. This yields two consequences:

(a) *Admissibility.* $\lambda_j \le B\delta/2 \le 1/2 \le 1/(2\overline{\mu}_0)$ since $B \le 1$, $\delta \le 1$, and $\overline{\mu}_0 \le 1$. Moreover, $\lambda_j \overline{\mu}_0 \le 1/2$, so $u_t := \lambda_j X_t \ge -\lambda_j \overline{\mu}_0$ implies $u_t \ge -1/2$. Also, $u_t \le \lambda_j(1/B - \overline{\mu}_0) \le \lambda_j/B = c_j\delta =: \overline{u} \le 1/2$.

(b) *Growth.* By Lemma B.9, equation 35, and equation 36, for $t > \nu$,

$$\begin{aligned}\mathbb{E}[\log(1+u_t) \mid \mathcal{F}_{t-1}] &\ge \lambda_j\,\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] - \lambda_j^2\,\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] \\ &\ge \lambda_j\delta - \frac{\lambda_j^2}{B} = c_j B\delta^2 - c_j^2 B\delta^2,\end{aligned}$$

i.e. $\mathbb{E}[\log(1+u_t) \mid \mathcal{F}_{t-1}] \geq g_j := c_j(1-c_j)B\delta^2 > 0$. Since $c \mapsto c(1-c)$ is increasing on $(0, 1/2]$ and $c_j \geq 1/(2r_{\mathrm{g}})$,

$$c_j(1-c_j) \;\geq\; \frac{1}{2r_{\mathrm{g}}}\Big(1-\frac{1}{2r_{\mathrm{g}}}\Big) \;\geq\; \frac{1}{4r_{\mathrm{g}}}, \tag{37}$$

using $1 - 1/(2r_{\mathrm{g}}) \geq 1/2$ for $r_{\mathrm{g}} \geq 1$.

Step 4 (a restart copy that is entirely post-change). Fix an onset $\nu \leq T$ and let

$$r^\star := \min\{r \in \mathcal{S} :\; r \geq \nu + 1\}.$$

This set is nonempty because its largest element is $RL + 1 \geq T + 1 \geq \nu + 1$. Since consecutive restarts are L apart, $r^\star \leq \nu + L$, hence

$$0 \;\leq\; (r^\star - 1) - \nu \;\leq\; L - 1. \tag{38}$$

Every factor of copy $r^\star$ is indexed by $s \geq r^\star > \nu$, so all its increments are post-change, and Step 3 applies to each of them.

Step 5 (hitting time of the selected component). Define

$$W_t := \omega_{r^\star}\pi_j \prod_{s=r^\star}^{t}(1+\lambda_j X_s), \qquad t \geq r^\star - 1, \qquad W_{r^\star-1} = \omega_{r^\star}\pi_j.$$

By construction, $\omega_{r^\star}\mathcal{E}^{\mathrm{mix}}_{r^\star,t} \geq W_t$ because all other mixture components are nonnegative. Thus, with $b := \log(1/\alpha_{\mathrm{mon}})$ and $N := \inf\{t \geq r^\star - 1 : W_t \geq e^b\}$, we have $\tau_{\mathrm{rmix}} \leq N$. Lemma B.10 applies with $t_0 = r^\star - 1$, drift g_j, and $\overline{u} = c_j\delta$ from Step 3, yielding

$$\mathbb{E}\big[N - (r^\star - 1) \,\big|\, \mathcal{F}_{r^\star-1}\big] \;\leq\; \Xi := \frac{\log(1/\alpha_{\mathrm{mon}}) + \log(1/\pi_j) + \log(1/\omega_{r^\star}) + \log(1+c_j\delta)}{c_j(1-c_j)B\delta^2},$$

because $(b - \log W_{r^\star-1})_+ = \log(1/\alpha_{\mathrm{mon}}) + \log(1/\pi_j) + \log(1/\omega_{r^\star})$ and all three terms are nonnegative. With $\pi_j = 1/m$ and $\omega_{r^\star} = 1/(R+1)$, the numerator is $\log(1/\alpha_{\mathrm{mon}}) + \log m + \log(R+1) + \log(1+c_j\delta)$. Each design choice contributes one logarithmic term: mixing over the unknown margin contributes $\log m$, restarting over the unknown onset contributes $\log(R+1)$, and the alarm level contributes $\log(1/\alpha_{\mathrm{mon}})$.

Step 6 (assembling the delay). By equation 38 and $\tau_{\mathrm{rmix}} \leq N$,

$$(\tau_{\mathrm{rmix}} - \nu)_+ \;\leq\; \big((r^\star - 1) - \nu\big) + \big(N - (r^\star - 1)\big) \;\leq\; (L-1) + \big(N - (r^\star - 1)\big).$$

Since $r^\star - 1 \geq \nu$, we have $\mathcal{F}_\nu \subseteq \mathcal{F}_{r^\star-1}$; therefore, the tower property and Step 5 give the deterministic bound

$$\mathbb{E}_\nu\big[(\tau_{\mathrm{rmix}} - \nu)_+ \,\big|\, \mathcal{F}_\nu\big] \;\leq\; (L-1) + \Xi. \tag{39}$$

The "monitoring continues past T" hypothesis is used here: N may exceed T, and Lemma B.10 bounds a hitting time of a process that must be allowed to run.

Step 7 (choice of L and the explicit constant C). Since $L \geq 1$, we have $R = \lceil T/L \rceil \leq T$, so $\log(R+1) \leq \log(T+1) \leq \log(2T)$ for $T \geq 1$; thus, the restart cost is bounded independently of L. Therefore, $\Xi \leq \Xi_0$, where

$$\Xi_0 := \frac{\log(2T/\alpha_{\mathrm{mon}}) + \log m + \log(1+c_j\delta)}{c_j(1-c_j)B\delta^2}.$$

Choose $L = \lceil \Xi_0 \rceil$, so that $L - 1 \leq \Xi_0$. Then equation 39 gives $\mathbb{E}_\nu[(\tau_{\mathrm{rmix}} - \nu)_+ \mid \mathcal{F}_\nu] \leq 2\Xi_0$ *uniformly over onsets* $\nu \leq T$, which is the supremum in equation 9. We now make the constant explicit. Using equation 37 and $\log(1+c_j\delta) \leq \log(3/2)$, since $c_j\delta \leq 1/2$, we obtain

$$D^\infty := 2\Xi_0 \;\leq\; \frac{8r_{\mathrm{g}}\big(\log(2T/\alpha_{\mathrm{mon}}) + \log(3/2) + \log m\big)}{B\delta^2} = \frac{8r_{\mathrm{g}}\big(\log(3T/\alpha_{\mathrm{mon}}) + \log m\big)}{B\delta^2},$$

and, provided $T/\alpha_{\mathrm{mon}} \geq 3$ (which holds in any nontrivial regime, e.g. $T \geq 3$ and $\alpha_{\mathrm{mon}} \leq 1$), $\log(3T/\alpha_{\mathrm{mon}}) \leq 2\log(T/\alpha_{\mathrm{mon}})$ and hence

$$D^\infty \;\leq\; C\,\frac{\log(T/\alpha_{\mathrm{mon}}) + \log m}{B\delta^2}, \qquad C = 16\,r_{\mathrm{g}}. \tag{40}$$

The constant in equation 9 is therefore explicit: $C = 16\, r_{\mathrm{g}}$, sixteen times the grid ratio, which is the "fixed factor" by which the grid brackets $B\delta$. (For the deployed grid, $r_{\mathrm{g}} \approx 1.35$, so $C \approx 22$.)

