No Free Best-of-Both-Worlds Learning in Repeated Bilateral Trade

Yutian Cheng
Shanghai Jiao Tong University
cyt2021@sjtu.edu.cn

Canzhe Zhao
National University of Singapore
canzhezhao@nus.edu.sg

Jingye Zhao
Shanghai Jiao Tong University
jingye_zhao@sjtu.edu.cn

Shuai Li
The University of Hong Kong
Shanghai Innovation Institute
shuaili@hku.hk

Abstract

We study best-of-both-worlds learning in repeated bilateral trade with two-bit feedback. When seller and buyer valuations are independent, minimax regret scales as $\Theta(T^{2/3})$; under general bounded-density distributions it degrades to $\Theta(T^{3/4})$. This regime-dependent gap naturally raises a question: can a single algorithm adapt optimally to both? We first resolve this negatively: any algorithm achieving $O(T^{\alpha})$ regret on independent-values instances necessarily suffers $\Omega(T^{1-\alpha/3})$ regret on some dependent-values instance. We then characterize the optimal tradeoff and construct a meta-algorithm achieving $\big(\widetilde{O}(T^{\alpha}), \widetilde{O}(T^{1-\alpha/3})\big)$ for every $\alpha \in [2/3, 3/4]$. Finally, we show that this tradeoff can be bypassed with mild side information: knowing the optimal same-price gain suffices to recover the classical best-of-both-worlds guarantee $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{3/4})\big)$.

1 Introduction

Repeated bilateral trade serves as a fundamental model for online posted pricing with partial feedback, attracting significant recent attention in its basic stochastic and smoothed-adversarial forms as well as in budget-balanced, fair, feature-based, and minimal-feedback variants [1–7]. In this setting, a learner repeatedly posts a bid to a seller and an ask to a buyer, receiving only binary acceptance feedback from each party. Remarkably, even in the stochastic i.i.d. setting, the achievable regret rate depends critically on the underlying distributional structure. When seller and buyer values are independent, the minimax rate is $\Theta(T^{2/3})$, but this deteriorates to $\Theta(T^{3/4})$ over the full bounded-density class [1]. This performance gap naturally raises a best-of-both-worlds question, in the same spirit as model-selection and best-of-both-worlds guarantees in online learning more broadly [8–12]:

Can we design a single algorithm that achieves the faster rate on independent-values instances while maintaining minimax optimality over the broader class?

This question is practically unavoidable for real-world market intermediaries, including online marketplaces, ad exchanges, and OTC trading platforms. On the one hand, independent valuations arise naturally in many markets (e.g., standardized commodities and display-ad impressions), and the faster $\widetilde{O}(T^{2/3})$ rate translates directly into higher platform revenue; any intermediary would strongly prefer to exploit this structure when available. On the other hand, value dependence is equally prevalent and well documented in economics: in used-good markets, unobserved item quality links buyer and seller valuations through the classic lemons mechanism [13]; in financial OTC

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

markets, common exposure to macroeconomic shocks and inventory costs creates correlated private values across counterparties [14]; and auction theory has long recognized that the distinction between private-value and common-value paradigms is fundamental yet notoriously difficult to identify from transaction data alone [15, 16]. The bilateral trade problem, already constrained by the celebrated Myerson–Satterthwaite impossibility [17], inherits all of these complications: the intermediary observes only two binary acceptance signals per round, never the actual valuations or their dependence structure. A dependent-values market can therefore masquerade as an independent one during early exploration, causing an algorithm tuned for independence to silently accumulate outsized regret. Conversely, an overly conservative algorithm that hedges against worst-case dependence forfeits the substantial gains available in benign independent-values markets. The platform therefore needs a single adaptive policy that automatically attains the fast rate whenever independence holds while gracefully degrading to minimax optimality otherwise, all without prior structural knowledge.

Our work provides a complete characterization of this adaptation problem with three main results:

- **Classical BoBW is impossible without additional information.** We prove that no single algorithm can simultaneously achieve the optimal rates of $\big(O(T^{2/3}), O(T^{3/4})\big)$ on independent and on general bounded-density instances. In fact, any algorithm attaining $O(T^{\alpha})$ regret on independent-values instances must incur $\Omega(T^{1-\alpha/3})$ regret on some dependent-values instance. We establish this via a novel *pure-copula perturbation* that preserves independent-values marginals while introducing dependence, combined with a localized χ^2 change-of-measure argument that rigorously quantifies the unavoidable cost of two-price exploration.
- **The achievable trade-off frontier is exactly characterized.** We show that the above lower bound is tight by constructing a meta-algorithm that achieves the optimal T^{α} versus $T^{1-\alpha/3}$ regret trade-off for every $\alpha \in [2/3, 3/4]$. The algorithm directly measures how far the true gain-from-trade curve deviates from its product-law surrogate (derived from marginals alone) on a fine grid, and switches strategies accordingly: if the deviation is small (i.e., the surrogate is valid), we use *temporal decoupling* to synthesize virtual product-law feedback from historical joint-law data, safely offloading to black-box independent-values subroutines; if the deviation is large, we fall back to a robust fixed-price strategy.
- **One scalar suffices to restore classical BoBW.** We demonstrate that the impossibility barrier vanishes with mild side information: knowing only the optimal same-price value g_ν^* suffices to recover the pair $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{3/4})\big)$. This is non-obvious, since g_ν^* alone reveals neither the optimal price location nor realized trade gains; we achieve it via a *sparse audit-and-switch wrapper* that runs a phase-local independent-values routine on same-price rounds and occasionally inserts unbiased two-price probes to estimate the cumulative deficit against g_ν^*, switching to a robust fallback once the deficit exceeds a threshold. A key enabling ingredient is a *prefix-stable phase learner* that strengthens the standard $\widetilde{O}(T^{2/3})$ guarantee to simultaneous all-prefix high-probability regret control, preventing spurious switches during the audit.

1.1 Techniques and challenges

Achieving best-of-both-worlds performance in bilateral trade poses unique challenges because the learner only observes two acceptance bits per round. Our goal is to adaptively match the regime-specific minimax rates of [1], and doing so requires new ideas at each stage of the problem.

Identifying the fundamental impossibility. To prove the fundamental impossibility of achieving the regime-specific rates simultaneously, we design a novel pure-copula perturbation combined with a localized χ^2 change-of-measure argument. The main challenge in establishing this impossibility is that one must construct a dependent-values instance that remains indistinguishable from an independent-values (IV) instance under standard single-price exploration, yet strictly penalizes any algorithm that trusts the IV structure too early. In particular, any visible change in the marginal acceptance probabilities would immediately reveal the dependence to the learner. Our pure-copula perturbation precisely overcomes this obstacle by altering the joint distribution while keeping the IV marginals exactly intact. Consequently, detecting the hidden dependence forces the learner to play strictly off-diagonal, two-price exploratory actions ($P_t \ll Q_t$) that inevitably fall into a costly

exploration strip. Our localized χ^2 change-of-measure argument then rigorously quantifies the inescapable cost of such two-price exploration, yielding the exact quantitative trade-off.

Achieving the optimal trade-off. To achieve the optimal trade-off frontier between the two regimes, we develop an online surrogate-consistency certificate and a temporal decoupling technique for virtual histories. The challenge here is that reaching the exact trade-off frontier is considerably subtler than standard change-detection [11, 18–20]. The learner cannot simply test for product-surrogate validity and switch to an IV routine; instead we bound the *surrogate mismatch*, namely the gap between the true gain-from-trade curve and the surrogate curve derived solely from marginals, uniformly over all prices. Estimating this object uniformly is difficult because the test must be robust enough to support future adaptive queries while consuming only a limited diagnostic budget. Furthermore, even if the diagnostic test passes, the historical data was generated under the true joint law rather than the product law, making it incompatible with black-box IV subroutines. Our online surrogate-consistency certificate addresses the first challenge by employing a carefully calibrated grid-based diagnostic stage that constructs an unbiased estimator of the joint gain curve and a plug-in estimator of its product surrogate within tight budget constraints. To resolve the second challenge, our temporal decoupling argument synthesizes fresh virtual feedback that perfectly emulates the correct product-law structure, allowing us to safely feed the subsequent interaction history to any standard IV learner.

Learning with extra information. To bypass the impossibility and recover the optimal $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{3/4})\big)$ rates with the optimal value g_ν^* as extra information, we design a sparse audit-and-switch wrapper powered by a prefix-stable phase IV learner. The challenge in utilizing g_ν^* is that, unlike prior hint-based settings with bandit feedback [21–25], knowing this scalar reveals neither the optimal price location nor the realized gain from trade, meaning the learner cannot directly compare its actions against the benchmark. Additionally, deciding when to switch away from the IV subroutine requires checking the accumulated deficit against g_ν^* at intermediate checkpoints, making the algorithm vulnerable to spurious switches if the base learner's performance fluctuates. Our sparse audit-and-switch wrapper overcomes the lack of observable gains by running a phase-local IV routine on reserved same-price rounds and inserting infrequent, unbiased two-price probes to estimate the cumulative deficit against g_ν^*, gracefully switching to a robust fallback once the deficit exceeds a threshold. To prevent spurious switches, our prefix-stable phase IV learner strengthens the standard $\widetilde{O}(T^{2/3})$ expected bound of [1] by providing simultaneous, all-prefix high-probability regret control within each phase, ensuring the audit mechanism only triggers when the product-surrogate is genuinely violated.

1.2 Related work

Bilateral trade in online learning. Repeated bilateral trade has emerged as the canonical model for online posted pricing under partial feedback, with established regret bounds of $\Theta(T^{2/3})$ for independent values and $\Theta(T^{3/4})$ for bounded-density distributions when the regime is known [1–3]. Extensions include global budget balance [4], fair gain-from-trade objectives [5], and feature-based trade [6]. However, all prior work assumes the learner knows the underlying regime structure, leaving unaddressed the fundamental challenge we tackle: determining from identical two-bit observations whether product-based reasoning is justified.

Best-of-both-worlds and adaptivity. Our problem connects to the broader literature on adaptive learning and best-of-both-worlds (BoBW) guarantees, which seek optimal performance across different environments while retaining worst-case guarantees. This line of work spans the classical stochastic-versus-adversarial setting [8, 9], instance-dependent and problem-dependent guarantees for Lipschitz, kernelized, and contextual bandits [26–29], and general approaches to adaptation and model selection [10, 11, 30]. Related work develops simultaneous stochastic and adversarial guarantees for multi-armed and semi-bandits, including Tsallis-INF and hybrid regularization [31–33]; extends FTRL and black-box reductions to linear bandits and richer feedback structures [34–38]; and studies robustness to heavy-tailed observations, corruption, and information acquisition [39, 40]. Research automation has also been explored in related settings [41]. In contrast to these works, our setting remains stochastic and i.i.d.; the unknown structure is within-round independence between seller and buyer values. The learner must certify this structure from limited two-bit feedback, and

our lower bound quantifies the resulting structural-identification cost rather than a generic stochastic-versus-adversarial adaptation cost.

Side information and adaptive learning. The impossibility of regime-agnostic adaptation motivates our study of side information, aligning with work showing how hints can resolve otherwise intractable adaptivity problems [21–23]. Known optimum values constitute a particularly powerful class of hints across diverse settings: thresholding bandits [24], smoothness-adaptive X-armed bandits [42], and Bayesian optimization [25]. Our contribution extends this paradigm to bilateral trade: we prove that the scalar hint (optimal single-price gain) suffices to restore classical best-of-both-worlds guarantees, despite the learner remaining ignorant of optimal price locations and never observing realized trade gains.

2 Problem Setting

We adopt the repeated bilateral trade protocol with two-bit feedback and weakly-budget-balanced two-price posting from [1], further studied in [2–4]. At each round $t = 1, \dots, T$, a fresh seller and buyer arrive with private valuations $(S_t, B_t) \in [0,1]^2$ drawn i.i.d. from an unknown law ν on $[0,1]^2$. Without observing (S_t, B_t), the learner posts a bid-ask pair $(P_t, Q_t) \in \mathcal{U} := \{(p,q) \in [0,1]^2 : p \le q\}$, where P_t is quoted to the seller and Q_t to the buyer. Trade occurs iff both agents accept, i.e. iff $S_t \le P_t$ and $Q_t \le B_t$; under $P_t \le Q_t$, this is equivalent to $S_t \le P_t \le Q_t \le B_t$.

Following the weakly-budget-balanced two-price convention of [1–3], the reward is the agents' residual surplus, equivalently the weakly-budget-balanced two-price gain from trade,

$$Y_t := \big((P_t - S_t) + (B_t - Q_t)\big)\mathbf{1}\{S_t \le P_t \le Q_t \le B_t\}.$$

Thus Y_t equals gross executed surplus minus the intermediary's spread. In the same-price case $P_t = Q_t = x$, the spread vanishes and this reward reduces to the standard single-price gain from trade, $Y_t = (B_t - S_t)\mathbf{1}\{S_t \le x \le B_t\}$. The learner observes neither (S_t, B_t) nor Y_t, and only receives the two acceptance bits

$$X_t^S := \mathbf{1}\{S_t \le P_t\}, \qquad X_t^B := \mathbf{1}\{B_t \ge Q_t\}.$$

Objective and benchmark. Define $r_\nu(p,q) := \mathbb{E}_\nu\big[\big((p-S)+(B-q)\big)\mathbf{1}\{S \le p \le q \le B\}\big]$, and let $g_\nu(x) := r_\nu(x,x) = \mathbb{E}_\nu[(B-S)\mathbf{1}\{S \le x \le B\}]$ be the same-price gain-from-trade curve. The best fixed two-price benchmark collapses to the best same-price benchmark:

$$g_\nu^* := \sup_{x \in [0,1]} g_\nu(x) = \sup_{(p,q) \in \mathcal{U}} r_\nu(p,q). \tag{1}$$

Indeed, on $\mathcal{U}$ we have $\{S \le p \le q \le B\} \subseteq \{S \le q \le B\}$ and $p - S + B - q \le B - S$, so $r_\nu(p,q) \le g_\nu(q) \le g_\nu^*$; the reverse direction follows by taking $p = q$. Accordingly, we measure the performance of a learning algorithm A by the realized regret and its expectation,

$$\mathfrak{R}_T(A;\nu) \;:=\; T\,g_\nu^* - \sum_{t=1}^{T} Y_t, \qquad R_T(A;\nu) \;:=\; \mathbb{E}_\nu[\mathfrak{R}_T(A;\nu)] \;=\; T\,g_\nu^* - \mathbb{E}_\nu\Big[\textstyle\sum_{t=1}^{T} Y_t\Big],$$

which, by (1), equals the regret relative to the best fixed two-price action in hindsight.

Instance classes. For $M > 0$, let $\mathcal{H}(M)$ be the class of laws on $[0,1]^2$ with joint density bounded by M, and let $\mathcal{E}_{\mathrm{IV}}(M) \subseteq \mathcal{H}(M)$ be the subclass in which S and B are independent. Under the weakly-budget-balanced two-price gain-from-trade reward above, the known minimax regret rates with two-bit feedback are $\Theta(T^{2/3})$ on $\mathcal{E}_{\mathrm{IV}}(M)$ and $\Theta(T^{3/4})$ on $\mathcal{H}(M)$ [1–3].

3 The Classical BoBW Lower Bound

We begin with the natural question: can one algorithm simultaneously match the two fixed-regime endpoint rates from [1], namely $O(T^{2/3})$ on $\mathcal{E}_{\mathrm{IV}}(M)$ and $O(T^{3/4})$ on the full class $\mathcal{H}(M)$? The first theorem shows that it is impossible and reveals the tradeoff frontier that will organize the rest of the paper.

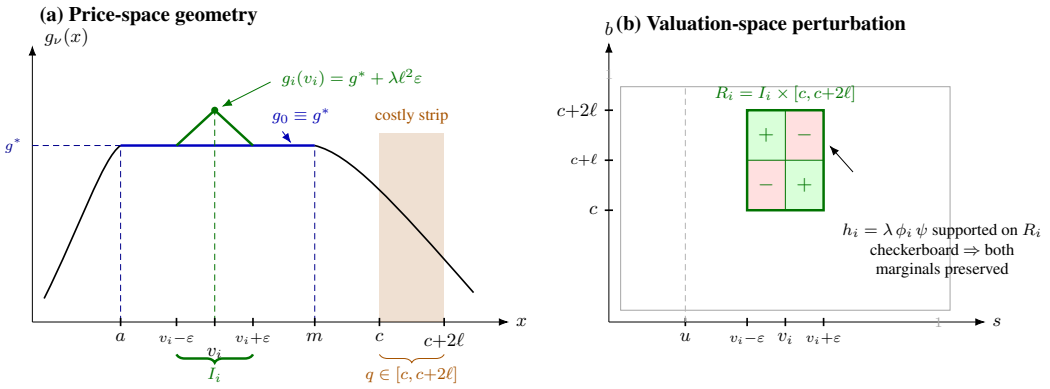

Figure 1: Geometry of the hard family underlying Theorem 1. **(a)** Under the IV base ν_0, the gain-from-trade curve g_0 has a flat optimal plateau on $[a, m]$ at level g^*, while any price $q \in [c, c + 2\ell]$ beyond the plateau lies in a *costly exploration strip* (shaded). The perturbation ν_i adds a small bump over $I_i = [v_i - \varepsilon, v_i + \varepsilon]$, creating a unique hard optimum $g_i(v_i) = g^* + \lambda\ell^2\varepsilon$. **(b)** The density perturbation $h_i = \lambda\phi_i\psi$ is supported on the rectangle $R_i = I_i \times [c, c + 2\ell]$ and has a checkerboard sign pattern $(+, -, -, +)$; this annihilates both the s- and b-marginals, so ν_i is indistinguishable from ν_0 on every one-dimensional test.

Theorem 1 (No-extra-information lower bound)**.** *There exists a density level $M_* > 0$ such that, for every exponent $\alpha \in [0, 1]$ and every algorithm A,*

$$\sup_{\nu \in \mathcal{E}_{\mathrm{IV}}(M_*)} R_T(A; \nu) = O(T^\alpha) \quad \Longrightarrow \quad \sup_{\mu \in \mathcal{H}(M_*) \setminus \mathcal{E}_{\mathrm{IV}}(M_*)} R_T(A; \mu) = \Omega\big(T^{1-\alpha/3}\big).$$

In particular, any algorithm with $O(T^{2/3})$ regret on $\mathcal{E}_{\mathrm{IV}}(M_*)$ must incur $\Omega(T^{7/9})$ regret on some non-IV instance, so the target pair $(2/3, 3/4)$ is impossible.