Step 8 (matching). Part (i) with the Bernoulli constants of Step 1 gives $\Omega(\log(1/\alpha)/(B\Delta^2))$, and equation 40 gives $O((\log(T/\alpha_{\mathrm{mon}}) + \log m)/(B\delta^2))$. Whenever $\delta \geq c_0\Delta$ for a fixed $c_0 > 0$ — in particular under exact calibration $\overline{\mu}_0 = \mu_0$, where $\delta = \Delta$ — the two bounds agree in their dependence on *both* the audit budget B and the signal strength Δ, up to a constant determined by ς, c_0 and r_{g} together with the additive logarithmic grid and restart terms $\log(Tm)/(B\delta^2)$. The restarted construction thus attains the minimax rate over outcome-oblivious audit policies and all detectors that use full audited cost vectors, for regret-sufficient drift. For a general admissible $\overline{\mu}_0$ the upper bound carries the explicit factor $(\Delta/\delta)^2$, so the rate is governed by the calibrated margin δ throughout. Validity of the restarted mixture is supplied by Theorem 1, which under (A1) applies because $\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = \mu_0 \leq \overline{\mu}_0$ before the onset. □

PROOF OF PART (III)

Assume (A2): the hypotheses of part (ii) hold with the ledger finite at cap K_{det}, and D^∞ is the delay bound equation 40 for the ledger-free companion process.

Proof. Step 1 (coupling the finite-ledger and ledger-free monitors). Construct both processes on the same probability space, driven by the same audit coins (U_t) and the same outcomes (Y_t). The ledger-free process uses $A_t^\infty = \mathbf{1}\{U_t \leq B\}$; the finite-ledger process uses $A_t = I_t\mathbf{1}\{U_t \leq B\}$. As long as the ledger remains active ($I_t = 1$), the two processes have *identical* audit indicators, and hence identical $Z_t = A_tY_t/B$, identical X_t, identical grid bets (which are constants), and therefore identical restart-copy wealths. Thus, the two monitors have identical trajectories and, in particular, identical alarm times strictly before exhaustion. The coupling requires the coins to be exogenous and conditionally fresh: under (A2), the $(U_t)_{t>\nu}$ are conditionally i.i.d. $\mathrm{Unif}[0,1]$ given $\mathcal{F}_\nu$, and $p_t \equiv B$ makes both processes use the same threshold. A policy whose *marginal* rate is B but whose realized p_t differs between the two processes would not admit this coupling.

Step 2 (the good event). Let N denote the alarm time of the ledger-free restarted mixture and, for an integer $x \geq 1$, define $M_x := \sum_{t=\nu+1}^{\nu+x} A_t^\infty$. Conditionally on $\mathcal{F}_\nu$, $M_x \sim \mathrm{Bin}(x, B)$ by Step 1. We combine the tail bounds for M_x and N with a union bound. Recall that $K_\nu = K_{\mathrm{det}} - \sum_{s\leq\nu} A_s$ is the number of audits remaining *after* round ν, and write $S_t = \sum_{s\leq t} A_s$. On the event $\{N - \nu \leq x\} \cap \{M_x \leq K_\nu - 1\}$, the cumulative audit count through round $\nu + x$ satisfies

$$S_{\nu+x} \;=\; S_\nu + M_x \;\leq\; (K_{\mathrm{det}} - K_\nu) + (K_\nu - 1) \;=\; K_{\mathrm{det}} - 1,$$

so $\tau_{\mathrm{cap}} > \nu + x \geq N$. The ledger remains active throughout, the coupling from Step 1 holds through round $\nu + x$, and the finite-ledger monitor alarms at $\tau_{\mathrm{rmix}} = N < \tau_{\mathrm{cap}}$. Consequently,

$$\mathbb{P}_\nu\big(\tau_{\mathrm{rmix}} > \tau_{\mathrm{cap}} \,\big|\, \mathcal{F}_\nu\big) \;\leq\; \mathbb{P}_\nu(N - \nu > x \mid \mathcal{F}_\nu) + \mathbb{P}_\nu(M_x \geq K_\nu \mid \mathcal{F}_\nu). \tag{41}$$

Step 3 (first term: Markov). By part (ii), $\mathbb{E}_\nu[(N-\nu)_+ \mid \mathcal{F}_\nu] \leq D^\infty$, so Lemma B.11 gives $\mathbb{P}_\nu(N - \nu > x \mid \mathcal{F}_\nu) \leq D^\infty/x$. (If the threshold was already crossed by round ν the conclusion is trivial.)

Step 4 (second term: Bernstein). Take

$$x := \Big\lfloor \frac{K_\nu}{2B} \Big\rfloor,$$

which is at least 1 because $K_\nu \geq 2B$. Then $\mathbb{E}[M_x \mid \mathcal{F}_\nu] = xB \leq K_\nu/2$ and $\mathrm{Var}(M_x \mid \mathcal{F}_\nu) = xB(1-B) \leq K_\nu/2$. Apply Fact B.4 with $s := K_\nu - xB \geq K_\nu/2$ and $V := K_\nu/2$. The map $s \mapsto s^2/(2(V + s/3))$ is increasing on $(0,\infty)$ because its numerator derivative is $2s(V + s/3) - s^2/3 = 2sV + s^2/3 > 0$. We may therefore replace s by the smaller value $K_\nu/2$:

$$\begin{aligned}\mathbb{P}_\nu(M_x \geq K_\nu \mid \mathcal{F}_\nu) &\leq \exp\left(-\frac{(K_\nu/2)^2}{2\big(\frac{K_\nu}{2} + \frac{1}{3}\cdot\frac{K_\nu}{2}\big)}\right) = \exp\left(-\frac{K_\nu^2/4}{2\cdot\frac{2K_\nu}{3}}\right)\\ &= \exp\left(-\frac{K_\nu^2/4}{\frac{4K_\nu}{3}}\right) = \exp\Big(-\frac{3K_\nu}{16}\Big),\end{aligned}$$

which is the second term of equation 10; the constant $3/16$ follows from $\frac{1}{4}/\frac{4}{3}$.

Step 5 (floor bookkeeping: the constant 4). Put $a := K_\nu/(2B) \geq 1$. For $a \geq 1$, we have $\lfloor a \rfloor \geq a/2$ (if $a \in [1,2)$, then $\lfloor a \rfloor = 1 \geq a/2$; if $a \geq 2$, then $\lfloor a \rfloor > a - 1 \geq a/2$). Hence, $x \geq K_\nu/(4B)$ and

$$\frac{D^\infty}{x} \;\leq\; \frac{4BD^\infty}{K_\nu}.$$

Substituting Steps 3–5 into equation 41 and taking complements yields

$$\mathbb{P}_\nu(\tau_{\mathrm{rmix}} \leq \tau_{\mathrm{cap}} \mid \mathcal{F}_\nu) \;\geq\; 1 - \frac{4BD^\infty}{K_\nu} - \exp\Big(-\frac{3K_\nu}{16}\Big),$$

which is equation 10. For $K_\nu < 2B$, the stated bound is vacuous (the right-hand side is negative once $4BD^\infty/K_\nu > 1$), which explains the condition $K_\nu \geq 2B$.

Step 6 (the sufficient ledger size). Fix a target failure probability $\beta \in (0,1)$. If

$$K_\nu \;\geq\; \max\Big\{\frac{8BD^\infty}{\beta},\; \frac{16}{3}\log\frac{2}{\beta},\; 2B\Big\},$$

then each of the two error terms is at most $\beta/2$:

$$\frac{4BD^\infty}{K_\nu} \leq \frac{4BD^\infty\beta}{8BD^\infty} = \frac{\beta}{2}, \qquad \exp\Big(-\frac{3K_\nu}{16}\Big) \leq \exp\Big(-\frac{3}{16}\cdot\frac{16}{3}\log\frac{2}{\beta}\Big) = \exp\Big(-\log\frac{2}{\beta}\Big) = \frac{\beta}{2},$$

so $\mathbb{P}_\nu(\tau_{\mathrm{rmix}} \leq \tau_{\mathrm{cap}} \mid \mathcal{F}_\nu) \geq 1-\beta$, as asserted after the theorem. The audit *rate* B sets the achievable delay D^∞, whereas the audit *stock* K_ν determines whether that delay can be realized: it suffices for the remaining ledger to exceed $B \times$ (delay bound) by a constant factor, with additional logarithmic slack for binomial fluctuations in audit arrivals. □

Remark B.2 (domain of validity). Part (i) holds for every outcome-oblivious audit policy and every detector that uses the full audited cost vectors, uniformly over the product-form class $\mathcal{D}(\kappa, M)$ of cost-kernel alternatives and all onsets, under a finite-horizon false-alarm constraint. Part (ii) is proved for the restarted mixture in equation 6 with $\lceil T/L \rceil + 1$ copies under (A1)–(A2) and regret-sufficient drift, for which the cost-kernel constants are controlled by those of the regret kernel. The construction uses the same two components as the deployment loop, uniform auditing and mixture betting, and thus attains the rate within the method's design space. Part (iii) holds under the hypotheses of part (ii), and its coupling argument applies unchanged to any monitor for which a delay bound D is available, with D substituted for D^∞.

B.4 Proof of Theorem 3 (separation between error and regret monitoring)

Restatement. Theorem 3 (restated). *For any $\gamma > d > 0$, there exist pre- and post-change distributions P_0 and P_1 over two-action decision problems such that: (i) for every change time, the entire observation path $(d_t, |g_t|)_{t\geq 1}$ has the same law under the all-P_0 process and the changed process; (ii) with P_0 as the source law, source regret vanishes, so $R_q = r_0$ and the mean clipped regret increases by $\Delta_{01} = \min\{(\gamma - d)/R_q, 1\} > 0$ as in equation 12; and (iii) every level-α sequential test measurable with respect to $(\mathcal{H}_t^{\mathrm{acc}})$ of equation 11 has detection probability at most α. If $\Delta_{01} > \overline{\mu}_0$ and the upper-bound conditions of Theorem 2(ii) hold (uniform rate B, unbounded ledger, grid coverage, monitoring past T), the audited-regret monitor has the corresponding delay upper bound at margin $\Delta_{01} - \overline{\mu}_0$.*

Proof. Step 1 (the construction). Take the two-action feasible set $\mathcal{W} = \{e_1, e_2\} \subset \mathbb{R}^2$ and the deterministic prediction $\widehat{c} = (0, d)$ with $d > 0$. The solver returns $w^\star(\widehat{c}) \in \arg\min_{w\in\mathcal{W}} \widehat{c}^\top w = \{e_1\}$, so action 1 is deployed on every round. The *predicted cost gap* is $d_t := \widehat{c}_2 - \widehat{c}_1 = d$. Let the second coordinate be predicted exactly, $c_2 = d$, and let the signed prediction error on the first coordinate be $g := c_1 - \widehat{c}_1 = c_1$, with fixed magnitude $|g| = \gamma > d$. Define

$$P_0: \quad g = -\gamma, \;\; \text{so } c = (-\gamma,\, d); \qquad\qquad P_1: \quad g = +\gamma, \;\; \text{so } c = (\gamma,\, d).$$

Step 2 (regret before and after the change). The realized regret is $\rho = c^\top w^\star(\widehat{c}) - \min_{w\in\mathcal{W}} c^\top w = c_1 - \min\{c_1, c_2\}$.

- Under P_0: the true gap is $c_2 - c_1 = d + \gamma > 0$, so $\min\{c_1, c_2\} = c_1$ and $\rho = 0$. The deployed action is optimal despite an error of magnitude γ — the error is "harmless" because it does not cross the decision boundary.
- Under P_1: the true gap is $c_2 - c_1 = d - \gamma < 0$, so $\min\{c_1, c_2\} = c_2 = d$ and $\rho = c_1 - c_2 = \gamma - d > 0$. The same error magnitude is now "consequential".

Take P_0 as the source law. Source regret is then identically zero, so the cal-A quantile is $\widehat{r}_q = 0$ and the fallback stated after equation 1 sets $R_q = r_0 > 0$, a constant declared before any calibration data are drawn and hence fixed before cal-B and deployment. The monitored functional $Y = \min\{\rho/R_q, 1\}$ of equation 1 equals 0 under P_0 and $\min\{(\gamma - d)/R_q, 1\} = \Delta_{01}$ under P_1. Hence the mean clipped regret increases by exactly $\Delta_{01} > 0$, which is (ii) and equation 12.

Step 3 (the observed path is unchanged). The scalar error-magnitude monitor observes only $\mathcal{H}_t^{\mathrm{acc}} = \sigma\{d_s, |g_s| : s \le t\}$. Under both P_0 and P_1, the observation is the deterministic pair $(d_s, |g_s|) = (d, \gamma)$ at every round. Therefore, for *every* change time ν, the process that emits P_0 for all t and the process that switches to P_1 at ν induce the same degenerate law on $\mathcal{H}_\infty^{\mathrm{acc}} = \sigma\big(\bigcup_t \mathcal{H}_t^{\mathrm{acc}}\big)$. This establishes (i). Only the *sign* of g changes, and this sign is, by definition, invisible to the filtration.

Step 4 (blindness of every level-α test). Let T be the rejection time of any sequential test measurable with respect to $(\mathcal{H}_t^{\mathrm{acc}})$, with $\mathbb{P}_0(T < \infty) \le \alpha$. Randomized and adaptive tests are also covered. For a randomized test, augment the filtration to $\mathcal{H}_t^{\mathrm{acc}} \vee \sigma(V_1, \dots, V_t)$ with auxiliary variables V_s whose law is fixed and identical under both hypotheses and that are independent of the data; the augmented path laws remain equal because they are products of equal factors. Since $\{T \le t\} \in \mathcal{H}_t^{\mathrm{acc}}$ (augmented) for each t, and the two laws agree on each $\mathcal{H}_t^{\mathrm{acc}}$ by Step 3,

$$\mathbb{P}_{0\to1}(T < \infty) = \lim_{t\to\infty} \mathbb{P}_{0\to1}(T \le t) = \lim_{t\to\infty} \mathbb{P}_0(T \le t) = \mathbb{P}_0(T < \infty) \;\le\; \alpha,$$

where the limits exist by continuity from below along the increasing events $\{T \le t\}$. Thus, the detection probability is at most the false-alarm level, so the test has no power. This proves (iii).

Step 5 (what the regret monitor does instead). Under P_1 the audited clipped regret satisfies $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = \Delta_{01}$ deterministically. Assume $\delta := \Delta_{01} - \overline{\mu}_0 > 0$, uniform auditing at rate B with an unbounded ledger, and a bet grid covering $B\delta/2$. Then Step 2 of the proof of Theorem 2(ii) applies verbatim with Δ replaced by Δ_{01}: the drift is $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = \Delta_{01} - \overline{\mu}_0 = \delta > 0$; the second moment obeys $\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] \le \mathbb{E}[Z_t^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}[Y_t^2]/B \le 1/B$, the first inequality using $\overline{\mu}_0 < \Delta_{01} = \mathbb{E}[Z_t \mid \mathcal{F}_{t-1}]$. Steps 3–7 of that proof then give the delay bound equation 9 at margin $\delta = \Delta_{01} - \overline{\mu}_0$.

Scope of this last claim. The post-change outcome here is the *constant* Δ_{01}, not a Bernoulli variable with mean in $[\varsigma, 1 - \varsigma]$; the Bernoulli-kernel hypothesis of Theorem 2(ii) is used only to compute the class constants κ and M for the *lower* bound, and the upper bound of that theorem needs only the drift and second-moment conditions verified above. The margin $\delta > 0$ is essential and is not implied by $\Delta_{01} > 0$: with a conservative $\overline{\mu}_0 \ge \Delta_{01}$, every admissible nonnegative bet gives factors of conditional mean at most one and the regret monitor need not detect either — while the blindness of the $\mathcal{H}^{\mathrm{acc}}$ class persists regardless. □

Remark B.3 (the monitor class covered). The theorem covers monitors of the *scalar error magnitude*, which is the class deployed dashboards belong to: CUSUM and windowed-mean procedures driven by $|g_t|$ together with the predicted gap d_t are all measurable with respect to $\mathcal{H}_t^{\mathrm{acc}}$, and the blindness is uniform over that class rather than specific to one statistic. A monitor with access to the full *signed* residual vector lies in a different class: it can reconstruct $c_t = \widehat{c}_t - (\widehat{c}_t - c_t)$, hence the regret, and is therefore an audited-regret monitor by another name.