Proof sketch. The lower bound rests on a single explicit construction, sketched in Figure 1. The easy base ν_0 is an independent-values law whose gain-from-trade curve g_0 is *flat on a plateau* $[a, m]$ at the optimum value g^* and strictly decreases outside it (panel (a)). In particular, any price $q \geq c > m$ lies in a *costly exploration strip*: every round with $Q_t \in [c, c + 2\ell]$ contributes at least a constant regret Δ_J relative to g^*.

For each index i we then add a non-product perturbation $h_i = \lambda\phi_i\psi$ supported on a small rectangle $R_i = I_i \times [c, c + 2\ell]$ with $I_i = [v_i - \varepsilon, v_i + \varepsilon]$ (panel (b)). The perturbation has a *checkerboard sign pattern* $(+, -, -, +)$ on the four subrectangles of R_i, which makes it a pure copula perturbation: both the s- and b-marginals are annihilated, so ν_i and ν_0 agree on all one-dimensional observations. Inside the plateau, however, the bump raises the gain at the tip of I_i to $g_i(v_i) = g^* + \lambda\ell^2\varepsilon$, creating a *unique hard optimum* that is undetectable from marginal statistics.

Because identifying i requires separating ν_i from ν_0 under the full two-bit feedback vector, and the perturbation is concentrated on R_i, every informative sample must satisfy both $P_t \in I_i$ and $Q_t \in [c, c + 2\ell]$. Since $I_i \subset (a, u)$ and $[c, c + 2\ell] \subset (m, 1)$ are disjoint, any such action is *strictly off-diagonal* with $Q_t - P_t \geq m - u$; in particular, no single-price action can ever be informative, and the argument genuinely relies on the two-price action space $\mathcal{U}$. If the easy-side regret is $O(T^\alpha)$, the expected number N_T of rounds with $Q_t \geq c$ under ν_0 is at most $O(T^\alpha/\Delta_J)$. A localized χ^2/KL change-of-measure argument over the $K = \Theta(\varepsilon^{-1})$ rectangles then transfers this bound to a lower bound on the non-IV side, and optimizing over ε yields the exponent $1 - \alpha/3$. The full change-of-measure argument, including the density bound M_* and the precise χ^2 inequalities, is carried out in Appendix A.

4 The Optimal Tradeoff Frontier

Having identified the impossibility of classical BoBW, we now investigate what is achievable without extra information. The central object governing this question is the *surrogate mismatch*: the gap between the true same-price curve and the curve obtained by tensorizing the marginals.

For any one-round law ν, let $\nu^{\otimes}$ denote the product of its marginals and define the product surrogate $\tilde{g}_\nu(x) := g_{\nu^{\otimes}}(x)$. Writing $F_\nu(x) = \mathbb{P}_\nu(S \le x)$, $\overline{F}_\nu(x) = \mathbb{P}_\nu(B \ge x)$, $\Gamma_\nu(x) = \int_0^x F_\nu(p)\,dp$, and $H_\nu(x) = \int_x^1 \overline{F}_\nu(q)\,dq$, Lemma 8 gives

$$\tilde{g}_\nu(x) = \overline{F}_\nu(x)\Gamma_\nu(x) + F_\nu(x)H_\nu(x).$$

Thus $g_\nu = \tilde{g}_\nu$ for every product law, so the entire problem is to decide whether $\Delta_\nu := g_\nu - \tilde{g}_\nu$ is small enough that one may safely run an IV learner.

Algorithm 1 $\mathsf{SCM}^{(\gamma)}_{M,T}$: Surrogate-Consistency Meta-Strategy

Input: density bound M, horizon T, exponent $\gamma \in (0, 1/3)$.
Set $\varepsilon := T^{-\gamma}$, $K := \lceil \varepsilon^{-1} \rceil$, and $x_j := j/K$ for $j = 0, \dots, K$.
Set $m := \left\lceil \frac{\log(24KT^2)}{2\varepsilon^2} \right\rceil$ and $N_0 := 2m(K-1)$.
if $N_0 \ge T$ **then**
    Play any fixed same price for all T rounds and stop.
end if
Initialize $\widehat{g}_0 = \widehat{g}_K = 0$.
Diagnostic stage.
for each interior grid point x_j **do**
    Run BATCHPROBE(x_j, m) from Algorithm 4 and form $(\widehat{g}_j, \widetilde{g}_j)$ via (2).
end for
Set $D_{\text{test}} := \max_{1 \le j \le K-1} |\widehat{g}_j - \widetilde{g}_j|$.
if $D_{\text{test}} \le 8\varepsilon$ **then**
Accept branch.
    Initialize a fresh IV learner $\mathcal{L}_{M^2,H}$ by Algorithm 3, with $H := \lfloor (T - N_0)/2 \rfloor$.
    for $k = 1, 2, \dots, H$ **do**
        Query the learner for its next same price X_k; set $x_{\text{last}} \leftarrow X_k$.
        Play (X_k, X_k) on the next two real rounds and stream the synthetic feedback (3) back to $\mathcal{L}_{M^2,H}$.
    end for
    If one real round remains after those H virtual rounds, play $(x_{\text{last}}, x_{\text{last}})$ on that round; if $H = 0$, use the default price $x_{\text{last}} := 0$.
else
Reject branch.
    Choose $\widehat{j} \in \arg\max_{0 \le j \le K} \widehat{g}_j$, set $\widehat{x} := x_{\widehat{j}}$, and play $(\widehat{x}, \widehat{x})$ for all remaining rounds.
end if

Algorithm 1 implements exactly this decision. It first computes the diagnostic budget and, if that budget already exhausts the horizon, simply plays a fixed same price throughout. Otherwise, the diagnostic stage estimates both g_ν and $\tilde{g}_\nu$ on an ε-grid via BATCHPROBE from Algorithm 4; the algorithm accepts iff the empirical mismatch is uniformly small, then restarts a fresh IV learner and uses temporal decoupling to emulate product-law feedback, and otherwise commits to the empirical best grid price. In one line: test product-surrogate validity, then either reduce to IV or fall back to a robust same-price action. The surrogate-consistency certificate must be uniform because the IV routine from Theorem 11, accessed through Algorithm 3, later queries prices adaptively.

More specifically, fix the mesh $\varepsilon := T^{-\gamma}$ and let $K := \lceil \varepsilon^{-1} \rceil$. Set $x_j := j/K$ for $j = 0, \dots, K$. At each interior grid point, Stage I runs BATCHPROBE(x_j, m) with batch size $m := \left\lceil \frac{\log(24KT^2)}{2\varepsilon^2} \right\rceil$. The diagnostic budget is $N_0 := 2m(K-1)$, which is the total number of rounds used by Stage I. If $N_0 \ge T$, the policy simply plays a fixed same price; below we describe only the nontrivial case

$N_0 < T$. The resulting samples define the estimators

$$\widehat{g}_j := x_j \frac{1}{m}\sum_{\ell=1}^{m} Y_{j,\ell}^{L} + (1-x_j)\frac{1}{m}\sum_{\ell=1}^{m} Y_{j,\ell}^{R}, \quad \widetilde{g}_j := \widehat{\overline{F}}_j \widehat{\Gamma}_j + \widehat{F}_j \widehat{H}_j, \tag{2}$$

where the probe variables are those returned by Algorithm 4. The same batch estimates both g_ν and $\tilde{g}_\nu$: $Y_{j,\ell}^{L}$ and $Y_{j,\ell}^{R}$ target the joint-tail integrals in g_ν, while $(\widehat{\overline{F}}_j, \widehat{\Gamma}_j, \widehat{F}_j, \widehat{H}_j)$ target the factors of $\tilde{g}_\nu$. Lemma 12 identifies these targets, and Lemma 13 yields simultaneous gridwise accuracy with probability at least $1 - 1/(2T^2)$.

If the acceptance test in Algorithm 1 accepts, the algorithm restarts a fresh IV learner and feeds it temporally decoupled same-price feedback, thereby turning the accepted branch into a run against $\nu^{\otimes}$. Concretely, on the accept branch the learner receives the synthetic same-price feedback

$$\widetilde{Z}_k := \big(\mathbf{1}\{S_{N_0+2k-1} \le X_k\},\, \mathbf{1}\{B_{N_0+2k} \ge X_k\}\big). \tag{3}$$

Since the environment is i.i.d. across rounds, Lemma 14 shows that, conditionally on the diagnostic stage, this virtual history is exactly a fresh run of the IV learner against $\nu^{\otimes}$.

Theorem 2 (Achieving the no-extra-information frontier via surrogate consistency)**.** *Fix $M > 0$ and $\gamma \in (0, 1/3)$. There exists an algorithm, namely Algorithm 1, such that*

$$\sup_{\nu \in \mathcal{E}_{\mathrm{IV}}(M)} R_T\big(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu\big) \le C_1(M,\gamma)\, T^{\max\{2/3,\, 3\gamma\}} \log(eT),$$

and

$$\sup_{\nu \in \mathcal{H}(M)} R_T\big(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu\big) \le C_2(M,\gamma)\, T^{\max\{1-\gamma,\, 3\gamma\}} \log(eT).$$

Proof sketch. Let $\mathcal{F}_0$ be the sigma-field generated by Stage I, and let Acc, Rej be the acceptance and rejection events. We decompose the regret of $\mathsf{SCM}_{M,T}^{(\gamma)}$ as $\mathfrak{R}_T(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu) = \mathfrak{R}_{\mathrm{I}} + \mathfrak{R}_{\mathrm{II}}$, where the first term $\mathfrak{R}_{\mathrm{I}} := N_0 g_\nu^* - \sum_{t=1}^{N_0} Y_t$ captures the gap incurred in the diagnostic stage. By construction $N_0 = 2m(K-1)$, so $\mathfrak{R}_{\mathrm{I}} \le N_0 = \widetilde{O}(T^{3\gamma})$ pathwise.

We then define the event

$$\mathcal{E} := \Big\{ |\widehat{g}_j - g_\nu(x_j)| \le 2\varepsilon,\ |\widetilde{g}_j - \tilde{g}_\nu(x_j)| \le 4\varepsilon \text{ for all } j = 1, \dots, K-1 \Big\}.$$

By Lemma 13, $\mathbb{P}_\nu(\mathcal{E}^c) \le 1/(2T^2)$. On $\mathcal{E}$, Corollary 1 implies that every IV law is accepted and, more generally, that acceptance yields

$$D_\nu := \sup_{x \in [0,1]} |g_\nu(x) - \tilde{g}_\nu(x)| \le C_{\mathrm{test}}(M)\varepsilon.$$

On $\mathcal{E} \cap \mathsf{Acc}$, Lemma 15 gives

$$\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0] \le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 4HC_{\mathrm{test}}(M)\varepsilon + 1 = O(T^{2/3} + T^{1-\gamma}).$$

The $H^{2/3}$ term comes from Lemma 14, which reduces the accept branch to a fresh run of the IV subroutine against the product law; the $H\varepsilon$ term accounts for the mismatch $D_\nu \le C_{\mathrm{test}}(M)\varepsilon$ between the true and product gain curves (Corollary 1). On $\mathcal{E} \cap \mathsf{Rej}$, Lemma 16 gives

$$\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0] \le (2M+4)(T - N_0)\varepsilon = O(T^{1-\gamma}).$$

The intuition is grid commitment: on $\mathcal{E}$ each estimated gain is within $O(\varepsilon)$ of the true one, so the empirical maximizer is within $O(\varepsilon)$ of the optimum. Since Acc and Rej partition $\mathcal{E}$, the conditional regret on $\mathcal{E}$ is at most $\max\{O(T^{2/3} + T^{1-\gamma}),\, O(T^{1-\gamma})\} = O(T^{2/3} + T^{1-\gamma})$. Therefore, for any $\nu \in \mathcal{H}(M)$,

$$R_T(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu) \le N_0 + O\big(T^{2/3} + T^{1-\gamma}\big) + T\mathbb{P}_\nu(\mathcal{E}^c),$$

whereas for $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$ we have $D_\nu = 0$, so $\mathcal{E}$ forces acceptance and hence

$$R_T(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu) \le N_0 + O\big(T^{2/3}\big) + T\mathbb{P}_\nu(\mathcal{E}^c).$$

Using $N_0 = \widetilde{O}(T^{3\gamma})$ and $\mathbb{P}_\nu(\mathcal{E}^c) \le 1/(2T^2)$ yields the claimed hard- and easy-side rates. The complete proof is deferred to Appendix B. A direct specialization recovers the lower-bound frontier on the interval $\alpha \in [2/3, 3/4]$: choosing $\gamma = \alpha/3$ gives

$$\sup_{\nu \in \mathcal{E}_{\mathrm{IV}}(M)} R_T\big(\mathsf{SCM}_{M,T}^{(\alpha/3)}; \nu\big) = \widetilde{O}(T^{\alpha}), \qquad \sup_{\nu \in \mathcal{H}(M)} R_T\big(\mathsf{SCM}_{M,T}^{(\alpha/3)}; \nu\big) = \widetilde{O}\big(T^{1-\alpha/3}\big).$$

In particular, $\gamma = 2/9$ yields the endpoint pair $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{7/9})\big)$.

5 Restoring Classical BoBW with Value Information

Theorems 1 and 2 show that without extra information the classical pair $\big(O(T^{2/3}), O(T^{3/4})\big)$ is impossible. A natural next step is to determine what side information suffices. Inspired by the literature on known-optimum hints for bandits and optimization [21, 22, 24, 25, 42], we study the *exact-value* setting, where the learner is given only the scalar $V := g^*_\nu$. Despite its simplicity, exploiting V is nontrivial: it reveals neither the optimal price location nor the realized gain from trade, and one must detect IV-violations via online certification under a limited diagnostic budget. More importantly, it does not break the lower bound construction in [1, 2], so the minimax rates of $\big(\Theta(T^{2/3}), \Theta(T^{3/4})\big)$ can still be preserved on each regime. As we show, this single scalar suffices to bypass the impossibility and restore the classical best-of-both-worlds guarantee. Recovering the classical rates requires two new algorithmic ingredients.

The sparse audit-and-switch mechanism. Even given the optimal value $V = g^*_\nu$, the learner never observes the realized gain from trade, so it cannot directly compare its actions against V. We address this with a *sparse audit-and-switch wrapper* (Algorithm 2) that periodically checks whether the same-price trajectory remains consistent with V, and switches to a robust fallback if a persistent deficit is detected.

The meta-algorithm splits the horizon into phases each with budgets B_k. Each phase k consists of three parts: a *burn-in* period of w_k rounds where an IV routine runs freely to accumulate data, followed by J_k *audited blocks* of increasing lengths ($b_j = 2j - 1$ for $j = 1, \dots, J_k$), and finally an *unaudited tail* of r_k rounds. The block lengths grow linearly, so the total number of same-price rounds monitored grows quadratically with J_k.

After each audited block j, the wrapper inserts a single *fresh audit round* using an unbiased probe mechanism ValueProbe (Algorithm 5, Lemma 17). This probe samples a price uniformly from the current block and plays a carefully designed two-price action to obtain an unbiased estimate $Y_{k,j} \in [0, 2]$ of the average gain achieved in that block. The critical design constraint is that audit rounds are *disjoint* from the IV learner's trajectory and are never fed back to the IV routine; by Lemma 26 this ensures the audits do not contaminate the IV learner's data distribution. At each checkpoint j, the wrapper maintains a checkpoint statistic

$$\widehat{G}_{k,j} := \frac{1}{j^2} \sum_{i=1}^{j} b_i Y_{k,i}, \qquad \widehat{\delta}_{k,j} := V - \widehat{G}_{k,j}, \tag{4}$$

which aggregates the j audit estimates, weighting each by the corresponding block length b_i. By Lemma 19, $\widehat{G}_{k,j}$ is an unbiased estimator of the average gain over the first j^2 monitored same-price rounds, and concentrates around its expectation.

A prefix-stable phase IV learner. Making the audit-and-switch wrapper work requires a stronger IV subroutine than the original horizon-level $\widetilde{O}(n^{2/3})$ *expected* guarantee of [1]. The reason is that switching decisions are taken at many intermediate checkpoints within each phase, not only at the end. Inspired by the algorithm of [7], we design $\mathsf{PhaseScout}_{n,\delta_{\mathrm{ph}}}$ (Algorithm 7) to provide a *simultaneous all-prefix high-probability bound* (Proposition 4): with probability $\geq 1 - \delta_{\mathrm{ph}}$, every prefix length m satisfies

$$\sum_{t=w_n+1}^{w_n+m} \big(V - g_\nu(X_t)\big) \leq \widetilde{O}\big(mn^{-1/3} + n^{1/6}\sqrt{m}\big). \tag{5}$$

With the two ingredients ready, we now introduce the switch rule. At checkpoint j it compares $\widehat{\delta}_{k,j}$ against $\Theta_k + 8\beta_{k,j}$, where $\Theta_k = \widetilde{\Theta}(n_k^{-1/3})$ and $\beta_{k,j}$ is a confidence bound for $\widehat{G}_{k,j}$. If a switch is triggered, the wrapper abandons the IV strategy and delegates to a robust fallback RF_H (Algorithm 6, Proposition 2) for the remaining horizon.

Theorem 3 (Exact value restores classical BoBW)**.** *For every fixed $M > 0$, there exists an exact-value algorithm $\mathsf{VAM}_T(V)$ such that*

$$\sup_{\nu \in \mathcal{H}(M)} R_T(\mathsf{VAM}_T(V); \nu) \leq C_3(M)\,\mathrm{polylog}(T)\,T^{3/4},$$

Algorithm 2 $\mathsf{VAM}_T(V)$: Value-Auditing Meta-Strategy

Input: density bound M, horizon T, exact optimum value $V = g^*_\nu$.
Precompute $(B_k, J_k, n_k, \eta_k, w_k, r_k, \Theta_k, j_{0,k}, \beta_{k,j})_{k\geq 1}$ from (17)–(21).
for phases $k = 1, 2, \ldots$ until the horizon is exhausted **do**
  Initialize $\mathsf{PhaseScout}(n_k, \eta_k)$ via Algorithm 7 and run $\mathsf{PhaseScout}$ for w_k rounds (burn-in).
  for checkpoints $j = 1, \ldots, J_k$ **do**
    Run $\mathsf{PhaseScout}$ for $b_j := 2j - 1$ rounds; collect the price sequence $(x_1, \ldots, x_{b_j})$.
    Sample $\tau \sim \mathrm{Unif}\{1, \ldots, b_j\}$ and set $Y_{k,j} \leftarrow \textsc{ValueProbe}(x_\tau)$ via Algorithm 5.
    Update $(\widehat{G}_{k,j}, \widehat{\delta}_{k,j})$ via (4).
    if $j \geq j_{0,k}$ and $\widehat{\delta}_{k,j} > \Theta_k + 8\beta_{k,j}$ **then**
      Switch to RF_H (Algorithm 6) on the remaining horizon H and terminate.
    end if
  end for
  Run $\mathsf{PhaseScout}$ for r_k rounds (tail).
end for

while simultaneously

$$\sup_{\nu \in \mathcal{E}_{\mathrm{IV}}(M)} R_T(\mathsf{VAM}_T(V); \nu) \leq C_4(M)\,\mathrm{polylog}(T)\,T^{2/3}.$$

Thus the side information $V = g^*_\nu$ recovers the classical best-of-both-worlds pair.

Proof sketch. We analyze the regret by considering whether the algorithm ever triggers a switch. Let

$$\mathcal{E}_{\mathrm{aud}} := \bigcap_{k=1}^{K_T} \bigcap_{j=1}^{J_k} \{|\widehat{G}_{k,j} - G_{k,j}| \leq \beta_{k,j}\},$$

where $\widehat{G}_{k,j}$ and $G_{k,j}$ are the weighted checkpoint estimator and its expectation in phase k at checkpoint j (Lemma 19); this is the key concentration event underlying the analysis.

If the environment lies in $\mathcal{H}(M)$ and no switch occurs in phase k, then on the concentration event $\mathcal{E}_{\mathrm{aud}}$(Lemma 19) the true gap at the final checkpoint $j = J_k$ satisfies $\delta_{k,J_k} = V - G_{k,J_k} \leq \Theta_k + 9\beta_{k,J_k} = \widetilde{O}(n_k^{-1/4})$. The J_k^2 audited rounds contribute $\widetilde{O}(n_k^{3/4})$ regret (Lemma 23(ii)), and the burn-in and tail costs are of the same order; summing over the $K_T = O(\log T)$ geometric phases gives $\widetilde{O}(T^{3/4})$. If instead a switch is triggered at phase k, the fallback RF_H (Proposition 2) incurs at most $\widetilde{O}(T_k^{3/4})$ on the remaining horizon $T_k \leq T$, so the same bound holds. Hence $R_T^{\mathrm{gen}} = \widetilde{O}(T^{3/4})$.

If the environment lies in $\mathcal{E}_{\mathrm{IV}}(M)$, the reserved same-price trajectory of each phase effectively replicates an independent run of $\mathsf{PhaseScout}$ (Lemma 26). By the all-prefix bound of Proposition 4, for any reached checkpoint $j \geq j_{0,k}$ the true gap satisfies $\delta_{k,j} \leq \Theta_k/32 + \beta_{k,j}$, which together with the concentration event $\mathcal{E}_{\mathrm{aud}}$ implies $\widehat{\delta}_{k,j} \leq \Theta_k + 8\beta_{k,j}$; consequently the switch is never triggered. The total regret is therefore the sum of phase-IV costs $\sum_k \widetilde{O}(n_k^{2/3}) = \widetilde{O}(T^{2/3})$ plus the audit overhead $\sum_k O(J_k) = O(\sqrt{T})$ (Lemma 23(iii),(iv)), yielding $R_T^{\mathrm{iv}} = \widetilde{O}(T^{2/3})$.

The failure probability of $\mathcal{E}_{\mathrm{aud}}$ is controlled by a union bound over $O(\sqrt{T})$ checkpoints, giving $\mathbb{P}(\mathcal{E}_{\mathrm{aud}}^c) \leq 1/T^2$ (Lemma 19); the failure probability of the IV prefix events $\mathcal{E}_k^{\mathrm{IV}}$ is bounded via Proposition 4, and the $O(\delta)$ contributions are absorbed into the $\widetilde{O}(\cdot)$ terms. Full details are given in Section C.

6 Further Discussion

Two welfare conventions for gain from trade. The bilateral-trade literature uses two standard welfare conventions. We use *residual-surplus GFT*, which counts the utilities of the seller and buyer under a weakly budget-balanced bid–ask pair:

$$r^{\mathrm{res}}(p, q; s, b) = ((p - s) + (b - q))\mathbf{1}\{s \leq p \leq q \leq b\}.$$

Another common convention is *gross-surplus GFT*,

$$r^{\mathrm{gross}}(p, q; s, b) = (b - s)\mathbf{1}\{s \le p \le q \le b\}.$$

The former is used, for example, in the two-price formulation of [3]; the latter is standard in work that counts the intermediary's spread as part of social welfare [1, 7]. The two rewards coincide on every same-price action $p = q = x$, where both equal $(b - s)\mathbf{1}\{s \le x \le b\}$. They differ only for off-diagonal exploration, with

$$r^{\mathrm{res}}(p, q; s, b) \le r^{\mathrm{gross}}(p, q; s, b) \le r^{\mathrm{gross}}(q, q; s, b), \qquad p \le q.$$

Consequently, the optimal fixed-action benchmarks under the two conventions are identical: the best two-price gross-surplus benchmark equals the best same-price value g^*_ν. For any fixed action sequence, the gross-surplus regret is no larger than the residual-surplus regret, since the benchmark is the same and the gross reward is pointwise larger. Thus every upper bound proved here for residual-surplus regret, including the trade-off guarantee and the exact-value guarantee, transfers directly to gross-surplus GFT.

The lower bound transfers as well. Under the independent base instance ν_0, every informative off-diagonal action has its buyer price in the costly exploration strip, so its expected gross reward is bounded by the same plateau optimum minus a constant exploration gap. The checkerboard perturbation preserves both marginals; its buyer-coordinate component integrates to zero, so the additional spread-revenue term does not remove the hidden gain spike used in the change-of-measure argument. Finally, the feedback law and its KL divergence depend only on the two acceptance bits, not on which welfare convention is used. Hence the same lower-bound exponent and the same adaptation trade-off hold for gross-surplus GFT. In short, the convention changes how intermediary revenue is classified on exploratory rounds, but changes neither the benchmark nor any regret exponent.

7 Conclusion and Limitations

In this paper, we studied whether a single learner can adapt between independent-values and general bounded-density repeated bilateral trade under two-bit feedback. We first identified a fundamental barrier: without extra information, the classical best-of-both-worlds pair $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{3/4})\big)$ cannot be achieved, and more generally the easy-side and hard-side guarantees must obey the tradeoff T^α versus $T^{1-\alpha/3}$ (Theorem 1). We then developed a surrogate-consistency meta-algorithm showing that this frontier is achievable up to logarithmic factors, which gives a sharp characterization of what can be done without side information (Theorem 2). Finally, we showed that the obstruction disappears once the learner is given only the optimum same-price value g^*_ν: with this information, a value-auditing strategy recovers the classical best-of-both-worlds rates (Theorem 3).

Our study also leaves several natural limitations. First, while this paper resolves the IV-versus-general bounded-density dichotomy, it does not study best-of-both-worlds guarantees between IV instances and the more general smooth-adversarial models considered in prior work on online bilateral trade [2, 3]. Second, we do not investigate whether similar adaptive guarantees are possible under stricter feedback models, such as the minimal-feedback setting recently studied by [7]. Third, our analysis is tailored to the gain-from-trade objective from [1]; extending the same adaptive perspective to other economically meaningful objectives, such as trading volume [43] and fair gain from trade [5], remains open. These extensions would substantially broaden the scope of adaptive learning in bilateral trade and constitute promising directions for future work.

Acknowledgments and Disclosures

Funding. This work received no external funding. Shuai Li acknowledges support from the National Natural Science Foundation of China (Grant No. 62376154), which did not support this work. **Competing interests.** The authors declare none.

References

[1] Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Bilateral trade: A regret minimization perspective. *Mathematics of Operations Research*, 49(1):171–203, 2024. doi: 10.1287/moor.2023.1351. URL https://pubsonline.informs.org/doi/10.1287/moor.2023.1351.

[2] Nicolò Cesa-Bianchi, Tommaso R. Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Repeated bilateral trade against a smoothed adversary. In *Proceedings of the Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1095–1130. PMLR, 2023. URL https://proceedings.mlr.press/v195/cesa-bianchi23a.html.

[3] Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Regret analysis of bilateral trade with a smoothed adversary. *Journal of Machine Learning Research*, 25(234):1–36, 2024. URL https://jmlr.org/papers/v25/23-1627.html.

[4] Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. No-regret learning in bilateral trade via global budget balance. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 247–258. ACM, 2024. URL https://arxiv.org/abs/2310.12370.

[5] François Bachoc, Nicolò Cesa-Bianchi, Tommaso Cesari, and Roberto Colomboni. Fair online bilateral trade. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://openreview.net/forum?id=I90ypQpLgL&noteId=rlzTXRgHyf.

[6] Solenne Gaucher, Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Vianney Perchet. Feature-based online bilateral trade. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xnF2U0ro7b.

[7] Francesco Bacchiocchi, Matteo Castiglioni, Roberto Colomboni, and Alberto Marchesi. Online bilateral trade with minimal feedback: Don't waste seller's time. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[8] Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 42.1–42.23. PMLR, 2012. URL https://proceedings.mlr.press/v23/bubeck12b.html.

[9] Tor Lattimore. The pareto regret frontier for bandits. In *Advances in Neural Information Processing Systems*, volume 28, pages 208–216. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/hash/6974ce5ac660610b44d9b9fed0ff9548-Abstract.html.

[10] Dylan J. Foster, Akshay Krishnamurthy, and Haipeng Luo. Open problem: Model selection for contextual bandits. In *Proceedings of the Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3842–3846. PMLR, 2020. URL https://proceedings.mlr.press/v125/foster20a.html.

[11] Christoph Dann, Chen-Yu Wei, and Julian Zimmert. A blackbox approach to best of both worlds in bandits and beyond. In *Proceedings of the Thirty-Sixth Annual Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 5503–5570. PMLR, 2023. URL https://proceedings.mlr.press/v195/dann23a.html.

[12] Yuko Kuroki, Alberto Rumi, Taira Tsuchiya, Fabio Vitale, and Nicolò Cesa-Bianchi. Best-of-both-worlds algorithms for linear contextual bandits. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1216–1224. PMLR, 2024. URL https://proceedings.mlr.press/v238/kuroki24a.html.

[13] George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

[14] Darrell Duffie, Nicolae Gârleanu, and Lasse Heje Pedersen. Over-the-counter markets. *Econometrica*, 73(6):1815–1847, 2005.

[15] Paul R. Milgrom and Robert J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, 1982.

[16] Susan Athey and Philip A. Haile. Empirical models of auctions. In Richard Blundell, Whitney K. Newey, and Torsten Persson, editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 2, pages 1–45. Cambridge University Press, 2007.

[17] Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281, 1983.

[18] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory*, pages 174–188. Springer, 2011. doi: 10.1007/978-3-642-24412-4_16.

[19] Yifang Chen, Chung-Wei Lee, Haipeng Luo, and Chen-Yu Wei. A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 696–726. PMLR, 2019. URL https://proceedings.mlr.press/v99/chen19b.html.

[20] Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E. Schapire. Corralling a band of bandit algorithms. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 12–38. PMLR, 2017. URL https://proceedings.mlr.press/v65/agarwal17b.html.

[21] Chen-Yu Wei, Haipeng Luo, and Alekh Agarwal. Taking a hint: How to leverage loss predictors in contextual bandits? In *Proceedings of the Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3583–3634. PMLR, 2020. URL https://proceedings.mlr.press/v125/wei20a.html.

[22] Ashok Cutkosky, Chris Dann, Abhimanyu Das, and Qiuyi Zhang. Leveraging initial hints for free in stochastic linear bandits. In *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pages 282–318. PMLR, 2022. URL https://proceedings.mlr.press/v167/cutkosky22a.html.

[23] Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Bandit online linear optimization with hints and queries. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 2313–2336. PMLR, 2023. URL https://proceedings.mlr.press/v202/bhaskara23a.html.

[24] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1690–1698. PMLR, 2016. URL https://proceedings.mlr.press/v48/locatelli16.html.

[25] Vu Nguyen and Michael A. Osborne. Knowing The What But Not The Where in Bayesian optimization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7317–7326. PMLR, 2020. URL https://proceedings.mlr.press/v119/nguyen20d.html.

[26] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 681–690. ACM, 2008. doi: 10.1145/1374376.1374475.

[27] Aleksandrs Slivkins. Contextual bandits with similarity information. *Journal of Machine Learning Research*, 15(1):2533–2568, 2014. URL http://jmlr.org/papers/v15/slivkins14a.html.

[28] Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 844–853. PMLR, 2017. URL https://proceedings.mlr.press/v70/chowdhury17a.html.

[29] Dylan J. Foster, Alekh Agarwal, Miroslav Dudík, Haipeng Luo, and Robert E. Schapire. Practical contextual bandits with regression oracles. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1539–1548. PMLR, 2018. URL https://proceedings.mlr.press/v80/foster18a.html.

[30] Hédi Hadiji. Polynomial cost of adaptation for x-armed bandits. In *Advances in Neural Information Processing Systems*, volume 32, pages 1027–1036, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/d7a728a67d909e714c0774e22cb806f2-Abstract.html.

[31] Julian Zimmert and Yevgeny Seldin. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021. URL https://www.jmlr.org/papers/v22/19-753.html.

[32] Julian Zimmert, Haipeng Luo, and Chen-Yu Wei. Beating stochastic and adversarial semi-bandits optimally and simultaneously. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7683–7692. PMLR, 2019. URL https://proceedings.mlr.press/v97/zimmert19a.html.

[33] Shinji Ito. Hybrid regret bounds for combinatorial semi-bandits and adversarial linear bandits. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/15a50c8ba6a0002a2fa7e5d8c0a40bd9-Abstract.html.

[34] Fang Kong, Canzhe Zhao, and Shuai Li. Best-of-three-worlds analysis for linear bandits with follow-the-regularized-leader algorithm. In *Proceedings of the Thirty-Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 657–673. PMLR, 2023. URL https://proceedings.mlr.press/v195/kong23a.html.

[35] Shinji Ito, Taira Tsuchiya, and Junya Honda. Adaptive learning rate for follow-the-regularized-leader: Competitive analysis and best-of-both-worlds. In *Proceedings of the Thirty-Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 2522–2563. PMLR, 2024. URL https://proceedings.mlr.press/v247/ito24a.html.

[36] Shinji Ito, Taira Tsuchiya, and Junya Honda. Nearly optimal best-of-both-worlds algorithms for online learning with feedback graphs. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/b7aea253ab34a773967f1e4cdea9e4fb-Abstract-Conference.html.

[37] Fang Kong, Yichi Zhou, and Shuai Li. Simultaneously learning stochastic and adversarial bandits with general graph feedback. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 11473–11482. PMLR, 2022. URL https://proceedings.mlr.press/v162/kong22b.html.

[38] Shinji Ito and Kei Takemura. Best-of-three-worlds linear bandit algorithm with variance-adaptive regret bounds. In *Proceedings of the Thirty-Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 2653–2677. PMLR, 2023. URL https://proceedings.mlr.press/v195/ito23a.html.

[39] Canzhe Zhao, Shinji Ito, and Shuai Li. Heavy-tailed linear bandits: Adversarial robustness, best-of-both-worlds, and beyond. *arXiv preprint arXiv:2508.13679*, 2025. URL https://arxiv.org/abs/2508.13679.

[40] Canzhe Zhao and Vincent Y. F. Tan. Best-of-both-worlds ftrl for linear bandits: The interplay of information acquisition, corruption, and geometry. *arXiv preprint arXiv:2611.01299*, 2026. URL https://arxiv.org/abs/2611.01299.

[41] Ruofeng Yang, Yongcan Li, and Shuai Li. ARIS: Autonomous research via adversarial multi-agent collaboration. *arXiv preprint arXiv:2605.03042*, 2026. URL https://arxiv.org/abs/2605.03042.

[42] Andrea Locatelli and Alexandra Carpentier. Adaptivity to smoothness in x-armed bandits. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1463–1492. PMLR, 2018. URL https://proceedings.mlr.press/v75/locatelli18a.html.

[43] Tommaso Cesari and Roberto Colomboni. An online learning theory of trading-volume maximization. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=OvU9u6wS2J.

A Proof of Theorem 1

We use the stochastic i.i.d. model, the benchmark identity (1), and the class notation from Sections 2 and 3. The proof is built around one explicit product base law ν_0 and a localized family of non-product perturbations $\{\nu_i\}$ with the same marginals as ν_0. The common marginals make one-dimensional evidence uninformative, so identifying which perturbation is present requires the learner to enter a thin informative strip. The strip width is then chosen to balance information against regret.

A.1 The explicit easy base and the hard family

Fix parameters

$$0 < a < u < \tfrac{1}{2}, \qquad u > \tfrac{1}{3}, \qquad m := 4u - 1 \in (u, 1),$$

and then choose constants

$$a < a_- < a_+ < u, \qquad m < c < c + 2\ell < 1.$$

Write

$$L := a_+ - a_- > 0.$$

All constants below depend only on these fixed geometric parameters and on the fixed perturbation amplitude λ introduced later, unless explicitly stated otherwise.

Define the seller marginal density

$$f_S(s) = \frac{1}{u}\mathbf{1}_{[0,u]}(s),$$

so $S \sim \mathrm{Unif}[0, u]$. Define the buyer survival function

$$Q_B(y) := \mathbb{P}(B \geq y) = \begin{cases} 1, & 0 \leq y \leq a, \\ a^2/y^2, & a < y \leq u, \\ a^2/u^2, & u < y \leq m, \\ \dfrac{a^2}{u^2}\dfrac{1-y}{1-m}, & m < y \leq 1, \end{cases}$$

and let $f_B(y) := -Q_B'(y)$. Explicitly,

$$f_B(y) = \begin{cases} 0, & 0 \leq y < a, \\ 2a^2 y^{-3}, & a < y < u, \\ 0, & u < y < m, \\ \dfrac{a^2}{u^2(1-m)}, & m < y < 1. \end{cases}$$

Let

$$\nu_0 := \nu_S \otimes \nu_B$$

be the corresponding independent-values base law, and let $f_0(s, b) = f_S(s) f_B(b)$ denote its joint density.