B.5 Proof of Proposition 1 (oracle allocation and exact excess) and of the transfer decomposition

Restatement. Proposition 1 (restated). *Assume $0 < B \le 1$, $\sigma_Q > 0$ almost surely, $0 < \mathbb{E}_Q[\sigma_Q] < \infty$, and $s > 0$ with $0 < \mathbb{E}_Q[s] < \infty$ and $\mathbb{E}_Q[\sigma_Q^2/s] < \infty$. Suppose $p_s \le 1$ and $p_{\sigma_Q} \le 1$ almost surely. Among feasible proportional allocations, $V(p; Q)$ is minimized by $p^\star$ in*

equation 14, with $V^\star(Q) = \mathbb{E}_Q[\sigma_Q]^2/B$. For any feasible score s, write $\sigma_Q(x) = s(x)(1+\varepsilon(x))$ and $d\mu_s = s\,dQ/\mathbb{E}_Q[s]$. Then the exact identities in equation 15 are

$$V(p_s;Q) - V^\star(Q) = \frac{\mathbb{E}_Q[s]^2}{B}\,\mathrm{Var}_{\mu_s}(\varepsilon), \qquad \frac{V(p_{\mathrm{unif}};Q)}{V^\star(Q)} = 1 + \mathrm{CV}_Q(\sigma_Q)^2.$$

If the proportional oracle is infeasible, the constrained optimum is the capped rule $p^\star(x) = \min\{1, c\sigma_Q(x)\}$, where the constant c is chosen to meet the budget.

Proof. Step 1 (V depends on the outcome law only through σ_Q). By the tower property and the identity $\sigma_Q(x)^2 = \mathbb{E}_Q[Y^2 \mid X = x]$ from equation 13,

$$V(p;Q) = \mathbb{E}_Q\Big[\frac{Y^2}{p(X)}\Big] = \mathbb{E}_Q\Big[\frac{\mathbb{E}_Q[Y^2 \mid X]}{p(X)}\Big] = \mathbb{E}_Q\Big[\frac{\sigma_Q(X)^2}{p(X)}\Big], \tag{42}$$

using that p is a measurable function of X alone. Recall from the proof of Theorem 1, Step 2, that V is the conditional second moment of the Horvitz–Thompson term: on an active round with $p_t = p(x_t)$, $\mathbb{E}[Z_t^2 \mid \mathcal{G}_t, Y_t] = Y_t^2/p(x_t)$, so $\mathbb{E}[Z_t^2] = V(p;Q)$. This is why V, and not some other functional, is the relevant design objective: it is the quantity that enters the quadratic term of the log-growth bound in Lemma B.9.

Step 2 (value of a proportional allocation). For $p_s(x) = Bs(x)/\mathbb{E}_Q[s]$, equation 42 gives

$$V(p_s;Q) = \frac{\mathbb{E}_Q[s]}{B}\,\mathbb{E}_Q\Big[\frac{\sigma_Q^2}{s}\Big] \;<\; \infty, \tag{43}$$

which is finite under the assumptions $\mathbb{E}_Q[s] < \infty$ and $\mathbb{E}_Q[\sigma_Q^2/s] < \infty$. Taking $s = \sigma_Q$ (feasible by hypothesis) gives $p^\star = B\sigma_Q/\mathbb{E}_Q[\sigma_Q]$ and

$$V(p^\star;Q) = \frac{\mathbb{E}_Q[\sigma_Q]}{B}\,\mathbb{E}_Q\Big[\frac{\sigma_Q^2}{\sigma_Q}\Big] = \frac{\mathbb{E}_Q[\sigma_Q]^2}{B} = V^\star(Q),$$

which is the second display in equation 14.

Step 3 (the exact excess identity). Substituting $\sigma_Q = s(1+\varepsilon)$ into equation 43, we have $\sigma_Q^2/s = s(1+\varepsilon)^2$ and $\mathbb{E}_Q[\sigma_Q] = \mathbb{E}_Q[s(1+\varepsilon)]$, so

$$B\big(V(p_s;Q) - V^\star(Q)\big) = \mathbb{E}_Q[s]\,\mathbb{E}_Q[s(1+\varepsilon)^2] - \mathbb{E}_Q[s(1+\varepsilon)]^2.$$

Expanding both terms and writing $\mathbb{E}$ for $\mathbb{E}_Q$ throughout,

$$\mathbb{E}[s]\,\mathbb{E}[s(1+\varepsilon)^2] = \mathbb{E}[s]\big(\mathbb{E}[s] + 2\,\mathbb{E}[s\varepsilon] + \mathbb{E}[s\varepsilon^2]\big) = \mathbb{E}[s]^2 + 2\,\mathbb{E}[s]\,\mathbb{E}[s\varepsilon] + \mathbb{E}[s]\,\mathbb{E}[s\varepsilon^2],$$
$$\mathbb{E}[s(1+\varepsilon)]^2 = \big(\mathbb{E}[s] + \mathbb{E}[s\varepsilon]\big)^2 = \mathbb{E}[s]^2 + 2\,\mathbb{E}[s]\,\mathbb{E}[s\varepsilon] + \mathbb{E}[s\varepsilon]^2.$$

The first two terms cancel exactly, leaving

$$B\big(V(p_s;Q) - V^\star(Q)\big) = \mathbb{E}[s]\,\mathbb{E}[s\varepsilon^2] - \mathbb{E}[s\varepsilon]^2. \tag{44}$$

We now change the measure. Since $d\mu_s = s\,dQ/\mathbb{E}_Q[s]$ is a probability measure ($s > 0$, $0 < \mathbb{E}_Q[s] < \infty$), for any ψ satisfying $\mathbb{E}_Q[s|\psi|] < \infty$, we have $\mathbb{E}_{\mu_s}[\psi] = \mathbb{E}_Q[s\psi]/\mathbb{E}_Q[s]$. Applying this identity with $\psi = \varepsilon$ and $\psi = \varepsilon^2$ in equation 44 gives

$$B\big(V(p_s;Q) - V^\star(Q)\big) = \mathbb{E}[s]^2\,\mathbb{E}_{\mu_s}[\varepsilon^2] - \mathbb{E}[s]^2\,\mathbb{E}_{\mu_s}[\varepsilon]^2 = \mathbb{E}_Q[s]^2\,\mathrm{Var}_{\mu_s}(\varepsilon),$$

which is the first identity in equation 15. The required integrability follows from $\mathbb{E}_Q[s\varepsilon^2] \le 2\mathbb{E}_Q[\sigma_Q^2/s] + 2\mathbb{E}_Q[s] < \infty$ because $s\varepsilon^2 = (\sigma_Q - s)^2/s \le 2\sigma_Q^2/s + 2s$.

Step 4 (optimality within the feasible proportional family). Since $\mathrm{Var}_{\mu_s}(\varepsilon) \ge 0$, $V(p_s;Q) \ge V^\star(Q)$ for every feasible score s, and the proportional oracle is therefore the minimizer. Equality holds if and only if ε is μ_s-almost surely constant. Since $s > 0$, μ_s and Q are mutually absolutely continuous, so this condition is equivalent to σ_Q/s being Q-almost surely constant, i.e. $s \propto \sigma_Q$, in which case $p_s = p^\star$. The quantity $\mathrm{Var}_{\mu_s}(\varepsilon)$ is measured under the *score-tilted* law μ_s, not under Q; thus, misallocation is penalized in proportion to the audit mass that the score assigns to mis-scored contexts.