Lemma 4 (Explicit IV plateau base)**.** *The law ν_0 belongs to $\mathcal{E}_{\mathrm{IV}}(M_0)$ with*

$$M_0 := \max\Big\{\frac{2}{au}, \frac{a^2}{u^3(1-m)}\Big\} < \infty.$$

Moreover,

$$g_0(x) := g_{\nu_0}(x) \equiv g^* := \frac{3a^2}{2u}, \qquad \forall x \in [a, m],$$

and

$$g_0'(x) > 0 \quad \textit{for } x < a, \qquad g_0'(x) < 0 \quad \textit{for } x > m.$$

Hence

$$\operatorname*{argmax}_{x\in[0,1]} g_0(x) = [a, m].$$

Proof. Because $f_0(s, b) = f_S(s)f_B(b)$, the explicit formulas for f_S and f_B give

$$f_0(s, b) \le \max\Big\{\frac{2}{au}, \frac{a^2}{u^3(1-m)}\Big\} = M_0, \qquad \forall (s, b) \in [0, 1]^2.$$

Hence $\nu_0 \in \mathcal{E}_{\mathrm{IV}}(M_0)$.

Under independence and the tail-integration identities

$$\mathbb{E}\big[(x - S)\mathbf{1}\{S \le x\}\big] = \int_0^x F_S(t)\,dt, \qquad \mathbb{E}\big[(B - x)\mathbf{1}\{B \ge x\}\big] = \int_x^1 Q_B(t)\,dt,$$

we have

$$g_0(x) = Q_B(x)\int_0^x F_S(t)\,dt + F_S(x)\int_x^1 Q_B(t)\,dt.$$

Write

$$F_S(x) = \begin{cases} x/u, & x \le u, \\ 1, & x > u, \end{cases} \qquad \Gamma(x) := \int_0^x F_S(t)\,dt = \begin{cases} x^2/(2u), & x \le u, \\ x - u/2, & x \ge u. \end{cases}$$

For $x \in [a, u]$,

$$\int_x^1 Q_B(t)\,dt = \int_x^u \frac{a^2}{t^2}\,dt + \int_u^m \frac{a^2}{u^2}\,dt + \int_m^1 \frac{a^2}{u^2}\frac{1-t}{1-m}\,dt.$$

The first term is $a^2(1/x - 1/u)$, the second is $a^2(m - u)/u^2$, and the third is $a^2(1 - m)/(2u^2)$. Since $m = 4u - 1$, the last two terms sum to a^2/u, hence

$$\int_x^1 Q_B(t)\,dt = \frac{a^2}{x}.$$

Therefore, for $x \in [a, u]$,

$$g_0(x) = \frac{a^2}{x^2}\cdot\frac{x^2}{2u} + \frac{x}{u}\cdot\frac{a^2}{x} = \frac{3a^2}{2u}.$$

For $x \in [u, m]$, we have $F_S(x) = 1$, $\Gamma(x) = x - u/2$, $Q_B(x) = a^2/u^2$, and

$$\int_x^1 Q_B(t)\,dt = \int_x^m \frac{a^2}{u^2}\,dt + \int_m^1 \frac{a^2}{u^2}\frac{1-t}{1-m}\,dt = \frac{a^2}{u^2}(2u - x),$$

again using $m = 4u - 1$. Hence

$$g_0(x) = \frac{a^2}{u^2}\Big(x - \frac{u}{2}\Big) + \frac{a^2}{u^2}(2u - x) = \frac{3a^2}{2u},$$

so g_0 is constant on $[a, m]$.

For $x \in [0, a]$, since $Q_B(x) = 1$ and $\int_a^1 Q_B(t)\,dt = a$, we obtain

$$\int_x^1 Q_B(t)\,dt = (a - x) + a = 2a - x,$$

and therefore

$$g_0(x) = \frac{x^2}{2u} + \frac{x}{u}(2a - x) = \frac{2ax - x^2/2}{u},$$

so

$$g_0'(x) = \frac{2a - x}{u} > 0 \qquad (x < a).$$

For $x > m$, differentiating

$$g_0(x) = Q_B(x)\Gamma(x) + F_S(x) \int_x^1 Q_B(t)\, dt$$

gives

$$g_0'(x) = f_S(x) \int_x^1 Q_B(t)\, dt - f_B(x)\Gamma(x).$$

Since $x > m > u$, we have $f_S(x) = 0$, while $f_B(x) > 0$ and $\Gamma(x) > 0$, hence $g_0'(x) < 0$. The claim follows. □

Set

$$\rho_0 := \frac{a^2}{u^3(1-m)}.$$

On the rectangle $[0, u] \times [c, c + 2\ell]$, the base density equals ρ_0. Fix

$$0 < \lambda < \rho_0/4.$$

We also set

$$M_* := \max\{\sup f_0 + \lambda,\ \|f_S\|_\infty,\ \|f_B\|_\infty\}.$$

Then $\nu_0 \in \mathcal{E}_{\mathrm{IV}}(M_*)$.

For each $\varepsilon > 0$, define

$$K := \left\lfloor \frac{L}{4\varepsilon} \right\rfloor, \qquad v_i := a_- + (4i - 2)\varepsilon, \qquad i = 1, \dots, K,$$

and set

$$I_i := [v_i - \varepsilon, v_i + \varepsilon].$$

Whenever $K \geq 1$, the intervals I_i are pairwise disjoint and satisfy $I_i \subset (a_-, a_+) \subset (a, u)$.

Further define

$$\phi_i(s) := \mathbf{1}_{[v_i - \varepsilon, v_i)}(s) - \mathbf{1}_{[v_i, v_i + \varepsilon)}(s), \qquad \psi(b) := \mathbf{1}_{[c+\ell, c+2\ell)}(b) - \mathbf{1}_{[c, c+\ell)}(b),$$

and let

$$h_i(s, b) := \lambda\, \phi_i(s)\psi(b), \qquad f_i := f_0 + h_i.$$

Denote by ν_i the one-round law with density f_i.

Lemma 5 (Pure copula spike family)**.** *There exists $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0]$ and every $i \in \{1, \dots, K\}$:*

(i) f_i is a valid density and $0 \leq f_i \leq \sup f_0 + \lambda$;

(ii) ν_i has the same seller and buyer marginals as ν_0;

(iii) for any $(p, q) \in \mathcal{U}$, if we define

$$\alpha_i(p) := \int_0^p \phi_i(s)\, ds, \qquad \beta(q) := \int_q^1 \psi(b)\, db,$$

then the four feedback-cell probabilities under ν_i minus those under ν_0, written in the cell order $(11, 10, 01, 00)$, equal

$$\big(+\delta_i(p, q), -\delta_i(p, q), -\delta_i(p, q), +\delta_i(p, q)\big), \qquad \delta_i(p, q) = \lambda\, \alpha_i(p)\beta(q);$$

(iv) the functions α_i and β admit the explicit formulas

$$\alpha_i(p) = \begin{cases} 0, & p \le v_i - \varepsilon, \\ p - (v_i - \varepsilon), & v_i - \varepsilon < p \le v_i, \\ (v_i + \varepsilon) - p, & v_i < p < v_i + \varepsilon, \\ 0, & p \ge v_i + \varepsilon, \end{cases}$$

and

$$\beta(q) = \begin{cases} 0, & q \le c, \\ q - c, & c < q \le c + \ell, \\ c + 2\ell - q, & c + \ell < q < c + 2\ell, \\ 0, & q \ge c + 2\ell. \end{cases}$$

In particular, α_i and β are nonnegative tent functions with

$$\|\alpha_i\|_\infty = \varepsilon, \qquad \|\beta\|_\infty = \ell,$$

α_i has unique maximizer v_i, and

$$\alpha_i(p) = 0 \text{ for all } p \notin I_i, \qquad \beta(q) = 0 \text{ for all } q \notin [c, c + 2\ell].$$

Consequently,

$$\delta_i(p, q) = 0 \qquad \text{whenever } p \notin I_i \text{ or } q \notin [c, c + 2\ell].$$

Consequently, every ν_i lies in $\mathcal{H}(M_*)$, and every ν_i is non-IV (that is, $\nu_i \notin \mathcal{E}_{\mathrm{IV}}(M_*)$).

Proof. We first record the geometry of the intervals I_i. Since

$$v_i - \varepsilon = a_- + (4i - 3)\varepsilon \ge a_- + \varepsilon > a_-,$$

and, because $4K\varepsilon \le L$,

$$v_i + \varepsilon = a_- + (4i - 1)\varepsilon \le a_- + L - \varepsilon = a_+ - \varepsilon < a_+,$$

we indeed have

$$I_i = [v_i - \varepsilon, v_i + \varepsilon] \subset (a_-, a_+) \subset (a, u).$$

In particular, the intervals I_i are pairwise disjoint and lie entirely inside the seller support interval $[0, u]$.

For item (i), note that $\phi_i, \psi \in \{-1, 0, 1\}$, hence

$$|h_i(s, b)| \le \lambda \qquad \text{pointwise.}$$

Moreover,

$$\operatorname{supp}(h_i) \subset I_i \times [c, c + 2\ell] \subset [0, u] \times [c, c + 2\ell].$$

On the rectangle $[0, u] \times [c, c + 2\ell]$, the base density is constant and equal to

$$f_0(s, b) = f_S(s) f_B(b) = \frac{1}{u} \cdot \frac{a^2}{u^2(1 - m)} = \rho_0.$$

Therefore, on the support of h_i,

$$f_i(s, b) = f_0(s, b) + h_i(s, b) \in \{\rho_0 - \lambda, \rho_0, \rho_0 + \lambda\},$$

while off the support of h_i one has $f_i = f_0 \ge 0$. Since $0 < \lambda < \rho_0/4$, it follows that

$$f_i \ge \rho_0 - \lambda > 0 \qquad \text{and} \qquad f_i \le \sup f_0 + \lambda.$$

Also,

$$\iint h_i(s, b)\, ds\, db = \lambda \Big(\int \phi_i(s)\, ds \Big) \Big(\int \psi(b)\, db \Big) = 0,$$

so f_i integrates to one. Hence f_i is a valid density and item (i) follows.

For item (ii),

$$\int h_i(s,b)\,db = \lambda\,\phi_i(s)\int \psi(b)\,db = 0, \qquad \int h_i(s,b)\,ds = \lambda\,\psi(b)\int \phi_i(s)\,ds = 0.$$

Thus the seller and buyer marginals of ν_i agree with those of ν_0.

For item (iii), the perturbation of the $(1,1)$ feedback cell is

$$\delta_i(p,q) = \iint_{[0,p]\times[q,1]} h_i(s,b)\,ds\,db = \lambda\Big(\int_0^p \phi_i(s)\,ds\Big)\Big(\int_q^1 \psi(b)\,db\Big) = \lambda\,\alpha_i(p)\beta(q).$$

Because the marginals do not change, the changes in the four feedback-cell probabilities, in the order $(11,10,01,00)$, are

$$(\delta_i, -\delta_i, -\delta_i, +\delta_i).$$

This proves (iii).

For item (iv), a direct computation gives

$$\alpha_i(p) = \begin{cases} 0, & p \le v_i - \varepsilon, \\ p - (v_i - \varepsilon), & v_i - \varepsilon < p \le v_i, \\ (v_i + \varepsilon) - p, & v_i < p < v_i + \varepsilon, \\ 0, & p \ge v_i + \varepsilon, \end{cases}$$

and

$$\beta(q) = \begin{cases} 0, & q \le c, \\ q - c, & c < q \le c + \ell, \\ c + 2\ell - q, & c + \ell < q < c + 2\ell, \\ 0, & q \ge c + 2\ell. \end{cases}$$

Hence α_i and β are nonnegative tent functions, with

$$\|\alpha_i\|_\infty = \varepsilon, \qquad \|\beta\|_\infty = \ell,$$

α_i has unique maximizer v_i, and

$$\alpha_i(p) = 0 \text{ for all } p \notin I_i, \qquad \beta(q) = 0 \text{ for all } q \notin [c, c+2\ell].$$

Therefore

$$\delta_i(p,q) = \lambda\,\alpha_i(p)\beta(q) = 0 \qquad \text{whenever } p \notin I_i \text{ or } q \notin [c, c+2\ell].$$

Finally, because $f_i \le \sup f_0 + \lambda \le M_*$, every ν_i lies in $\mathcal{H}(M_*)$. Since the marginals of ν_i agree with those of ν_0 but $h_i \not\equiv 0$, one has $f_i \ne f_S f_B$, so ν_i is not a product law and hence $\nu_i \notin \mathcal{E}_{\mathrm{IV}}(M_*)$. □

Define the exploration regions

$$R_i := I_i \times [c, c+2\ell]$$

(which are automatically contained in $\mathcal{U}$ because $I_i \subset (a,u)$ and $c > m > u$), and the candidate regions

$$O_i := \{(p,q) \in \mathcal{U} : p \in I_i,\ q < c\}.$$

Let

$$N_T(i) := \sum_{t=1}^T \mathbf{1}\{(P_t, Q_t) \in R_i\}, \qquad M_T(i) := \sum_{t=1}^T \mathbf{1}\{(P_t, Q_t) \in O_i\},$$

$$N_T := \sum_{t=1}^T \mathbf{1}\{Q_t \in [c, c+2\ell]\}.$$

Because the intervals I_i are disjoint, the sets O_i are disjoint and

$$\sum_{i=1}^K M_T(i) \le T, \qquad \sum_{i=1}^K N_T(i) \le N_T.$$

Lemma 6 (The exploration strip has constant cost under the base)**.** *Let*

$$\Delta_J := g^* - g_0(c) > 0.$$

Then $r_0^* = g_0^* = g^*$, and every action $(p, q) \in \mathcal{U}$ with $q \geq c$ satisfies

$$r_0(p, q) \leq g_0(q) \leq g_0(c) = r_0^* - \Delta_J.$$

Consequently, every round with $Q_t \in [c, c + 2\ell]$ incurs at least Δ_J regret under ν_0, and hence

$$R_T(A; \nu_0) \geq \Delta_J \, \mathbb{E}_0[N_T].$$

In particular, if $R_T(A; \nu_0) \leq C_0 T^\alpha$, then

$$\mathbb{E}_0[N_T] \leq \frac{C_0}{\Delta_J} T^\alpha.$$

Proof. The benchmark equivalence (1) gives $r_0^* = g_0^* = g^*$. If $q \geq c > m$, then Lemma 4 implies that g_0 is strictly decreasing on $[m, 1]$, hence

$$r_0(p, q) \leq g_0(q) \leq g_0(c) = g^* - \Delta_J = r_0^* - \Delta_J.$$

Therefore every round with $Q_t \in [c, c + 2\ell]$ contributes at least Δ_J regret relative to the optimal fixed action, and summing over such rounds proves the claim. □

Lemma 7 (Hard-instance geometry and change of measure)**.** *There exist constants $c_{\mathrm{sub}} > 0$, $C_{\mathrm{KL}} > 0$, $C_{\mathrm{cm}} > 0$, and $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0]$ and every $i \in \{1, \ldots, K\}$ the following hold.*

(i) The unique maximizing same-price action under ν_i is $x = v_i$, and the unique maximizing fixed two-price action is (v_i, v_i). Equivalently,

$$g_i^* = r_i^* = g^* + \lambda \ell^2 \varepsilon.$$

(ii) Every action outside O_i is $\Omega(\varepsilon)$-suboptimal:

$$r_i^* - r_i(p, q) \geq c_{\mathrm{sub}} \, \varepsilon, \qquad \forall (p, q) \in \mathcal{U} \setminus O_i.$$

Consequently,

$$R_T(A; \nu_i) \geq c_{\mathrm{sub}} \, \varepsilon \big(T - \mathbb{E}_i[M_T(i)]\big).$$

(iii) The KL divergence between the full observable histories under ν_0 and ν_i satisfies

$$\mathrm{KL}\big(\mathbb{P}_0^{(T)} \,\|\, \mathbb{P}_i^{(T)}\big) \leq C_{\mathrm{KL}} \, \varepsilon^2 \mathbb{E}_0[N_T(i)].$$

Consequently,

$$\mathbb{E}_i[M_T(i)] - \mathbb{E}_0[M_T(i)] \leq C_{\mathrm{cm}} \, \varepsilon T \sqrt{\mathbb{E}_0[N_T(i)]}.$$

Proof. We write $r_i := r_{\nu_i}$ and $g_i := g_{\nu_i}$.

Step 1: same-price benchmark geometry. For any $x < c$, the perturbation contributes

$$g_i(x) - g_0(x) = \iint (b - s) \mathbf{1}\{s \leq x \leq b\} h_i(s, b) \, ds \, db.$$

Since the support of h_i is contained in $b \in [c, c + 2\ell]$, the condition $x < c$ implies $x \leq b$ automatically on the support. Using also $\int \psi = 0$, we obtain

$$g_i(x) - g_0(x) = \lambda \Big(\int_0^x \phi_i(s) \, ds \Big) \Big(\int b \, \psi(b) \, db \Big) = \lambda \ell^2 \alpha_i(x),$$

where

$$\int b \, \psi(b) \, db = \int_{c+\ell}^{c+2\ell} b \, db - \int_c^{c+\ell} b \, db = \ell^2.$$

Because $I_i \subset (a,u) \subset [a,m]$ and $g_0 \equiv g^*$ on $[a,m]$, the function g_i equals $g^* + \lambda\ell^2\alpha_i(x)$ on I_i and equals $g_0(x)$ off I_i for all $x < c$. Since α_i is a tent function with unique maximizer v_i and peak value ε, we obtain

$$g_i(v_i) = g^* + \lambda\ell^2\varepsilon,$$

and this is the unique maximum of g_i over $[0, c)$.