Step 5 (the uniform ratio). Take $s \equiv 1$, so $p_{\mathrm{unif}} \equiv B$, feasible since $B \le 1$. Then equation 43 gives $V(p_{\mathrm{unif}}; Q) = \mathbb{E}_Q[\sigma_Q^2]/B$ and

$$\frac{V(p_{\mathrm{unif}}; Q)}{V^\star(Q)} = \frac{\mathbb{E}_Q[\sigma_Q^2]/B}{\mathbb{E}_Q[\sigma_Q]^2/B} = \frac{\mathbb{E}_Q[\sigma_Q^2]}{\mathbb{E}_Q[\sigma_Q]^2} = \frac{\mathbb{E}_Q[\sigma_Q]^2 + \mathrm{Var}_Q(\sigma_Q)}{\mathbb{E}_Q[\sigma_Q]^2} = 1 + \mathrm{CV}_Q(\sigma_Q)^2,$$

the second identity in equation 15. Equivalently, via equation 44 with $s \equiv 1$: $\mu_1 = Q$, $\varepsilon = \sigma_Q - 1$, and $\mathrm{Var}_Q(\varepsilon) = \mathrm{Var}_Q(\sigma_Q)$. The clairvoyant headroom of any score-guided auditing over uniform is therefore exactly $\mathrm{CV}_Q(\sigma_Q)^2$ — an estimable, per-problem quantity.

Step 6 (the capped rule when the proportional oracle is infeasible). Feasibility is not automatic: if $B\sigma_Q/\mathbb{E}_Q[\sigma_Q] > 1$ on a set of positive measure, then $p^\star$ is not a probability. Consider the constrained problem

$$\text{minimize} \quad V(p; Q) = \mathbb{E}_Q[\sigma_Q^2/p] \qquad \text{over} \quad \{p : \mathcal{X} \to (0, 1] \text{ measurable}, \ \mathbb{E}_Q[p] \le B\}.$$

Define $\Phi(c) := \mathbb{E}_Q[\min\{1, c\sigma_Q\}]$ for $c > 0$. The function Φ is nondecreasing and continuous by dominated convergence, with dominating function 1, and satisfies $\Phi(0^+) = 0$ and $\Phi(\infty) = 1 \ge B$. Hence, there exists $c > 0$ such that $\Phi(c) = B$ (if $B = 1$, adopt the convention $p \equiv 1$). Set $p_{\mathrm{cap}} := \min\{1, c\sigma_Q\}$, which is the rule denoted by $p^\star$ in the statement, and $\theta := 1/c^2$.

Fix x and consider $L(u) := \sigma_Q(x)^2/u + \theta u$ on $(0, 1]$. The function L is strictly convex ($L''(u) = 2\sigma_Q(x)^2 u^{-3} > 0$), with unconstrained stationary point $L'(u) = 0$ at $u = \sigma_Q(x)/\sqrt{\theta} = c\,\sigma_Q(x)$. Therefore, the minimizer of L over $(0, 1]$ is $\min\{1, c\sigma_Q(x)\} = p_{\mathrm{cap}}(x)$, and $L(p(x)) \ge L(p_{\mathrm{cap}}(x))$ pointwise for every admissible p. Taking $\mathbb{E}_Q$ gives

$$V(p; Q) + \theta\,\mathbb{E}_Q[p] \ \ge \ V(p_{\mathrm{cap}}; Q) + \theta\,\mathbb{E}_Q[p_{\mathrm{cap}}] = V(p_{\mathrm{cap}}; Q) + \theta B,$$

and since $\mathbb{E}_Q[p] \le B$, rearranging the budget terms yields

$$V(p; Q) \ \ge \ V(p_{\mathrm{cap}}; Q) + \theta\big(B - \mathbb{E}_Q[p]\big) \ \ge \ V(p_{\mathrm{cap}}; Q).$$

Thus, the capped rule is optimal over *all* budgeted samplers with known Q, not merely within the proportional family. As a consistency check, when the proportional oracle is feasible, $c = B/\mathbb{E}_Q[\sigma_Q]$ satisfies $\Phi(c) = \mathbb{E}_Q[c\sigma_Q] = B$ and $p_{\mathrm{cap}} = p^\star$, so the capped rule reduces to equation 14. When it is infeasible, the capped optimum is strictly worse than $V^\star(Q)$ because $V^\star(Q)$ is the value of an infeasible relaxation. □

Proof of the transfer decomposition equation 16.

Proof. Let $\widehat{\sigma} > 0$ be a source-fitted score satisfying the score assumptions of Proposition 1, and let $d\mu_{\widehat{\sigma}} = \widehat{\sigma}\, dQ/\mathbb{E}_Q[\widehat{\sigma}]$ denote the associated tilted probability measure.

Step 1 (pointwise split). By definition $\varepsilon = (\sigma_Q - \widehat{\sigma})/\widehat{\sigma}$. Adding and subtracting the source volatility profile σ_{src} in the numerator,

$$\varepsilon = \frac{\sigma_Q - \widehat{\sigma}}{\widehat{\sigma}} = \underbrace{\frac{\sigma_{\mathrm{src}} - \widehat{\sigma}}{\widehat{\sigma}}}_{=:a \text{ (estimation)}} + \underbrace{\frac{\sigma_Q - \sigma_{\mathrm{src}}}{\widehat{\sigma}}}_{=:b \text{ (transfer)}} = a + b,$$

an identity holding pointwise on $\{\widehat{\sigma} > 0\}$, hence Q-almost surely and $\mu_{\widehat{\sigma}}$-almost surely. This is the first line of equation 16. The split is meaningful because a depends only on source quantities (it is what a better estimator could reduce) while b is a property of the drift (no amount of source data reduces it).

Step 2 (variance expansion). Assume $a, b \in L^2(\mu_{\widehat{\sigma}})$. Bilinearity of the covariance form on $L^2(\mu_{\widehat{\sigma}})$ gives

$$\mathrm{Var}_{\mu_{\widehat{\sigma}}}(\varepsilon) = \mathrm{Var}_{\mu_{\widehat{\sigma}}}(a) + \mathrm{Var}_{\mu_{\widehat{\sigma}}}(b) + 2\,\mathrm{Cov}_{\mu_{\widehat{\sigma}}}(a, b),$$

which is the third line of equation 16. By the Cauchy–Schwarz inequality and the arithmetic–geometric mean inequality,

$$2\,\mathrm{Cov}_{\mu_{\widehat{\sigma}}}(a, b) \le 2\sqrt{\mathrm{Var}_{\mu_{\widehat{\sigma}}}(a)\,\mathrm{Var}_{\mu_{\widehat{\sigma}}}(b)} \le \mathrm{Var}_{\mu_{\widehat{\sigma}}}(a) + \mathrm{Var}_{\mu_{\widehat{\sigma}}}(b),$$

and therefore $\mathrm{Var}_{\mu_{\widehat{\sigma}}}(\varepsilon) \le 2\,\mathrm{Var}_{\mu_{\widehat{\sigma}}}(a) + 2\,\mathrm{Var}_{\mu_{\widehat{\sigma}}}(b)$, which is the last line of equation 16.

Step 3 (the exact-estimation case). Suppose $\widehat{\sigma} = \sigma_{\mathrm{src}}$, so $a \equiv 0$ and $\mu_{\widehat{\sigma}} = \mu_{\sigma_{\mathrm{src}}}$. Then

$$\mathrm{Var}_{\mu_{\sigma_{\mathrm{src}}}}(\varepsilon) = \mathrm{Var}_{\mu_{\sigma_{\mathrm{src}}}}(b) = \mathrm{Var}_{\mu_{\sigma_{\mathrm{src}}}}\big(\sigma_Q/\sigma_{\mathrm{src}} - 1\big) = \mathrm{Var}_{\mu_{\sigma_{\mathrm{src}}}}\big(\sigma_Q/\sigma_{\mathrm{src}}\big),$$

which vanishes if and only if the ratio $\sigma_Q/\sigma_{\mathrm{src}}$ is almost surely constant. Combining with equation 15: the allocation excess is zero exactly under a *uniform rescaling* of the volatility profile and strictly positive under any context-dependent *reshaping*. This is the precise sense in which even a perfectly estimated source score fails to transfer.

Cancellation between a and b can allow estimation bias to offset transfer error. Both equation 15 and equation 16 are exact identities for the *unclipped* proportional family; the deployed sampler is clipped and is instead covered by Theorem 4, which imposes no proportionality. □

B.6 Proof of Theorem 4 (uniform auditing is minimax under reshaping)

Restatement. Theorem 4 (restated). *Let $(\mathcal{X}, \mu)$ be finite with equal weights or nonatomic standard, fix $\overline{\sigma} : \mathcal{X} \to [0,1]$ with $\mathbb{E}_\mu[\overline{\sigma}^2] > 0$, and let $\mathcal{Q}(\overline{\sigma})$ be the deployment laws with feature marginal μ for which σ_Q has the same distribution as $\overline{\sigma}$. Then for every $p \in \mathcal{P}(B)$ of equation 17, equation 18 holds:*

$$\sup_{Q \in \mathcal{Q}(\overline{\sigma})} V(p;Q) \;\ge\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p] \;\ge\; \frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B} \;=\; V(p_{\mathrm{unif}};Q), \qquad Q \in \mathcal{Q}(\overline{\sigma}).$$

Equality throughout holds if and only if $p(x) = B$ for μ-almost every x; uniform auditing is the unique minimax sampler in this drift class; and every nonuniform sampler has worst-case excess at least equation 19, namely $\big(B\mathbb{E}_\mu[1/p] - 1\big)\mathbb{E}_\mu[\overline{\sigma}^2]/B$.

Proof. Throughout, equation 42 and the fact that every $Q \in \mathcal{Q}(\overline{\sigma})$ has feature marginal μ imply that

$$V(p;Q) = \mathbb{E}_\mu\big[\sigma_Q^2/p\big], \qquad 0 \le \sigma_Q \le 1. \tag{45}$$

The class $\mathcal{P}(B)$ imposes only measurability, $0 < p \le 1$, and the population budget $\mathbb{E}_\mu[p] \le B$; it requires neither proportionality nor an unclipped structure. Note also that $\mathcal{Q}(\overline{\sigma})$ is nonempty: the Bernoulli construction in Step 4 below provides a member for every admissible $\overline{\sigma}$.

Step 1 (the uniform sampler is constant across the entire class). Take $p_{\mathrm{unif}} \equiv B$, which lies in $\mathcal{P}(B)$ because $0 < B \le 1$. For every $Q \in \mathcal{Q}(\overline{\sigma})$, σ_Q and $\overline{\sigma}$ are equal in distribution under μ and therefore have equal second moments. Thus, by equation 45,

$$V(p_{\mathrm{unif}};Q) = \frac{\mathbb{E}_\mu[\sigma_Q^2]}{B} = \frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B}. \tag{46}$$

The uniform sampler is thus *indifferent* to reshaping: the adversary's rearrangement of where volatility sits cannot change its value. This is the structural reason it turns out to be minimax.

Step 2 (the degenerate branch $\mathbb{E}_\mu[1/p] = \infty$). This case must be handled separately because the rearrangement argument in Step 3 invokes Ryff's theorem for an L^1 function. For $n \ge 1$, set

$$f_n := \min\{1/p,\, n\},$$

which is measurable, bounded by n, and satisfies $f_n \ge 1$ since $p \le 1$. Because it is bounded on a probability space, $f_n \in L^1_+(\mu)$. Applying the construction in Steps 3–4 below with f_n in place of $1/p$ produces a profile $\sigma_{*,n}$ equimeasurable with $\overline{\sigma}$ and a law $Q_n \in \mathcal{Q}(\overline{\sigma})$ such that $\mathbb{E}_\mu[\sigma_{*,n}^2 f_n] \ge \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[f_n]$. Since $1/p \ge f_n$ pointwise,

$$V(p;Q_n) = \mathbb{E}_\mu[\sigma_{*,n}^2/p] \;\ge\; \mathbb{E}_\mu[\sigma_{*,n}^2 f_n] \;\ge\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[f_n].$$

By monotone convergence, $\mathbb{E}_\mu[f_n] \uparrow \mathbb{E}_\mu[1/p] = \infty$, so $\sup_Q V(p;Q) = +\infty$, and both inequalities in equation 18 hold immediately: the first as $\infty \ge \infty$ and the second as $\infty \ge \mathbb{E}_\mu[\overline{\sigma}^2]/B$, which is finite because $\overline{\sigma} \le 1$. In this branch, p is not μ-a.e. equal to B, consistent with the equality characterization. Henceforth, assume $1/p \in L^1(\mu)$.

Step 3 (rearrangement lower bound). We establish

$$\exists\, \sigma_* \text{ equimeasurable with } \overline{\sigma} : \qquad \mathbb{E}_\mu\big[\sigma_*^2 \cdot (1/p)\big] \;\geq\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p]. \tag{47}$$

The adversary places the largest volatilities exactly where the audit probabilities are lowest.

(a) Finite $\mathcal{X}$ with equal weights. Order the n points so that $1/p$ is nonincreasing, and let σ_*^2 assign the n values of $\overline{\sigma}^2$ in nonincreasing order according to the same ordering. Then $\sigma_* := \sqrt{\sigma_*^2}$ takes exactly the values of $\overline{\sigma}$ with the same multiplicities and is therefore equimeasurable with $\overline{\sigma}$ under the uniform μ. Moreover, the two sequences (σ_*^2) and $(1/p)$ are similarly ordered. Fact B.7(a) gives equation 47.

(b) Nonatomic standard $(\mathcal{X}, \mu)$. Identify $(\mathcal{X}, \mu)$ with $((0,1), \mathrm{Leb})$. Since $p > 0$ everywhere and $1/p \in L^1_+(0,1)$, Fact B.6 provides a measure-preserving map $\sigma_0 : (0,1) \to (0,1)$ such that $1/p = h \circ \sigma_0$ almost everywhere, where $h := (1/p)^*$ is the nonincreasing rearrangement of $1/p$. Let q be the nonincreasing rearrangement of $\overline{\sigma}^2$ and define

$$\sigma_* := \sqrt{q \circ \sigma_0}.$$

Then σ_* is measurable and $[0,1]$-valued because $0 \leq q \leq 1$, and it has the same distribution as $\overline{\sigma}$. Indeed, σ_0 is measure-preserving, so $q \circ \sigma_0$ has the same law as q, which by construction has the same law as $\overline{\sigma}^2$; taking square roots preserves equality in distribution. Using measure preservation once more to change variables,

$$\begin{aligned}\mathbb{E}_\mu\big[\sigma_*^2 \cdot (1/p)\big] &= \int_0^1 q(\sigma_0(v))\, h(\sigma_0(v))\, dv = \int_0^1 q(u) h(u)\, du \\ &\geq \Big(\int_0^1 q\Big)\Big(\int_0^1 h\Big) = \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p],\end{aligned}$$

where the inequality follows from Fact B.7(b), applied to the two nonincreasing nonnegative functions q and h. The integrability requirement $qh \in L^1$ holds because $0 \leq q \leq 1$ and $h \in L^1(0,1)$. The last equality follows because a nonincreasing rearrangement preserves the integral.