It remains to rule out maximizers with $x \geq c$. Since the same-price reward satisfies

$$0 \leq (B - S)\mathbf{1}\{S \leq x \leq B\} \leq 1,$$

we have

$$|g_i(x) - g_0(x)| \leq \iint |h_i(s,b)|\, ds\, db = 4\lambda\varepsilon\ell.$$

On the other hand, Lemma 4 implies that for every $x \geq c$,

$$g_0(x) \leq g_0(c) = g^* - \Delta_J.$$

Choose ε_0 small enough that $4\lambda\varepsilon_0\ell \leq \Delta_J/2$. Then, for every $\varepsilon \leq \varepsilon_0$,

$$g_i(x) \leq g^* - \Delta_J + 4\lambda\varepsilon\ell \leq g^* - \Delta_J/2 < g^* + \lambda\ell^2\varepsilon = g_i(v_i), \qquad x \geq c.$$

Therefore $g_i^* = g^* + \lambda\ell^2\varepsilon$, and the unique same-price maximizer is $x = v_i$.

Step 2: gap outside O_i and uniqueness of the optimal fixed two-price action. Suppose first that $q < c$. Then on the support of h_i we have $q \leq b$, so using again $\int \psi = 0$,

$$r_i(p,q) - r_0(p,q) = \lambda\Big(\int_0^p \phi_i(s)\, ds\Big)\Big(\int (b - q)\, \psi(b)\, db\Big).$$

A direct calculation gives

$$\int (b - q)\, \psi(b)\, db = \ell^2,$$

independently of q, hence

$$r_i(p,q) - r_0(p,q) = \lambda\ell^2\alpha_i(p).$$

If $(p,q) \notin O_i$ and $q < c$, then necessarily $p \notin I_i$, so $\alpha_i(p) = 0$ and thus

$$r_i(p,q) = r_0(p,q) \leq r_0^* = g^*.$$

Since $g_i^* = g^* + \lambda\ell^2\varepsilon$, this yields

$$g_i^* - r_i(p,q) \geq \lambda\ell^2\varepsilon.$$

Now suppose $q \geq c$. By Lemma 6,

$$r_0(p,q) \leq r_0^* - \Delta_J = g^* - \Delta_J.$$

Also the two-price reward is pointwise bounded by 1, because on the event $S \leq p \leq q \leq B$ one has

$$0 \leq p - S + B - q \leq p + (1 - q) \leq 1.$$

Therefore

$$|r_i(p,q) - r_0(p,q)| \leq \iint |h_i(s,b)|\, ds\, db = 4\lambda\varepsilon\ell.$$

For $\varepsilon \leq \varepsilon_0$ with $4\lambda\varepsilon_0\ell \leq \Delta_J/2$,

$$g_i^* - r_i(p,q) \geq \Delta_J + \lambda\ell^2\varepsilon - 4\lambda\varepsilon\ell \geq \Delta_J/2.$$

Thus, with

$$c_{\mathrm{sub}} := \min\Big\{\lambda\ell^2, \frac{\Delta_J}{2\varepsilon_0}\Big\},$$

we obtain for all $\varepsilon \leq \varepsilon_0$,

$$g_i^* - r_i(p,q) \geq c_{\mathrm{sub}}\,\varepsilon, \qquad \forall (p,q) \notin O_i.$$

This proves the gap statement in (ii).

To identify the global optimizer, the benchmark equivalence (1) gives

$$r_i(p,q) \le g_i(q) \le g_i^* \qquad \forall (p,q) \in \mathcal{U}.$$

Since v_i is the unique maximizer of g_i, any maximizing fixed two-price action must satisfy $q = v_i$. For this value of q, note that $v_i < u$ and the product formula under ν_0 yields

$$r_0(p, v_i) = Q_B(v_i)\Gamma(p) + F_S(p) \int_{v_i}^{1} Q_B(t)\, dt.$$

The coefficients are strictly positive, while both Γ and F_S are strictly increasing on $[0, u]$; therefore $p \mapsto r_0(p, v_i)$ is strictly increasing on $[0, v_i]$. Also,

$$r_i(p, v_i) = r_0(p, v_i) + \lambda \ell^2 \alpha_i(p).$$

If $p < v_i - \varepsilon$, then $\alpha_i(p) = 0$, so

$$r_i(p, v_i) = r_0(p, v_i) < r_0(v_i, v_i) = g^* < g_i^*.$$

If $p \in [v_i - \varepsilon, v_i)$, then $r_0(p, v_i) < g^*$ and $\alpha_i(p) < \varepsilon$, hence

$$r_i(p, v_i) < g^* + \lambda \ell^2 \varepsilon = g_i^*.$$

Therefore the unique maximizing fixed two-price action is (v_i, v_i), and item (i) follows because (1) already gives $r_i^* = g_i^*$.

To convert the gap estimate into a regret bound, note that every action in O_i has nonnegative regret relative to $r_i^* = g_i^*$, while each action outside O_i contributes at least $c_{\mathrm{sub}}\varepsilon$. Therefore

$$R_T(A; \nu_i) \ge c_{\mathrm{sub}}\, \varepsilon \big(T - \mathbb{E}_i[M_T(i)]\big).$$

Step 3: local KL and transfer. For each round t, let

$$Y_t := (X_t^S, X_t^B) \in \{0,1\}^2.$$

For $t \ge 0$, write

$$H_t := (P_1, Q_1, Y_1, \ldots, P_t, Q_t, Y_t),$$

with H_0 the null history, and for $j \in \{0, i\}$ let $\mathbb{P}_j^{(t)}$ denote the law of H_t under instance P_j. Fix an algorithm A. Conditional on the past history H_{t-1}, the algorithm uses some action kernel

$$K_t(\cdot \mid H_{t-1})$$

on $\mathcal{U}$; this kernel is the same under all instances. Conditional on H_{t-1} and on the chosen action $(P_t, Q_t) = (p, q)$, the law of Y_t under instance P_j is a four-point distribution that we denote by $\pi^j(p, q)$.

By Lemma 5(iv), if $(p, q) \notin R_i$, then $\pi^i(p, q) = \pi^0(p, q)$. Hence only rounds in R_i contribute to the KL divergence.

For $(p, q) \in R_i$, the four cell probabilities under ν_0 are

$$\pi_{11}^0(p,q) = F_S(p)Q_B(q), \qquad \pi_{10}^0(p,q) = F_S(p)\big(1 - Q_B(q)\big),$$
$$\pi_{01}^0(p,q) = \big(1 - F_S(p)\big)Q_B(q), \qquad \pi_{00}^0(p,q) = \big(1 - F_S(p)\big)\big(1 - Q_B(q)\big).$$

Because $p \in I_i \subset (a_-, a_+) \subset (a, u)$ and $q \in [c, c + 2\ell] \subset (m, 1)$, we have

$$F_S(p) = \frac{p}{u} \in \Big[\frac{a_-}{u}, \frac{a_+}{u}\Big], \qquad Q_B(q) = \frac{a^2}{u^2}\frac{1-q}{1-m} \in \Big[\frac{a^2}{u^2}\frac{1-c-2\ell}{1-m}, \frac{a^2}{u^2}\frac{1-c}{1-m}\Big].$$

Define

$$\kappa_S := \min\Big\{\frac{a_-}{u}, 1 - \frac{a_+}{u}\Big\} > 0, \qquad \kappa_B := \min\Big\{\frac{a^2}{u^2}\frac{1-c-2\ell}{1-m}, 1 - \frac{a^2}{u^2}\frac{1-c}{1-m}\Big\} > 0,$$

and set

$$\kappa := \kappa_S \kappa_B.$$

Then, for every $(p, q) \in R_i$,

$$\pi^0_{11}(p,q),\ \pi^0_{10}(p,q),\ \pi^0_{01}(p,q),\ \pi^0_{00}(p,q) \geq \kappa.$$

Moreover, by Lemma 5(iii)–(iv), the perturbation magnitude satisfies

$$|\delta_i(p,q)| \leq \lambda \|\alpha_i\|_\infty \|\beta\|_\infty \leq \lambda \varepsilon \ell.$$

Shrinking ε_0 if needed, we may assume $\lambda \varepsilon_0 \ell \leq \kappa/2$. Then for every $(p, q) \in R_i$,

$$\pi^i(p,q) = (\pi^0_{11} + \delta_i,\ \pi^0_{10} - \delta_i,\ \pi^0_{01} - \delta_i,\ \pi^0_{00} + \delta_i),$$

with all coordinates at least $\kappa/2$. Using the standard bound $\mathrm{KL}(\mu\|\nu) \leq \sum_z (\mu_z - \nu_z)^2/\nu_z$ for finite distributions, we get

$$\mathrm{KL}\big(\pi^0(p,q) \,\|\, \pi^i(p,q)\big) \leq \sum_{z \in \{0,1\}^2} \frac{(\pi^0_z - \pi^i_z)^2}{\pi^i_z} \leq \frac{8}{\kappa}\, \delta_i(p,q)^2 \leq \frac{8\lambda^2 \ell^2}{\kappa}\, \varepsilon^2.$$

Therefore there exists $C_{\mathrm{KL}} > 0$ such that

$$\mathrm{KL}\big(\pi^0(p,q) \,\|\, \pi^i(p,q)\big) \leq C_{\mathrm{KL}} \varepsilon^2 \mathbf{1}\{(p,q) \in R_i\}.$$

By the chain rule for KL under adaptive interaction, and because the action kernels K_t are identical under ν_0 and ν_i,

$$\mathrm{KL}\big(\mathbb{P}_0^{(T)} \,\|\, \mathbb{P}_i^{(T)}\big) = \sum_{t=1}^{T} \mathbb{E}_0\Big[\mathrm{KL}\big(\pi^0(P_t,Q_t) \,\|\, \pi^i(P_t,Q_t)\big)\Big] \leq C_{\mathrm{KL}} \varepsilon^2 \mathbb{E}_0[N_T(i)].$$

To compare the candidate counts, note that for each t,

$$\mathbb{P}_i\big((P_t,Q_t) \in O_i\big) - \mathbb{P}_0\big((P_t,Q_t) \in O_i\big) = \int K_t(O_i \mid h)\, d\big(\mathbb{P}_i^{(t-1)} - \mathbb{P}_0^{(t-1)}\big)(h).$$

Since $0 \leq K_t(O_i \mid h) \leq 1$, this implies

$$\mathbb{P}_i\big((P_t,Q_t) \in O_i\big) - \mathbb{P}_0\big((P_t,Q_t) \in O_i\big) \leq \mathrm{TV}\big(\mathbb{P}_i^{(t-1)}, \mathbb{P}_0^{(t-1)}\big) \leq \sqrt{\frac{1}{2}\mathrm{KL}\big(\mathbb{P}_0^{(t-1)} \,\|\, \mathbb{P}_i^{(t-1)}\big)}$$

by Pinsker's inequality. Since $N_{t-1}(i) \leq N_T(i)$ pointwise, the KL bound above yields

$$\mathbb{P}_i\big((P_t,Q_t) \in O_i\big) - \mathbb{P}_0\big((P_t,Q_t) \in O_i\big) \leq \sqrt{\frac{C_{\mathrm{KL}}}{2}}\, \varepsilon \sqrt{\mathbb{E}_0[N_T(i)]}.$$

Summing over $t = 1, \ldots, T$ gives

$$\mathbb{E}_i[M_T(i)] - \mathbb{E}_0[M_T(i)] \leq C_{\mathrm{cm}}\, \varepsilon T \sqrt{\mathbb{E}_0[N_T(i)]}$$

for some constant $C_{\mathrm{cm}} > 0$. □

A.2 Completion of the no-go proof

Proposition 1 (Fixed-base witness theorem)**.** *Fix an exponent $\alpha \in [0, 1]$. For the fixed density level M_* and every constant $C_0 > 0$, there exists a constant $c_{\alpha,C_0} > 0$ such that the following holds for every horizon $T \geq 1$ and every algorithm A: if*

$$R_T(A; \nu_0) \leq C_0 T^\alpha,$$

then there exists a witness instance $\nu^{(T)} \in \mathcal{H}(M_*) \setminus \mathcal{E}_{\mathrm{IV}}(M_*)$, possibly depending on T, such that

$$R_T(A; \nu^{(T)}) \geq c_{\alpha,C_0} T^{1-\alpha/3}.$$

Each such witness is i.i.d., has the same marginals as ν_0, and has joint density bounded by M_*.

Proof. Fix $\alpha \in [0,1]$ and $C_0 > 0$. Choose

$$\varepsilon := \eta T^{-\alpha/3}$$

with a small constant $\eta > 0$ to be specified, and let $P_1, \dots, P_K$ be the corresponding hard family constructed above.

Assume that

$$R_T(A; \nu_0) \le C_0 T^\alpha.$$

By Lemma 6,

$$\mathbb{E}_0[N_T] \le \frac{C_0}{\Delta_J} T^\alpha.$$

Since $\sum_{i=1}^K N_T(i) \le N_T$, we get

$$\sum_{i=1}^K \mathbb{E}_0[N_T(i)] \le \frac{C_0}{\Delta_J} T^\alpha.$$

Next, Lemma 7 implies that for every i,

$$R_T(A; \nu_i) \ge c_{\mathrm{sub}}\, \varepsilon \big(T - \mathbb{E}_i[M_T(i)]\big).$$

Averaging over $i = 1, \dots, K$ and applying the transfer bound from Lemma 7,

$$\begin{aligned}
\frac{1}{K} \sum_{i=1}^K R_T(A; \nu_i) &\ge c_{\mathrm{sub}}\, \varepsilon \left(T - \frac{1}{K} \sum_{i=1}^K \mathbb{E}_i[M_T(i)] \right) \\
&\ge c_{\mathrm{sub}}\, \varepsilon \left(T - \frac{1}{K} \sum_{i=1}^K \mathbb{E}_0[M_T(i)] - \frac{C_{\mathrm{cm}} \varepsilon T}{K} \sum_{i=1}^K \sqrt{\mathbb{E}_0[N_T(i)]} \right).
\end{aligned}$$

Because the O_i are disjoint,

$$\sum_{i=1}^K M_T(i) \le T, \qquad \text{hence} \qquad \frac{1}{K} \sum_{i=1}^K \mathbb{E}_0[M_T(i)] \le \frac{T}{K}.$$

Also, by Jensen's inequality for the concave function $x \mapsto \sqrt{x}$,

$$\frac{1}{K} \sum_{i=1}^K \sqrt{\mathbb{E}_0[N_T(i)]} \le \sqrt{\frac{1}{K} \sum_{i=1}^K \mathbb{E}_0[N_T(i)]} \le \sqrt{\frac{C_0}{\Delta_J}}\, \frac{T^{\alpha/2}}{\sqrt{K}}.$$

Therefore,

$$\frac{1}{K} \sum_{i=1}^K R_T(A; \nu_i) \ge c_{\mathrm{sub}}\, \varepsilon T \left(1 - \frac{1}{K} - C_{\mathrm{cm}} \sqrt{\frac{C_0}{\Delta_J}}\, \varepsilon \frac{T^{\alpha/2}}{\sqrt{K}} \right).$$

For every $\varepsilon \le L/8$, the definition of K gives

$$\frac{L}{8\varepsilon} \le K \le \frac{L}{4\varepsilon}.$$

Indeed, if $x := L/(4\varepsilon)$, then $x \ge 2$ and therefore $\lfloor x \rfloor \ge x/2 = L/(8\varepsilon)$; the upper bound is immediate from $\lfloor x \rfloor \le x$. Hence

$$\frac{1}{K} \le \frac{8}{L} \varepsilon = \frac{8\eta}{L} T^{-\alpha/3} \le \frac{8\eta}{L},$$

and

$$\varepsilon \frac{T^{\alpha/2}}{\sqrt{K}} \le \sqrt{\frac{8}{L}}\, \varepsilon^{3/2} T^{\alpha/2} = \sqrt{\frac{8}{L}}\, \eta^{3/2}.$$

Choose $\eta > 0$ sufficiently small so that

$$\eta \le \min\Big\{ \varepsilon_0, \frac{L}{8} \Big\}, \qquad \frac{8\eta}{L} \le \frac{1}{4}, \qquad C_{\mathrm{cm}} \sqrt{\frac{C_0}{\Delta_J}} \sqrt{\frac{8}{L}}\, \eta^{3/2} \le \frac{1}{4}.$$

Then the bracketed factor is at least $1/2$, and therefore

$$\frac{1}{K}\sum_{i=1}^{K} R_T(A;\nu_i) \geq \frac{c_{\mathrm{sub}}}{2}\,\varepsilon T = \frac{c_{\mathrm{sub}}\eta}{2} T^{1-\alpha/3}.$$

Hence at least one index i satisfies

$$R_T(A;\nu_i) \geq c_{\alpha,C_0} T^{1-\alpha/3}, \qquad c_{\alpha,C_0} := \frac{c_{\mathrm{sub}}\eta}{2}.$$

Setting $\nu^{(T)} := \nu_i$ completes the proof. □

Proof of Theorem 1. Let ν_0 be the explicit product base law constructed above, and let $M_* > 0$ be the corresponding density level. By Lemma 4, we have $\nu_0 \in \mathcal{E}_{\mathrm{IV}}(M_*)$.

Fix $\alpha \in [0,1]$ and an algorithm A. Assume that

$$\sup_{\nu \in \mathcal{E}_{\mathrm{IV}}(M_*)} R_T(A;\nu) = O(T^\alpha).$$

Then there exists a constant $C_0 > 0$ such that

$$R_T(A;\nu_0) \leq C_0 T^\alpha$$

for all sufficiently large T. Proposition 1 therefore yields, for all sufficiently large T, a witness instance

$$\nu^{(T)} \in \mathcal{H}(M_*) \setminus \mathcal{E}_{\mathrm{IV}}(M_*)$$

such that

$$R_T(A;\nu^{(T)}) \geq c_{\alpha,C_0} T^{1-\alpha/3}.$$

Hence

$$\sup_{\mu \in \mathcal{H}(M_*)\setminus\mathcal{E}_{\mathrm{IV}}(M_*)} R_T(A;\mu) = \Omega\big(T^{1-\alpha/3}\big).$$

Taking $\alpha = 2/3$ gives the endpoint obstruction

$$\sup_{\mu \in \mathcal{H}(M_*)\setminus\mathcal{E}_{\mathrm{IV}}(M_*)} R_T(A;\mu) = \Omega(T^{7/9}),$$

which rules out the pair $(2/3, 3/4)$. □

B Proof of Theorem 2

The proof has three steps. We first show that the diagnostic batches estimate both the true same-price curve and its product surrogate uniformly on the grid. We then prove that, on the accept branch, temporal decoupling produces an exact fresh run of the IV learner against the product law $\nu^{\otimes}$. Finally, we combine these two ingredients with a simple grid-commit argument for the reject branch.

We use the model, benchmark identity, class notation, and regret notation from Sections 2 and 4. In particular, $\mathfrak{R}_T(A;\nu)$ denotes the realized regret random variable from Section 2. The only new objects introduced in this appendix are the product surrogate $\tilde{g}_\nu(x) := g_{\nu^\otimes}(x)$, its mismatch function $\Delta_\nu := g_\nu - \tilde{g}_\nu$, and the shorthand $\tilde{g}^*_\nu := \sup_{x\in[0,1]} \tilde{g}_\nu(x)$.

B.1 The product surrogate

Fix a one-round law ν on $[0,1]^2$. Write

$$F_\nu(x) := \mathbb{P}_\nu(S \leq x), \qquad \overline{F}_\nu(x) := \mathbb{P}_\nu(B \geq x),$$

$$\Gamma_\nu(x) := \int_0^x F_\nu(p)\,dp, \qquad H_\nu(x) := \int_x^1 \overline{F}_\nu(q)\,dq.$$

Let $\nu^\otimes := \nu_S \otimes \nu_B$ be the product of the seller and buyer marginals. Define the product surrogate

$$\tilde{g}_\nu(x) := g_{\nu^\otimes}(x).$$

Thus $\tilde{g}_\nu = g_\nu$ whenever $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$.

We will use the bivariate tail function

$$C_\nu(p,q) := \mathbb{P}_\nu(S \leq p,\ B \geq q), \qquad p \leq q.$$

Lemma 8 (Integral identities)**.** *For every law ν and every $x \in [0,1]$,*

$$g_\nu(x) = \int_0^x C_\nu(p,x)\,dp + \int_x^1 C_\nu(x,q)\,dq.$$

Moreover,

$$\tilde{g}_\nu(x) = \overline{F}_\nu(x)\Gamma_\nu(x) + F_\nu(x)H_\nu(x).$$

Proof. For every realization (S,B),

$$\begin{aligned}(B-S)\mathbf{1}\{S \le x \le B\} &= \left(\int_S^x dp + \int_x^B dq\right)\mathbf{1}\{S \le x \le B\} \\ &= \int_0^x \mathbf{1}\{S \le p,\ B \ge x\}\,dp + \int_x^1 \mathbf{1}\{S \le x,\ B \ge q\}\,dq.\end{aligned}$$

Taking expectations yields the first identity. For the second identity, apply the first identity to the product law $\nu^{\otimes}$, for which

$$C_{\nu^{\otimes}}(p,q) = F_\nu(p)\overline{F}_\nu(q).$$

Then

$$\tilde{g}_\nu(x) = \overline{F}_\nu(x)\int_0^x F_\nu(p)\,dp + F_\nu(x)\int_x^1 \overline{F}_\nu(q)\,dq. \qquad \square$$

Lemma 9 (Lipschitz control)**.** *Let $M > 0$ and let $\nu \in \mathcal{H}(M)$. Then the marginal tails F_ν and $\overline{F}_\nu$ are M-Lipschitz on $[0,1]$, the integral terms Γ_ν and H_ν are 1-Lipschitz, g_ν is $2M$-Lipschitz, and $\tilde{g}_\nu$ is $2(M+1)$-Lipschitz. Consequently,*

$$\Delta_\nu(x) := g_\nu(x) - \tilde{g}_\nu(x)$$

is L_M-Lipschitz with

$$L_M := 4M + 2.$$

Proof. Since $\nu \in \mathcal{H}(M)$, its marginal densities are bounded by M. Hence for $0 \le x < y \le 1$,

$$\begin{aligned}|F_\nu(y) - F_\nu(x)| &= \mathbb{P}_\nu(S \in [x,y]) \le M(y-x), \\ |\overline{F}_\nu(y) - \overline{F}_\nu(x)| &= \mathbb{P}_\nu(B \in [x,y]) \le M(y-x).\end{aligned}$$

Also $F_\nu, \overline{F}_\nu \in [0,1]$, so

$$\begin{aligned}|\Gamma_\nu(y) - \Gamma_\nu(x)| &= \left|\int_x^y F_\nu(p)\,dp\right| \le y - x, \\ |H_\nu(y) - H_\nu(x)| &= \left|\int_x^y \overline{F}_\nu(q)\,dq\right| \le y - x.\end{aligned}$$

For g_ν, the events $\{S \le x \le B\}$ and $\{S \le y \le B\}$ differ only when $S \in [x,y]$ or $B \in [x,y]$. Since $0 \le B - S \le 1$ on their union,

$$|g_\nu(y) - g_\nu(x)| \le \mathbb{P}_\nu(S \in [x,y]) + \mathbb{P}_\nu(B \in [x,y]) \le 2M(y-x).$$

For the surrogate, use

$$\tilde{g}_\nu(x) = \overline{F}_\nu(x)\Gamma_\nu(x) + F_\nu(x)H_\nu(x)$$

and the product rule together with the bounds above:

$$\begin{aligned}|\overline{F}_\nu(y)\Gamma_\nu(y) - \overline{F}_\nu(x)\Gamma_\nu(x)| &\le |\overline{F}_\nu(y) - \overline{F}_\nu(x)|\,|\Gamma_\nu(y)| + \overline{F}_\nu(x)|\Gamma_\nu(y) - \Gamma_\nu(x)| \\ &\le (M+1)(y-x),\end{aligned}$$

and the same estimate holds for the term $F_\nu H_\nu$. Thus

$$|\tilde{g}_\nu(y) - \tilde{g}_\nu(x)| \le 2(M+1)(y-x).$$

Subtracting the last two bounds gives the claim for Δ_ν. $\square$

B.2 The IV subroutine and its interface

We will use the independent-values algorithm of [1]. The point that matters for us is not merely the regret rate but also the *interaction pattern*: after the accept decision, we restart a *fresh* horizon-H IV algorithm and interact with it only through streaming two-bit feedback. In particular, $\mathcal{L}_{U,H}$ denotes the full IV routine together with whatever internal state it chooses to maintain. To make this explicit, we encode any randomized learner by a seed and measurable update maps.

Definition 10 (Causal same-price learner)**.** Fix a horizon $H \geq 1$. A randomized same-price learner of horizon H is called *causal* if there exist a measurable seed space W, an internal random seed $W \in \mathsf{W}$ independent of the environment, and measurable maps

$$\pi_k : \mathsf{W} \times (\{0,1\}^2)^{k-1} \to [0,1], \qquad k = 1, \dots, H,$$

such that, when the learner is run for H virtual rounds, its round-k price is

$$X_k = \pi_k\big(W, Z_1, \dots, Z_{k-1}\big), \qquad Z_\ell \in \{0,1\}^2.$$

Theorem 11 (Concrete IV subroutine)**.** *For every density parameter $U > 0$ and every horizon $H \geq 1$, there exists a causal same-price learner $\mathcal{L}_{U,H}$ and a constant $C_{\mathrm{IV}}(U)$ such that, for every product law $\mu \in \mathcal{E}_{\mathrm{IV}}(U)$,*

$$H g^*_\mu - \mathbb{E}_\mu\left[\sum_{k=1}^{H} g_\mu(X_k)\right] \leq C_{\mathrm{IV}}(U) H^{2/3}.$$

The learner uses only the usual two-bit feedback $\big(\mathbf{1}\{S \leq X_k\}, \mathbf{1}\{B \geq X_k\}\big)$.

Algorithm 3 Interface of the IV learner $\mathcal{L}_{U,H}$

1: **Init:** sample a fresh seed W and set the internal bit history to empty.
2: **NextPrice:** on virtual round k, return $X_k = \pi_k(W, Z_1, \dots, Z_{k-1})$.
3: **Update:** after receiving $Z_k \in \{0,1\}^2$, append Z_k to the bit history.

B.3 The Surrogate-Consistency Meta-Strategy

We then analyze the main-text meta-strategy in Algorithm 1. Fix $M > 0$, a horizon T, and a resolution exponent $\gamma \in (0, 1/3)$. Set

$$\varepsilon := T^{-\gamma}, \qquad K := \lceil \varepsilon^{-1} \rceil, \qquad x_j := \frac{j}{K}, \quad j = 0, 1, \dots, K,$$

$$m := \left\lceil \frac{\log(24KT^2)}{2\varepsilon^2} \right\rceil, \qquad N_0 := 2m(K-1).$$

Since $\gamma < 1/3$, there exists an integer $T_{\mathrm{sm}}(\gamma) \geq 3$ such that, for every $T \geq T_{\mathrm{sm}}(\gamma)$, one has $K \geq 2$ and $N_0 \leq T/2$. This threshold is used only to isolate the asymptotically nontrivial regime in the analysis below. In particular, for $T \geq T_{\mathrm{sm}}(\gamma)$ we have $H := \lfloor (T - N_0)/2 \rfloor \geq 1$. Moreover, if $\nu \in \mathcal{H}(M)$ then both marginals of ν are bounded by M, so the product law $\nu^{\otimes}$ belongs to $\mathcal{E}_{\mathrm{IV}}(M^2)$. This is why the accept branch always restarts the IV learner with density parameter M^2.

The diagnostic batch below also builds on the left/right probing template from the smoothed-adversary bilateral-trade line [2, 3], but repurposes it to estimate simultaneously the true same-price curve and the product surrogate at a diagnostic grid point. In the earlier fixed-regime use of that probe, one estimates a same-price value. Here the same randomization has to support a more delicate dual task: the products of the two bits recover the joint-tail terms needed for g_ν, while the single bits recover the marginals needed for $\tilde{g}_\nu$.

Algorithm 4 BATCHPROBE(x, m): one diagnostic probe batch at price x

for $\ell = 1$ to m **do**
  Sample $P_\ell \sim \mathrm{Unif}[0, x]$, play (P_ℓ, x), and set $(X_\ell^{S,L}, X_\ell^{B,L}) := (\mathbf{1}\{S \le P_\ell\}, \mathbf{1}\{B \ge x\})$.
  Sample $Q_\ell \sim \mathrm{Unif}[x, 1]$, play (x, Q_ℓ), and set $(X_\ell^{S,R}, X_\ell^{B,R}) := (\mathbf{1}\{S \le x\}, \mathbf{1}\{B \ge Q_\ell\})$.
end for
Form the linear summaries
  $\widehat{A} := x\frac{1}{m}\sum_{\ell=1}^m X_\ell^{S,L} X_\ell^{B,L}$, $\widehat{B} := (1-x)\frac{1}{m}\sum_{\ell=1}^m X_\ell^{S,R} X_\ell^{B,R}$,
  $\widehat{\Gamma} := x\frac{1}{m}\sum_{\ell=1}^m X_\ell^{S,L}$, $\widehat{\overline{F}} := \frac{1}{m}\sum_{\ell=1}^m X_\ell^{B,L}$, $\widehat{F} := \frac{1}{m}\sum_{\ell=1}^m X_\ell^{S,R}$, $\widehat{H} := (1-x)\frac{1}{m}\sum_{\ell=1}^m X_\ell^{B,R}$.
Set $\widehat{g} := \widehat{A} + \widehat{B}$ and $\widetilde{g} := \widehat{\overline{F}}\widehat{\Gamma} + \widehat{F}\widehat{H}$.
Return $(\widehat{A}, \widehat{B}, \widehat{\Gamma}, \widehat{\overline{F}}, \widehat{F}, \widehat{H}, \widehat{g}, \widetilde{g})$.

For each interior grid point x_j, Stage I of Algorithm 1 calls Algorithm 4. For the later concentration lemmas we denote the raw products by

$$Y_{j,\ell}^L := X_{j,\ell}^{S,L} X_{j,\ell}^{B,L}, \qquad Y_{j,\ell}^R := X_{j,\ell}^{S,R} X_{j,\ell}^{B,R},$$

and write the returned summaries as

$$(\widehat{A}_j, \widehat{B}_j, \widehat{\Gamma}_j, \widehat{\overline{F}}_j, \widehat{F}_j, \widehat{H}_j, \widehat{g}_j, \widetilde{g}_j).$$

This is precisely the point at which the inherited probe becomes a new certificate. The quantities $\widehat{g}_j$ and $\widetilde{g}_j$ are not two unrelated estimates collected side by side. They are built from the same left/right batches so that one concentration event simultaneously controls the true same-price curve and its product surrogate on the whole grid, which is what later allows a single diagnostic to justify both the accept and reject branches.

Set $\widehat{g}_0 = \widehat{g}_K = \widetilde{g}_0 = \widetilde{g}_K = 0$ and define

$$D_{\mathrm{test}} := \max_{1 \le j \le K-1} |\widehat{g}_j - \widetilde{g}_j|, \qquad \mathsf{Acc} := \{D_{\mathrm{test}} \le 8\varepsilon\}, \qquad \mathsf{Rej} := \mathsf{Acc}^c.$$

On Acc, Algorithm 1 starts a fresh copy of Algorithm 3 with density parameter M^2 on budget

$$H := \Big\lfloor \frac{T - N_0}{2} \Big\rfloor,$$

plays (X_k, X_k) on each real pair of rounds $(N_0 + 2k - 1, N_0 + 2k)$, and returns the temporally decoupled synthetic feedback

$$\widetilde{Z}_k := \big(\mathbf{1}\{S_{N_0+2k-1} \le X_k\},\ \mathbf{1}\{B_{N_0+2k} \ge X_k\}\big).$$

If $T - N_0$ is odd, the single remaining real round after these H emulated pairs is played at an arbitrary fixed same price; the proof below charges it regret at most 1. On Rej, it commits to the empirical grid maximizer

$$\widehat{x} := x_{\widehat{j}}, \qquad \widehat{j} \in \arg\max_{0 \le j \le K} \widehat{g}_j.$$

B.4 What the diagnostic stage estimates

Lemma 12 (The diagnostic batches estimate both objectives)**.** *Fix any i.i.d. law ν on $[0,1]^2$ and $j \in \{1, \dots, K-1\}$. Let*

$$A_j := \int_0^{x_j} C_\nu(p, x_j)\, dp, \qquad B_j := \int_{x_j}^1 C_\nu(x_j, q)\, dq.$$

Then

$$\mathbb{E}_\nu[\widehat{A}_j] = A_j, \qquad \mathbb{E}_\nu[\widehat{B}_j] = B_j, \qquad \mathbb{E}_\nu[\widehat{g}_j] = g_\nu(x_j).$$

Also,

$$\mathbb{E}_\nu[\widehat{\Gamma}_j] = \Gamma_\nu(x_j), \quad \mathbb{E}_\nu[\widehat{\overline{F}}_j] = \overline{F}_\nu(x_j), \quad \mathbb{E}_\nu[\widehat{F}_j] = F_\nu(x_j), \quad \mathbb{E}_\nu[\widehat{H}_j] = H_\nu(x_j).$$

Proof. For the left batch,

$$\mathbb{E}_\nu[\widehat{A}_j] = x_j \cdot \mathbb{E}_\nu[Y^L_{j,1}] = x_j \cdot \frac{1}{x_j}\int_0^{x_j} \mathbb{P}_\nu(S \le p,\ B \ge x_j)\,dp = A_j.$$

Likewise,

$$\mathbb{E}_\nu[\widehat{\Gamma}_j] = x_j \cdot \mathbb{E}_\nu[X^{S,L}_{j,1}] = \int_0^{x_j} F_\nu(p)\,dp = \Gamma_\nu(x_j),$$

and

$$\mathbb{E}_\nu[\widehat{\overline{F}}_j] = \mathbb{E}_\nu[X^{B,L}_{j,1}] = \overline{F}_\nu(x_j).$$

For the right batch,

$$\mathbb{E}_\nu[\widehat{B}_j] = (1-x_j)\cdot \mathbb{E}_\nu[Y^R_{j,1}] = (1-x_j)\cdot \frac{1}{1-x_j}\int_{x_j}^1 \mathbb{P}_\nu(S \le x_j,\ B \ge q)\,dq = B_j,$$

while

$$\mathbb{E}_\nu[\widehat{F}_j] = F_\nu(x_j), \qquad \mathbb{E}_\nu[\widehat{H}_j] = (1-x_j)\cdot \mathbb{E}_\nu[X^{B,R}_{j,1}] = \int_{x_j}^1 \overline{F}_\nu(q)\,dq.$$

The identity $\mathbb{E}_\nu[\widehat{g}_j] = g_\nu(x_j)$ follows from Lemma 8. □

Lemma 13 (Uniform concentration event)**.** *Fix any i.i.d. law ν on $[0,1]^2$. Let $\mathcal{E}$ be the event that, simultaneously for every $j = 1, \dots, K-1$,*

$$|\widehat{A}_j - A_j| \le \varepsilon, \quad |\widehat{B}_j - B_j| \le \varepsilon, \quad |\widehat{\Gamma}_j - \Gamma_\nu(x_j)| \le \varepsilon, \quad |\widehat{\overline{F}}_j - \overline{F}_\nu(x_j)| \le \varepsilon,$$
$$|\widehat{F}_j - F_\nu(x_j)| \le \varepsilon, \qquad |\widehat{H}_j - H_\nu(x_j)| \le \varepsilon.$$

Then

$$\mathbb{P}_\nu(\mathcal{E}) \ge 1 - \frac{1}{2T^2}.$$

On $\mathcal{E}$, simultaneously for every $j = 1, \dots, K-1$,

$$|\widehat{g}_j - g_\nu(x_j)| \le 2\varepsilon, \qquad |\widetilde{g}_j - \tilde{g}_\nu(x_j)| \le 4\varepsilon.$$

Proof. Fix j. The stage-I probe rounds used for this j are independent, and the auxiliary price draws are also independent. Hence the following six collections are i.i.d.:

$$\xi^A_{j,\ell} := x_j Y^L_{j,\ell} \in [0,1], \qquad \xi^B_{j,\ell} := (1-x_j) Y^R_{j,\ell} \in [0,1],$$
$$\xi^\Gamma_{j,\ell} := x_j X^{S,L}_{j,\ell} \in [0,1], \qquad \xi^{\overline{F}}_{j,\ell} := X^{B,L}_{j,\ell} \in [0,1],$$
$$\xi^F_{j,\ell} := X^{S,R}_{j,\ell} \in [0,1], \qquad \xi^H_{j,\ell} := (1-x_j) X^{B,R}_{j,\ell} \in [0,1].$$

Moreover,

$$\widehat{A}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^A_{j,\ell}, \quad \widehat{B}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^B_{j,\ell}, \quad \widehat{\Gamma}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^\Gamma_{j,\ell},$$
$$\widehat{\overline{F}}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^{\overline{F}}_{j,\ell}, \quad \widehat{F}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^F_{j,\ell}, \quad \widehat{H}_j = \frac{1}{m}\sum_{\ell=1}^m \xi^H_{j,\ell}.$$

Therefore Hoeffding's inequality gives, for any one of these six estimators,

$$\mathbb{P}_\nu\big(|\widehat{\theta} - \theta| > \varepsilon\big) \le 2e^{-2m\varepsilon^2} \le \frac{1}{12KT^2}$$

by the definition of m. A union bound over at most $6(K-1) < 6K$ estimates proves $\mathbb{P}_\nu(\mathcal{E}) \ge 1 - 1/(2T^2)$.