Step 4 (realizability: the rearranged profile is achieved by an admissible deployment law). Given σ_* from Step 3, define Q_* by taking the feature marginal μ and the conditional kernel $Y \mid X = x \;\sim\; \mathrm{Bern}\big(\sigma_*(x)^2\big)$, which is a valid probability because $\sigma_*(x)^2 \in [0,1]$. Then $Y \in [0,1]$ as required of a monitored functional, and since $Y^2 = Y$ for a $\{0,1\}$-valued variable, $\mathbb{E}_{Q_*}[Y^2 \mid X = x] = \sigma_*(x)^2$, i.e. the volatility profile of Q_* is exactly σ_*. Being equimeasurable with $\overline{\sigma}$, this puts $Q_* \in \mathcal{Q}(\overline{\sigma})$. Therefore, by equation 45 and equation 47,

$$\sup_{Q \in \mathcal{Q}(\overline{\sigma})} V(p; Q) \;\geq\; V(p; Q_*) = \mathbb{E}_\mu[\sigma_*^2/p] \;\geq\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p], \tag{48}$$

the first inequality in equation 18.

Step 5 (the budget step). The map $u \mapsto 1/u$ is strictly convex on $(0,1]$. Applying Fact B.8 and then using $\mathbb{E}_\mu[p] \leq B$ yields

$$\mathbb{E}_\mu[1/p] \;\geq\; \frac{1}{\mathbb{E}_\mu[p]} \;\geq\; \frac{1}{B}. \tag{49}$$

Chaining equation 48, equation 49, and equation 46 yields the full display in equation 18:

$$\sup_Q V(p; Q) \;\geq\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p] \;\geq\; \frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B} \;=\; V(p_{\mathrm{unif}}; Q) \quad \text{for every } Q \in \mathcal{Q}(\overline{\sigma}).$$

Step 6 (equality characterization). Suppose all three quantities in equation 18 coincide. Then in particular $\mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p] = \mathbb{E}_\mu[\overline{\sigma}^2]/B$, and since $\mathbb{E}_\mu[\overline{\sigma}^2] > 0$ we may divide to get $\mathbb{E}_\mu[1/p] = 1/B$. This forces equality at both steps of equation 49. Equality in the Jensen step forces p to be μ-almost surely constant (Fact B.8, strict convexity); equality in $1/\mathbb{E}_\mu[p] \geq 1/B$ forces $\mathbb{E}_\mu[p] = B$, so that constant is B. Conversely, if $p \equiv B$ μ-a.e. then $\mathbb{E}_\mu[1/p] = 1/B$, the rearrangement bound equation 47 becomes an identity (both sides equal $\mathbb{E}_\mu[\overline{\sigma}^2]/B$, since $1/p$ is constant and equimeasurable profiles have equal second moments), and by Step 1 the supremum equals $\mathbb{E}_\mu[\overline{\sigma}^2]/B$ exactly; so all three coincide. Uniform auditing is therefore the *unique* minimax sampler in $\mathcal{P}(B)$ against this reshaping adversary.

Step 7 (the overconfidence penalty). For any $p \in \mathcal{P}(B)$, subtracting equation 46 from equation 48,

$$\sup_{Q \in \mathcal{Q}(\overline{\sigma})} V(p;Q) - V(p_{\mathrm{unif}};Q) \;\geq\; \mathbb{E}_\mu[\overline{\sigma}^2]\,\mathbb{E}_\mu[1/p] - \frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B} \;=\; \big(B\,\mathbb{E}_\mu[1/p] - 1\big)\frac{\mathbb{E}_\mu[\overline{\sigma}^2]}{B},$$

which is equation 19; it is strictly positive whenever $p \not\equiv B$, by Step 6 and $\mathbb{E}_\mu[\overline{\sigma}^2] > 0$. The dimensionless factor $B\,\mathbb{E}_\mu[1/p] - 1$ is outcome-free: it depends on the sampler and the feature law only, never on the regret distribution. □

Remark B.4 (the sampler class covered, and condition (A3)). Three points fix the reach of the theorem. (a) It concerns the Horvitz–Thompson second moment V, the functional that drives the monitor's growth, over *static* samplers, i.e. measurable functions of x; this is the same policy class as in Theorem 2(i). (b) Because $\mathcal{P}(B)$ imposes no proportionality, no unclipped structure and no feasibility condition beyond $p \leq 1$, the optimality statement covers the clipped, $\widehat{\eta}$-calibrated samplers of the deployed form in equation 2 — a strictly larger class than the unclipped proportional family over which the exact identity equation 15 is stated. (c) Membership in $\mathcal{P}(B)$ is the population budget $\mathbb{E}_\mu[\mathrm{clip}(\widehat{\eta}s(X), p_{\min}, p_{\max})] \leq B$, which is the quantity the calibration of $\widehat{\eta}$ targets; we carry it as assumption (A3), so that a sampler enters the comparison only when it spends at most the budget it was granted. The penalty $B\,\mathbb{E}_\mu[1/p] - 1$ is dimensionless and is evaluated from calibration features by plug-in.

B.7 Proof of Theorem 5 (per-candidate and lifetime promotion control)

Restatement. Theorem 5 (restated). *Let $\mathcal{F}^{\mathrm{sh}}_{i-1}$ denote the selected-sequence history before the ith audited disagreement, including the fixed candidate and its training randomness. (i) Per candidate: under the conditional null $H^{\mathrm{sh}}_{0,k}$: $\mathbb{E}[D_i \mid \mathcal{F}^{\mathrm{sh}}_{i-1}] \leq 0$ for every i, the probability of promoting candidate k is at most α_k. (ii) Deployment lifetime: suppose candidates are generated sequentially and may depend on all data collected before their shadow phases. If the levels follow a predeclared spending rule $\sum_{k\geq 1}\alpha_k \leq \alpha_{\mathrm{dep}}$, then the probability of ever promoting any candidate satisfying its null $H^{\mathrm{sh}}_{0,k}$ is at most α_{dep}. The expected number of such promotions is also at most α_{dep}.*

Proof of part (i)

Proof. Step 1 (the selected sequence and its filtration). Let $t_1 < t_2 < \cdots$ enumerate the shadow rounds on which the two models disagree, i.e. $S_t = 1$ in equation 20, and which are therefore fully audited. Let $\mathcal{G}^{\mathrm{sh}}_t$ be the shadow information available at round t before the cost c_t is revealed; it contains x_t and both predictions $\widehat{c}_t, \widehat{c}'_t$, hence contains S_t. Since $\{t_i = t\}$ is determined by $S_1, \ldots, S_t$, each t_i is a stopping time of $(\mathcal{G}^{\mathrm{sh}}_t)$. Put $\mathcal{F}^{\mathrm{sh}}_{i-1} := \mathcal{G}^{\mathrm{sh}}_{t_i}$, the pre-outcome field at the ith selected round. It contains the previous selected differences $D_1, \ldots, D_{i-1}$, the candidate f' and its training randomness (all fixed before shadowing begins, hence in $\mathcal{F}^{\mathrm{sh}}_0$), and the current round's features and predictions — but not c_{t_i}, hence not D_i. The bet λ^{sh}_i is required to be $\mathcal{F}^{\mathrm{sh}}_{i-1}$-measurable, i.e. measurable before D_i is observed, exactly as in equation 21. If only finitely many disagreements occur, set $D_i := 0$ for the missing indices, which leaves every factor equal to 1.

Step 2 (factors are nonnegative). $D_i = Y_i(f) - Y_i(f') \in [-1, 1]$ because both clipped regrets lie in $[0, 1]$ by equation 1. With $0 \leq \lambda^{\mathrm{sh}}_i \leq 1$,

$$1 + \lambda^{\mathrm{sh}}_i D_i \;\geq\; 1 - \lambda^{\mathrm{sh}}_i \;\geq\; 0.$$

The sign convention matters: a positive D_i favors the candidate, so a nonnegative bet accumulates wealth only when the evidence favors the candidate.