On $\mathcal{E}$,

$$|\widehat{g}_j - g_\nu(x_j)| \le |\widehat{A}_j - A_j| + |\widehat{B}_j - B_j| \le 2\varepsilon.$$

Also, since every factor lies in $[0,1]$,

$$|\widetilde{g}_j - \tilde{g}_\nu(x_j)| = \big|\widehat{\overline{F}}_j\widehat{\Gamma}_j - \overline{F}_\nu(x_j)\Gamma_\nu(x_j) + \widehat{F}_j\widehat{H}_j - F_\nu(x_j)H_\nu(x_j)\big|$$

$$
\begin{aligned}
&\le |\widehat{\overline{F}}_j\widehat{\Gamma}_j - \overline{F}_\nu(x_j)\widehat{\Gamma}_j| + |\overline{F}_\nu(x_j)\widehat{\Gamma}_j - \overline{F}_\nu(x_j)\Gamma_\nu(x_j)| \\
&\quad + |\widehat{F}_j\widehat{H}_j - F_\nu(x_j)\widehat{H}_j| + |F_\nu(x_j)\widehat{H}_j - F_\nu(x_j)H_\nu(x_j)| \\
&\le |\widehat{\overline{F}}_j - \overline{F}_\nu(x_j)| + |\widehat{\Gamma}_j - \Gamma_\nu(x_j)| \\
&\quad + |\widehat{F}_j - F_\nu(x_j)| + |\widehat{H}_j - H_\nu(x_j)| \\
&\le 4\varepsilon.
\end{aligned}
$$

□

Corollary 1 (What the test guarantees)**.** *Assume $\mathcal{E}$.*

(i) If $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$, then the test accepts.

(ii) If $\nu \in \mathcal{H}(M)$ and the test accepts, then

$$
D_\nu := \sup_{x\in[0,1]} |\Delta_\nu(x)| \le C_{\mathrm{test}}(M)\varepsilon, \qquad C_{\mathrm{test}}(M) := 14 + L_M.
$$

Proof. If $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$, then $g_\nu = \tilde{g}_\nu$, so by Lemma 13,

$$
D_{\mathrm{test}} \le \max_{1\le j\le K-1} \big(|\widehat{g}_j - g_\nu(x_j)| + |\widetilde{g}_j - \tilde{g}_\nu(x_j)|\big) \le 6\varepsilon < 8\varepsilon.
$$

Hence the test accepts.

Now suppose the test accepts under an arbitrary $\nu \in \mathcal{H}(M)$. Then for every j,

$$
\begin{aligned}
|\Delta_\nu(x_j)| &\le |g_\nu(x_j) - \widehat{g}_j| + |\widehat{g}_j - \widetilde{g}_j| + |\widetilde{g}_j - \tilde{g}_\nu(x_j)| \\
&\le 2\varepsilon + 8\varepsilon + 4\varepsilon = 14\varepsilon.
\end{aligned}
$$

Since the grid mesh is at most ε and Δ_ν is L_M-Lipschitz by Lemma 9, every $x \in [0,1]$ has a grid point x_j with $|x - x_j| \le \varepsilon$, so

$$
|\Delta_\nu(x)| \le |\Delta_\nu(x_j)| + L_M|x - x_j| \le (14 + L_M)\varepsilon.
$$

□

B.5 Temporal decoupling as a generic interface

Let $\mathcal{F}_0$ be the sigma-field generated by the entire diagnostic stage: all stage-I random choices, the first N_0 environment draws, and the resulting observations. In the accept branch, a fresh copy of the causal learner $\mathcal{L}_{M^2,H}$ is started after time N_0, with an internal seed W independent of $\mathcal{F}_0$ and of all future environment draws.

Lemma 14 (Exact emulation of any causal IV learner)**.** *Fix any i.i.d. law ν on $[0,1]^2$ and let $\nu^{\otimes}$ be the product of its marginals. On the accept branch, conditionally on $\mathcal{F}_0$ and on the seed W, the complete virtual history*

$$
(X_1, \widetilde{Z}_1, \ldots, X_H, \widetilde{Z}_H)
$$

has exactly the same law as the history produced by running the causal learner $\mathcal{L}_{M^2,H}$ for H rounds against the product environment $\nu^{\otimes}$.

Proof. Fix $\mathcal{F}_0$ and W. For each virtual round $k = 1, \ldots, H$, define

$$
\bar{S}_k := S_{N_0+2k-1}, \qquad \bar{B}_k := B_{N_0+2k}.
$$

Because the environment is i.i.d. across real rounds and the pairs above use disjoint post-stage-I rounds, the sequence

$$
(\bar{S}_k, \bar{B}_k)_{k=1}^{H}
$$

is, conditionally on $\mathcal{F}_0$, i.i.d. with common law

$$
\nu_S \otimes \nu_B = \nu^{\otimes},
$$

and it is independent of the seed W. In the accept branch, the meta-strategy queries the learner for X_k, plays (X_k, X_k) on the two real rounds corresponding to $(\bar{S}_k, \bar{B}_k)$, and returns the synthetic feedback

$$
\widetilde{Z}_k = \big(\mathbf{1}\{\bar{S}_k \le X_k\},\ \mathbf{1}\{\bar{B}_k \ge X_k\}\big), \qquad X_k = \pi_k(W, \widetilde{Z}_1, \ldots, \widetilde{Z}_{k-1}).
$$

Thus, after conditioning on $(\mathcal{F}_0, W)$, the pair $(X_k, \widetilde{Z}_k)$ is generated from the i.i.d. product samples $(\bar{S}_k, \bar{B}_k) \sim \nu^{\otimes}$ by *exactly the same measurable recursion* as in a fresh run of $\mathcal{L}_{M^2,H}$ against the product environment $\nu^{\otimes}$. Consequently every downstream internal quantity of the learner is pathwise identical in the emulated execution and in the fresh run on $\nu^{\otimes}$, because each is a measurable function of the same seed W and the same past virtual feedback.

Equivalently, if one imagines revealing the product-law sample path

$$(\bar{S}_1, \bar{B}_1), \ldots, (\bar{S}_H, \bar{B}_H)$$

to a fresh copy of the learner, then at each virtual round the learner would request the same price X_k and would receive exactly the feedback $\widetilde{Z}_k$. Hence the whole history

$$(X_1, \widetilde{Z}_1, \ldots, X_H, \widetilde{Z}_H)$$

has the same conditional law as in a fresh run on $\nu^{\otimes}$.

This proof is pathwise in the exact queried prices X_k. No step requires X_k to lie on the outer grid $\{x_j\}$ or to share any discretization with Stage I. □

B.6 Analysis of the two branches

Let

$$\mathfrak{R}_{\mathrm{I}} := N_0 g_\nu^* - \sum_{t=1}^{N_0} Y_t$$

be the realized diagnostic-stage regret and

$$\mathfrak{R}_{\mathrm{II}} := \mathfrak{R}_T - \mathfrak{R}_{\mathrm{I}}$$

be the realized post-diagnostic regret.

Lemma 15 (Conditional regret of the accept branch)**.** *Let $\nu \in \mathcal{H}(M)$. Define*

$$D_\nu := \sup_{x\in[0,1]} |g_\nu(x) - \tilde{g}_\nu(x)|, \qquad H := \Big\lfloor \frac{T - N_0}{2} \Big\rfloor.$$

Then on the accept event Acc*,*

$$\mathbb{E}_\nu\big[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0\big] \le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 4HD_\nu + 1.$$

In particular, if additionally $\mathcal{E}$ holds and $\nu \in \mathcal{H}(M)$, then on $\mathcal{E} \cap$ Acc,

$$\mathbb{E}_\nu\big[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0\big] \le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 4HC_{\mathrm{test}}(M)\varepsilon + 1.$$

If $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$, then on Acc*,*

$$\mathbb{E}_\nu\big[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0\big] \le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 1.$$

Proof. Assume that Acc occurs, so the accept branch is the one actually executed. By Lemma 14, after conditioning on $\mathcal{F}_0$ the complete virtual history seen by $\mathcal{L}_{M^2,H}$ has the same law as a fresh run of the randomized learner $\mathcal{L}_{M^2,H}$ against the product environment $\nu^{\otimes}$: the seed W is sampled independently after Stage I and the synthetic feedback stream has exactly the product-law recursion required by the learner. Since $\nu^{\otimes} \in \mathcal{E}_{\mathrm{IV}}(M^2)$, we may therefore apply Theorem 11 directly under the conditional law given $\mathcal{F}_0$. This gives

$$H\tilde{g}_\nu^* - \mathbb{E}_\nu\left[\sum_{k=1}^{H} \tilde{g}_\nu(X_k) \,\middle|\, \mathcal{F}_0\right] \le C_{\mathrm{IV}}(M^2)H^{2/3}.$$

Each virtual action X_k is implemented on two real rounds, so the conditional expected cumulative reward of those $2H$ real rounds equals

$$2\mathbb{E}_\nu\left[\sum_{k=1}^{H} g_\nu(X_k) \,\middle|\, \mathcal{F}_0\right].$$

Since

$$g_\nu^* \le \tilde{g}_\nu^* + D_\nu, \qquad g_\nu(X_k) \ge \tilde{g}_\nu(X_k) - D_\nu,$$

the conditional regret over the $2H$ emulation rounds is at most

$$\begin{aligned}
&2Hg_\nu^* - 2\mathbb{E}_\nu\left[\sum_{k=1}^{H} g_\nu(X_k) \,\middle|\, \mathcal{F}_0\right] \\
&\le 2H\big(\tilde{g}_\nu^* + D_\nu\big) - 2\mathbb{E}_\nu\left[\sum_{k=1}^{H}(\tilde{g}_\nu(X_k) - D_\nu) \,\middle|\, \mathcal{F}_0\right] \\
&= 2\left(H\tilde{g}_\nu^* - \mathbb{E}_\nu\left[\sum_{k=1}^{H}\tilde{g}_\nu(X_k) \,\middle|\, \mathcal{F}_0\right]\right) + 4HD_\nu \\
&\le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 4HD_\nu.
\end{aligned}$$

If one real round remains after the $2H$ emulation rounds, we upper-bound its conditional regret by 1. This proves the first display. The second follows from Corollary 1; the third follows from $D_\nu = 0$ for $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$. □

Lemma 16 (Conditional regret of the reject branch)**.** *Let $\nu \in \mathcal{H}(M)$. On the event $\mathcal{E} \cap \mathsf{Rej}$,*

$$\mathbb{E}_\nu\big[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0\big] \le (2M+4)(T - N_0)\varepsilon.$$

Proof. Assume $\mathcal{E} \cap \mathsf{Rej}$. The chosen price $\widehat{x}$ is $\mathcal{F}_0$-measurable, and the future rounds are i.i.d. with one-round expected reward $g_\nu(\widehat{x})$. Hence

$$\mathbb{E}_\nu\big[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0\big] = (T - N_0)\big(g_\nu^* - g_\nu(\widehat{x})\big).$$

It remains to bound the one-round suboptimality of $\widehat{x}$. Let $x^* \in \arg\max_{x\in[0,1]} g_\nu(x)$. Choose a grid point x_{j^*} with $|x^* - x_{j^*}| \le \varepsilon$. Since g_ν is $2M$-Lipschitz by Lemma 9,

$$g_\nu^* = g_\nu(x^*) \le g_\nu(x_{j^*}) + 2M\varepsilon.$$

Because $\widehat{j} \in \arg\max_{0\le j\le K} \widehat{g}_j$ and $\mathcal{E}$ implies $|\widehat{g}_j - g_\nu(x_j)| \le 2\varepsilon$ for every j,

$$g_\nu(x_{j^*}) \le \widehat{g}_{j^*} + 2\varepsilon \le \widehat{g}_{\widehat{j}} + 2\varepsilon \le g_\nu(\widehat{x}) + 4\varepsilon.$$

Combining the two displays gives

$$g_\nu^* - g_\nu(\widehat{x}) \le (2M+4)\varepsilon.$$

Multiplying by $T - N_0$ proves the claim. □

B.7 Completion of the proof of Theorem 2

We now keep the resolution exponent γ symbolic. The construction above therefore defines the full family

$$\mathsf{SCM}_{M,T}^{(\gamma)}, \qquad \gamma \in (0, 1/3),$$

appearing in the statement of Theorem 2.

Proof of Theorem 2. Fix $M > 0$, $\gamma \in (0, 1/3)$, and let $\mathsf{SCM}_{M,T}^{(\gamma)}$ be the meta-strategy defined above.

If $T < T_{\mathrm{sm}}(\gamma)$, then regardless of how the algorithm behaves on this finite set of horizons we always have the trivial bound

$$R_T\big(\mathsf{SCM}_{M,T}^{(\gamma)}; \nu\big) \le T \le T_{\mathrm{sm}}(\gamma)$$

for every ν. Since $T^{\max\{1-\gamma,\,3\gamma\}}\log(eT) \ge 1$ for all $T \ge 1$, enlarging the final constants by the factor $T_{\mathrm{sm}}(\gamma)$ absorbs this regime. Hence it suffices to treat $T \ge T_{\mathrm{sm}}(\gamma)$.

Since

$$K \asymp \varepsilon^{-1} = T^\gamma, \qquad m \asymp \varepsilon^{-2}\log(eT) = T^{2\gamma}\log(eT),$$

the diagnostic budget satisfies

$$N_0 = 2m(K-1) \le C_{\mathrm{diag}}(M,\gamma)\, T^{3\gamma}\log(eT)$$

for all $T \ge T_{\mathrm{sm}}(\gamma)$.

Fix any $\nu \in \mathcal{H}(M)$ and let $\mathcal{E}$ be the concentration event from Lemma 13. Pathwise,

$$\mathfrak{R}_{\mathrm{I}} \le N_0, \qquad \mathfrak{R}_{\mathrm{II}} \le T,$$

because $g^*_\nu \in [0,1]$ and every realized reward belongs to $[0,1]$. Therefore

$$\begin{aligned} R_T\big(\mathsf{SCM}^{(\gamma)}_{M,T};\nu\big) &= \mathbb{E}_\nu[\mathfrak{R}_{\mathrm{I}}] + \mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}}\mathbf{1}_{\mathcal{E}}] + \mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}}\mathbf{1}_{\mathcal{E}^c}] \\ &\le N_0 + \mathbb{E}_\nu[\mathbf{1}_{\mathcal{E}\cap\mathsf{Acc}}\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0]] + \mathbb{E}_\nu[\mathbf{1}_{\mathcal{E}\cap\mathsf{Rej}}\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0]] + T\mathbb{P}_\nu(\mathcal{E}^c). \end{aligned}$$

On $\mathcal{E}\cap\mathsf{Acc}$, Lemma 15 and Corollary 1 give

$$\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0] \le 2C_{\mathrm{IV}}(M^2)H^{2/3} + 4HC_{\mathrm{test}}(M)\varepsilon + 1 \le C_{\mathrm{acc}}(M)\big(T^{2/3} + T^{1-\gamma} + 1\big).$$

On $\mathcal{E}\cap\mathsf{Rej}$, Lemma 16 yields

$$\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0] \le (2M+4)(T-N_0)\varepsilon \le C_{\mathrm{rej}}(M)\, T^{1-\gamma}.$$

Finally, Lemma 13 gives

$$T\mathbb{P}_\nu(\mathcal{E}^c) \le \frac{1}{2T}.$$

Combining the previous displays,

$$R_T\big(\mathsf{SCM}^{(\gamma)}_{M,T};\nu\big) \le C(M,\gamma)\big(T^{3\gamma}\log(eT) + T^{2/3} + T^{1-\gamma}\big).$$

Because $\gamma < 1/3$, we have $1-\gamma > 2/3$, so the $T^{2/3}$ term is dominated by $T^{1-\gamma}$. Hence

$$R_T\big(\mathsf{SCM}^{(\gamma)}_{M,T};\nu\big) \le C_2(M,\gamma)\, T^{\max\{1-\gamma,\,3\gamma\}}\log(eT).$$

Taking the supremum over $\nu \in \mathcal{H}(M)$ proves the hard-side bound.

Now suppose that $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$. Then $D_\nu = 0$, and on $\mathcal{E}$ Corollary 1(i) forces acceptance. Therefore

$$\begin{aligned} R_T\big(\mathsf{SCM}^{(\gamma)}_{M,T};\nu\big) &\le N_0 + \mathbb{E}_\nu[\mathbf{1}_{\mathcal{E}}\mathbb{E}_\nu[\mathfrak{R}_{\mathrm{II}} \mid \mathcal{F}_0]] + T\mathbb{P}_\nu(\mathcal{E}^c) \\ &\le N_0 + 2C_{\mathrm{IV}}(M^2)H^{2/3} + 1 + \frac{1}{2T} \\ &\le C_1(M,\gamma)\big(T^{3\gamma}\log(eT) + T^{2/3}\big). \end{aligned}$$

Since $\log(eT) \ge 1$, this implies

$$R_T\big(\mathsf{SCM}^{(\gamma)}_{M,T};\nu\big) \le C_1'(M,\gamma)\, T^{\max\{2/3,\,3\gamma\}}\log(eT).$$

Taking the supremum over $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$ proves the easy-side bound and completes the proof. □

A useful specialization. Choosing $\gamma = \alpha/3$ for $\alpha \in [2/3, 3/4]$ gives

$$\max\{2/3, 3\gamma\} = \alpha, \qquad \max\{1-\gamma, 3\gamma\} = 1 - \frac{\alpha}{3},$$

so Theorem 2 immediately yields the frontier stated in the main text. In particular, choosing $\gamma = 2/9$ gives the endpoint pair $\big(\widetilde{O}(T^{2/3}), \widetilde{O}(T^{7/9})\big)$.