Step 3 (conditional mean at most one). Under $H^{\mathrm{sh}}_{0,k}$, the $\mathcal{F}^{\mathrm{sh}}_{i-1}$-measurability and boundedness of λ^{sh}_i imply

$$\mathbb{E}\big[1 + \lambda^{\mathrm{sh}}_i D_i \,\big|\, \mathcal{F}^{\mathrm{sh}}_{i-1}\big] = 1 + \lambda^{\mathrm{sh}}_i\,\mathbb{E}\big[D_i \mid \mathcal{F}^{\mathrm{sh}}_{i-1}\big] \leq 1.$$

Step 4 (supermartingale and Ville). Hence $\mathcal{G}_m = \prod_{i\leq m}(1 + \lambda^{\mathrm{sh}}_i D_i)$ of equation 21 is a nonnegative supermartingale with respect to $(\mathcal{F}^{\mathrm{sh}}_m)$ started at $\mathcal{G}_0 = 1$, by the same product computation as in Step

5 of the proof of Theorem 1. The promotion rule for candidate k is $\mathcal{G}_m \geq 1/\alpha_k$ for some m before the shadow ledger is exhausted, so by Fact B.1

$$\mathbb{P}(\text{promote } k) \;\leq\; \mathbb{P}\Big(\sup_{m\geq 0} \mathcal{G}_m \geq 1/\alpha_k\Big) \;\leq\; \alpha_k.$$

Stopping at the shadow budget only shrinks the crossing event, so the bound is unaffected by the finite K_{shadow}.

Step 5 (why the raw-time null suffices, and why a marginal one does not). Define the raw-time difference $D_t^{\text{raw}} = S_t(Y_t(f) - Y_t(f'))$, which is 0 on agreement rounds. Suppose the raw-time conditional null $\mathbb{E}[D_t^{\text{raw}} \mid \mathcal{G}_t^{\text{sh}}] \leq 0$ holds. Because S_t is $\mathcal{G}_t^{\text{sh}}$-measurable, as established by equation 20 because S_t is computed from x_t and the two predictions before any cost is revealed, it factors out:

$$\mathbb{E}\big[D_t^{\text{raw}} \mid \mathcal{G}_t^{\text{sh}}\big] = S_t\, \mathbb{E}\big[Y_t(f) - Y_t(f') \mid \mathcal{G}_t^{\text{sh}}\big],$$

so on $\{S_t = 1\}$, $\mathbb{E}[Y_t(f) - Y_t(f') \mid \mathcal{G}_t^{\text{sh}}] \leq 0$. Evaluating this relation at the stopping time $t = t_i$ and recalling that $\mathcal{F}_{i-1}^{\text{sh}} = \mathcal{G}_{t_i}^{\text{sh}}$, with $S_{t_i} = 1$ by construction, yields

$$\mathbb{E}\big[D_i \mid \mathcal{F}_{i-1}^{\text{sh}}\big] = \mathbb{E}\big[Y_{t_i}(f) - Y_{t_i}(f') \mid \mathcal{G}_{t_i}^{\text{sh}}\big] \leq 0,$$

which is $H_{0,k}^{\text{sh}}$. Thus, the raw-time conditional null implies the selected-sequence null; disagreement selection does not affect validity precisely because it uses no outcome information. Equivalently, in terms of the raw filtration $\mathcal{F}_{t-1}^{\text{raw}}$, on $\{\mathbb{P}(S_t = 1 \mid \mathcal{F}_{t-1}^{\text{raw}}) > 0\}$, one has $\mathbb{E}[D_t \mid \mathcal{F}_{t-1}^{\text{raw}}, S_t = 1] = \mathbb{E}[D_t^{\text{raw}} \mid \mathcal{F}_{t-1}^{\text{raw}}]/\mathbb{P}(S_t = 1 \mid \mathcal{F}_{t-1}^{\text{raw}}) \leq 0$. By contrast, a merely *marginal* assumption that the candidate is "no better on average" is insufficient for anytime validity: a latent variable that gives the D_i a common sign within a run is compatible with a nonpositive marginal mean but allows the wealth process to drift upward along that run. Ville's inequality does not apply because the conditional-mean-one property does not hold. □

PROOF OF PART (II)

Proof. Step 1 (regularity and the null events). Assume the observation spaces are standard Borel and fix versions of all conditional kernels. Let $\mathcal{F}_0^{\text{sh},k}$ be the σ-field of everything known when the shadow phase of candidate k starts: all previous alarms, training data, promotions, the identity of candidate k and its training randomness. The event

$$N_k := \{H_{0,k}^{\text{sh}} \text{ holds for candidate } k\}$$

is a property of the conditional law of the shadow sequence given $\mathcal{F}_0^{\text{sh},k}$, and with fixed kernel versions it is therefore $\mathcal{F}_0^{\text{sh},k}$-measurable. This regularity is what allows conditioning on N_k below. Adopt the convention $\alpha_k = 0$ for candidates that are not gated, which are never promoted; only k with $\alpha_k > 0$ need be considered. The index k ranges over a countable set, and the number of alarms (hence of candidates) may itself be random and unbounded.

Step 2 (per-candidate bound, conditionally). Fix k with $\alpha_k > 0$. Conditionally on $\mathcal{F}_0^{\text{sh},k}$ and on the event N_k, part (i) applies verbatim: candidate k is fixed before its shadow phase begins, its shadow e-process is a nonnegative supermartingale started at 1, and Fact B.1 gives

$$\mathbb{P}\big(\text{promote } k \,\big|\, \mathcal{F}_0^{\text{sh},k}\big) \;\leq\; \alpha_k \qquad \text{on } N_k.$$

Crucially, the candidate may be chosen *adaptively* using all data collected before its shadow phase, because all of that information is contained in the conditioning field $\mathcal{F}_0^{\text{sh},k}$; the argument never requires the candidate to be independent of the past.

Step 3 (removing the conditioning). Since $N_k \in \mathcal{F}_0^{\text{sh},k}$,

$$\mathbb{P}\big(\{\text{promote } k\} \cap N_k\big) = \mathbb{E}\Big[\mathbf{1}_{N_k}\, \mathbb{P}\big(\text{promote } k \mid \mathcal{F}_0^{\text{sh},k}\big)\Big] \;\leq\; \mathbb{E}[\mathbf{1}_{N_k}\alpha_k] \;\leq\; \alpha_k.$$

Step 4 (family-wise bound). By countable subadditivity and the predeclared spending rule,

$$\mathbb{P}\Big(\bigcup_{k\geq 1} \{\text{promote } k\} \cap N_k\Big) \;\leq\; \sum_{k\geq 1} \mathbb{P}\big(\{\text{promote } k\} \cap N_k\big) \;\leq\; \sum_{k\geq 1} \alpha_k \;\leq\; \alpha_{\text{dep}}.$$

That is, the probability that *any* candidate satisfying its own null is *ever* promoted is at most α_{dep}, uniformly over the random number of alarms.

Step 5 (expected number of erroneous promotions). The summands $\mathbf{1}\{\text{promote } k\}\mathbf{1}_{N_k}$ are nonnegative, so monotone convergence (Tonelli's theorem for counting measure) gives

$$\mathbb{E}\Big[\sum_{k\geq 1}\mathbf{1}\{\text{promote } k\}\mathbf{1}_{N_k}\Big] = \sum_{k\geq 1}\mathbb{P}\big(\{\text{promote } k\}\cap N_k\big) \;\leq\; \sum_{k\geq 1}\alpha_k \;\leq\; \alpha_{\mathrm{dep}}.$$

Both quantities are therefore at most α_{dep}. The levels α_k must be *predeclared*: if they were chosen after observing the data, Step 3 would compare $\mathbb{P}(\text{promote } k \mid \mathcal{F}_0^{\mathrm{sh},k})$ with a random level, and the sum in Step 4 would no longer be bounded by α_{dep}. In the single-intervention runs of the experiments, $\alpha_1 = \alpha_{\mathrm{dep}}$ and all later levels are zero; the multi-intervention runs use $\alpha_k = \alpha_{\mathrm{dep}}2^{-k}$. □