C Proof of Theorem 3

The proof is organized to mirror the algorithm. We first isolate the two primitives that the meta-strategy relies on: direct value probes and a robust fallback. We then analyze the phase learner PhaseScout, whose all-prefix guarantee is the invariant later used on the IV side. Finally, we analyze the meta-strategy itself, first on the generic side and then on the IV side, and combine the two conclusions at the end.

We use the model, exact-value formulation, and regret notation from Sections 2 and 5. In particular, the side information is the scalar

$$V := g_\nu^*.$$

For some bookkeeping steps it is convenient to introduce the auxiliary benchmark-gap process

$$\mathfrak{B}_T(A;\nu) := Tg_\nu^* - \sum_{t=1}^{T} r_\nu(P_t, Q_t),$$

so that

$$R_T(A;\nu) = \mathbb{E}_\nu\big[\mathfrak{B}_T(A;\nu)\big].$$

We keep the main-text class notation: $\mathcal{H}(M)$ is the bounded-density i.i.d. class and $\mathcal{E}_{\mathrm{IV}}(M) \subseteq \mathcal{H}(M)$ is the independent-values subclass. Throughout, $M > 0$ is fixed and known to the algorithm, and $\mathrm{polylog}(T)$ abbreviates a quantity bounded by $C(\log(T+2))^c$ for constants C, c depending only on M.

We repeatedly use the first claim of Lemma 9: for every $\nu \in \mathcal{H}(M)$, the same-price objective g_ν is $2M$-Lipschitz on $[0,1]$.

C.1 Auditing primitives and the robust fallback

C.1.1 Direct point probes and block audits

The direct one-round probe below is a two-bit reformulation of the unbiased same-price estimator used in the smoothed-adversary bilateral-trade literature [2, 3]. We isolate it here because, in the present exact-value setting, it will be used as a sparse audit rather than as a continuously repeated exploration primitive. The point of this appendix is therefore not to claim the probe itself as new, but to show how exact-value side information lets one wrap that inherited probe into a qualitatively different auditing architecture.

Algorithm 5 VALUEPROBE(x): direct value probe at price x

Sample $U \sim \mathrm{Unif}\{\mathrm{L}, \mathrm{R}\}$.
if $U = \mathrm{L}$ **then**
  Sample $P \sim \mathrm{Unif}[0, x]$, play (P, x), and return $Y = 2x\,\mathbf{1}\{S \le P\}\mathbf{1}\{B \ge x\}$.
else
  Sample $Q \sim \mathrm{Unif}[x, 1]$, play (x, Q), and return $Y = 2(1-x)\,\mathbf{1}\{S \le x\}\mathbf{1}\{B \ge Q\}$.
end if

We analyze the direct audit primitive recorded in Algorithm 5.

Lemma 17 (Direct point probe)**.** *Fix $x \in [0,1]$ and run Algorithm 5. Then $Y \in [0,2]$ and*

$$\mathbb{E}_\nu[Y] = g_\nu(x).$$

Proof. For $p, q \in [0,1]$, write $C_\nu(p,q) := \mathbb{P}_\nu(S \le p,\ B \ge q)$. Lemma 8 gives

$$g_\nu(x) = \int_0^x C_\nu(p,x)\,dp + \int_x^1 C_\nu(x,q)\,dq.$$

If $x \in (0,1)$, conditioning on the coin flip yields

$$\mathbb{E}[Y \mid U = \mathrm{L}] = 2\int_0^x C_\nu(p,x)\,dp, \qquad \mathbb{E}[Y \mid U = \mathrm{R}] = 2\int_x^1 C_\nu(x,q)\,dq,$$

so averaging over $U \sim \mathrm{Unif}\{\mathrm{L}, \mathrm{R}\}$ gives $\mathbb{E}[Y] = g_\nu(x)$.

If $x = 0$, the left branch outputs 0 deterministically and the right branch is also zero almost surely because bounded density implies $\mathbb{P}(S = 0) = 0$; moreover $g_\nu(0) = 0$. The case $x = 1$ is symmetric. Hence the identity holds for all $x \in [0,1]$. □

Lemma 18 (Block audit)**.** *Consider a consecutive block of same-price play consisting of prices*

$$x_1, \dots, x_b \in [0, 1].$$

Choose an index $\tau \sim \mathrm{Unif}\{1, \dots, b\}$ independently of everything else, and on a fresh environment draw run Algorithm 5 at x_τ. Let Y be the resulting output. Then $Y \in [0, 2]$ and

$$\mathbb{E}[Y \mid x_1, \dots, x_b] = \frac{1}{b} \sum_{t=1}^{b} g_\nu(x_t).$$

Proof. Condition on the block history. By Lemma 17,

$$\mathbb{E}[Y \mid \tau = t, x_1, \dots, x_b] = g_\nu(x_t).$$

Averaging over the uniform index τ gives the claim. □

C.1.2 Weighted checkpoint concentration

Inside a phase we use odd block lengths

$$b_1 = 1, \quad b_2 = 3, \quad b_3 = 5, \ \dots, \quad b_j = 2j - 1,$$

so that

$$\sum_{i=1}^{j} b_i = j^2.$$

Lemma 19 (Weighted checkpoint estimator)**.** *Fix a phase and suppose its first j audited blocks have lengths $b_i = 2i - 1$, $i = 1, \dots, j$. Let $\bar{g}_i$ be the true block-average value in block i and let Y_i be the corresponding block audit from Lemma 18. Define*

$$\widehat{G}_j := \frac{1}{j^2} \sum_{i=1}^{j} b_i Y_i, \qquad G_j := \frac{1}{j^2} \sum_{i=1}^{j} b_i \bar{g}_i.$$

Then

$$\mathbb{E}[\widehat{G}_j \mid \mathcal{F}_j^{\mathrm{play}}] = G_j,$$

where $\mathcal{F}_j^{\mathrm{play}}$ is the sigma-field generated by the same-price play in the first j blocks. Moreover, there exists a universal constant $c_* > 0$ such that for every $u > 0$,

$$\mathbb{P}\big(|\widehat{G}_j - G_j| > u \mid \mathcal{F}_j^{\mathrm{play}}\big) \le 2 \exp\Big(-\frac{j u^2}{c_*}\Big).$$

In particular, for

$$\beta_j := c_0 \sqrt{\frac{L}{j}},$$

where $L \ge 1$ and c_0 is a sufficiently large universal constant,

$$\mathbb{P}\big(|\widehat{G}_j - G_j| > \beta_j \mid \mathcal{F}_j^{\mathrm{play}}\big) \le 2e^{-L}.$$

Proof. Unbiasedness follows by summing the conditional expectation in Lemma 18 over the j blocks.

For concentration, define

$$Z_i := b_i(Y_i - \bar{g}_i), \qquad i = 1, \dots, j.$$

Conditional on $\mathcal{F}_j^{\mathrm{play}}$, the variables $Z_1, \dots, Z_j$ are independent, centered, and satisfy

$$|Z_i| \le 2b_i.$$

Hence Azuma–Hoeffding (or Hoeffding for independent bounded variables) gives

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{j} Z_i\Big| > u j^2 \ \Big|\ \mathcal{F}_j^{\mathrm{play}}\Big) \le 2 \exp\Big(-\frac{u^2 j^4}{2 \sum_{i=1}^{j} (2b_i)^2}\Big).$$

Now

$$\sum_{i=1}^{j} b_i^2 = \sum_{i=1}^{j} (2i-1)^2 = \frac{4j^3 - j}{3} \le \frac{4}{3} j^3.$$

Therefore

$$\mathbb{P}\big(|\widehat{G}_j - G_j| > u \mid \mathcal{F}_j^{\text{play}}\big) \le 2\exp\Big(-\frac{3}{32} j u^2\Big).$$

This proves the first inequality. The last display follows by plugging in $u = \beta_j$ and choosing c_0 sufficiently large. □

C.1.3 Robust fallback

For the post-rejection fallback we do not design a new generic $T^{3/4}$ routine. Instead, for every horizon H, we use the two-price weak-budget-balance algorithm *Scouting Blindits* of Cesa-Bianchi et al. [1]. In our notation, RF_H denotes that horizon-H routine.

Proposition 2 (Robust fallback)**.** *For every horizon $H \ge 0$ there exists an algorithm $\mathsf{RF}_H = \mathsf{RF}_{H,M}$ such that for every $\nu \in \mathcal{H}(M)$,*

$$R_H(\mathsf{RF}_H; \nu) \le C_{\text{gen}}(M)\,\text{polylog}(H+2)\,(H+1)^{3/4}.$$

Proof. We use the weakly-budget-balanced two-price version of Scouting Blindits from [1, Sec. 7, Thm. 8], whose regret is stated for the same reward $\text{GFT}^{\text{WBB}}(p,q,s,b) = ((p-s)+(b-q))\mathbf{1}\{s \le p \le q \le b\}$. Hence the bound applies directly to our regret notion; no comparison with the gross two-price surplus $(b-s)\mathbf{1}\{s \le p \le q \le b\}$ is used. □

Algorithm 6 Interface of the robust fallback RF_H

1: **Input:** remaining horizon H.
2: Run the horizon-H robust routine guaranteed by Proposition 2.

C.2 Phase IV guarantee

We now analyze the phase learner that will later be used inside the Value-Auditing Meta-Strategy. Because our feedback already reveals both same-price bits $\mathbf{1}\{S \le x\}$ and $\mathbf{1}\{B \ge x\}$ on every same-price round, the learner can be written directly in the present two-bit model. Conceptually, PhaseScout is a phase-local repackaging of the independent-values same-price machinery from [1]; the genuinely new work here is not another fixed-regime $\widetilde{O}(n^{2/3})$ guarantee, but the stronger simultaneous high-probability all-prefix guarantee later formalized in Proposition 4 and adapted from [7]. Combined with Lemma 26 and the audit concentration event from Lemma 19, that guarantee is exactly what prevents the meta-strategy from spuriously switching in the IV regime. The only structural point we use is that audits occupy disjoint real rounds and are never fed back to PhaseScout, so the same-price trajectory of each phase can be analyzed on its own.

C.2.1 Phase model and learner

Fix a product environment $\nu \in \mathcal{E}_{\text{IV}}(M)$, a phase horizon $n \ge 16$, and a phase confidence level $\delta_{\text{ph}} \in (0, 1/4)$. Set the grid parameters by

$$K_n = H_n := \lceil n^{1/3} \rceil, \qquad \Delta_n := K_n^{-1}, \qquad p_k := k\Delta_n, \quad k \in [K_n]. \tag{6}$$

$$L_n := \log \frac{512 K_n n^3}{\delta_{\text{ph}}}. \tag{7}$$

Write

$$q_k := \mathbb{P}(B \ge p_k), \qquad s_k := \mathbb{P}(S \le p_k),$$

$$J_k^B := \int_{p_k}^{1} \mathbb{P}(B \ge \lambda)\, d\lambda, \qquad J_k^S := \int_{0}^{p_k} \mathbb{P}(S \le \lambda)\, d\lambda.$$

Then

$$g_\nu(p_k) = s_k J_k^B + q_k J_k^S.$$

Indeed, on $\{S \le p_k \le B\}$ one has

$$(B - S) = (B - p_k) + (p_k - S),$$

so independence and Fubini give

$$\mathbb{E}[(B - p_k)\mathbf{1}\{S \le p_k \le B\}] = \mathbb{P}(S \le p_k) \int_{p_k}^{1} \mathbb{P}(B \ge \lambda)\, d\lambda = s_k J_k^B,$$

and likewise

$$\mathbb{E}[(p_k - S)\mathbf{1}\{S \le p_k \le B\}] = \mathbb{P}(B \ge p_k) \int_{0}^{p_k} \mathbb{P}(S \le \lambda)\, d\lambda = q_k J_k^S.$$

For each arm $k \in [K_n]$ and each visit index $r \ge 1$, let

$$Y_{k,r}^B := \mathbf{1}\{B^{(k,r)} \ge p_k\}, \qquad Y_{k,r}^S := \mathbf{1}\{S^{(k,r)} \le p_k\}$$

be the buyer and seller bits observed on the r-th play of price p_k. The deterministic exploration prefix has length

$$w_n := H_n K_n. \tag{8}$$

Write X_t for the price played at phase round t. After this exploration prefix, for every phase round $t \ge w_n$ define

$$T_{k,t} := \sum_{s=1}^{t} \mathbf{1}\{X_s = p_k\}, \qquad \widehat{\nu}_{k,t} := \frac{1}{T_{k,t}} \sum_{r=1}^{T_{k,t}} Y_{k,r}^B, \qquad \widehat{\mu}_{k,t} := \frac{1}{T_{k,t}} \sum_{r=1}^{T_{k,t}} Y_{k,r}^S. \tag{9}$$

The exploration prefix also yields the integral estimates

$$\widehat{J}_k^B := \frac{1}{H_n K_n} \sum_{i=k}^{K_n - 1} \sum_{r=1}^{H_n} Y_{i,r}^B, \qquad \widehat{J}_k^S := \frac{1}{H_n K_n} \sum_{i=1}^{k} \sum_{r=1}^{H_n} Y_{i,r}^S. \tag{10}$$

Introduce the clipped optimistic factors

$$\overline{\mu}_{k,t} := \min\left\{1,\ \widehat{\mu}_{k,t} + \sqrt{\frac{L_n}{2T_{k,t}}}\right\}, \qquad \overline{\nu}_{k,t} := \min\left\{1,\ \widehat{\nu}_{k,t} + \sqrt{\frac{L_n}{2T_{k,t}}}\right\},$$

$$\overline{J}_k^B := \min\left\{1,\ \widehat{J}_k^B + \sqrt{\frac{L_n}{2H_n K_n}} + \frac{M}{K_n}\right\}, \qquad \overline{J}_k^S := \min\left\{1,\ \widehat{J}_k^S + \sqrt{\frac{L_n}{2H_n K_n}} + \frac{M}{K_n}\right\}.$$

The exploitation index is then

$$\mathrm{UCB}_{k,t} := \overline{\mu}_{k,t}\, \overline{J}_k^B + \overline{\nu}_{k,t}\, \overline{J}_k^S. \tag{11}$$

Algorithm 7 records the phase learner $\mathsf{PhaseScout}_{n,\delta_{\mathrm{ph}}}$. Inside the meta-strategy we use it only through the induced stateful calls *Init*, *NextPrice*, and *Update*: the meta-strategy requests the next same price, plays it, streams back the two observed bits, and never feeds any audit outcome into PhaseScout.

Algorithm 7 $\mathsf{PhaseScout}_{n,\delta_{\mathrm{ph}}}$: phase IV learner

Input: phase horizon $n \ge 16$, confidence level $\delta_{\mathrm{ph}} \in (0, 1/4)$.
Set $(K_n, H_n, \Delta_n, p_k)$ by (6) and L_n by (7).
Exploration.
for $k = 1$ to K_n **do**
    Play price p_k for H_n phase rounds and record both acceptance bits on each round.
end for
Set w_n by (8) and $(\widehat{J}_k^B, \widehat{J}_k^S)$ by (10).
Optimistic exploitation.
for $t = w_n + 1$ to n **do**
    for $k = 1$ to K_n **do**
        Compute $(T_{k,t-1}, \widehat{\nu}_{k,t-1}, \widehat{\mu}_{k,t-1})$ from (9) and $\mathrm{UCB}_{k,t-1}$ from (11).
    end for
    Play the *smallest* maximizer of $\mathrm{UCB}_{k,t-1}$ over $k \in [K_n]$.
end for

The following latent-variable representation will be used repeatedly. For each arm $k \in [K_n]$, let

$$(U_{k,r})_{r\geq 1} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(q_k), \qquad (V_{k,r})_{r\geq 1} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(s_k),$$

all sequences being mutually independent. Interpret $U_{k,r}$ as the buyer bit and $V_{k,r}$ as the seller bit on the r-th play of price p_k inside the phase learner. By the i.i.d. product structure of ν, this latent construction has exactly the same law as the feedback process generated by Algorithm 7.

C.2.2 Concentration and structural lemmas

Lemma 20 (Online empirical means for both coordinates)**.** *With probability at least* $1 - \delta_{\text{ph}}/16$*, simultaneously for every arm* $k \in [K_n]$ *and every phase round* $t \geq w_n$*,*

$$|\widehat{\nu}_{k,t} - q_k| \leq \sqrt{\frac{L_n}{2T_{k,t}}}, \qquad |\widehat{\mu}_{k,t} - s_k| \leq \sqrt{\frac{L_n}{2T_{k,t}}}.$$

Proof. Fix an arm k and an integer $s \in [H_n, n]$. Whenever $T_{k,t} = s$, the pair $(\widehat{\nu}_{k,t}, \widehat{\mu}_{k,t})$ consists of the empirical means of the Bernoulli samples $(U_{k,1}, \ldots, U_{k,s})$ and $(V_{k,1}, \ldots, V_{k,s})$. Hence Hoeffding's inequality gives

$$\mathbb{P}\left(|\widehat{\nu}_{k,t} - q_k| > \sqrt{\frac{L_n}{2s}} \text{ and } T_{k,t} = s\right) \leq 2e^{-L_n},$$

$$\mathbb{P}\left(|\widehat{\mu}_{k,t} - s_k| > \sqrt{\frac{L_n}{2s}} \text{ and } T_{k,t} = s\right) \leq 2e^{-L_n}.$$

A union bound over both coordinates, all $k \in [K_n]$, all $t \in \{w_n, \ldots, n\}$, and all $s \in [H_n, n]$ yields total failure probability at most

$$4K_n n^2 e^{-L_n} \leq \delta_{\text{ph}}/16.$$

□

Lemma 21 (Exploration integral estimates)**.** *With probability at least* $1 - \delta_{\text{ph}}/16$*, the following two statements hold simultaneously.*

(i) For every $k \in [K_n]$*,*

$$\left|\widehat{J}_k^B - J_k^B\right| \leq \sqrt{\frac{L_n}{2H_nK_n}} + \frac{M}{K_n}.$$

(ii) For every $k \in [K_n]$*,*

$$\left|\widehat{J}_k^S - J_k^S\right| \leq \sqrt{\frac{L_n}{2H_nK_n}} + \frac{M}{K_n}.$$

Proof. Because $\nu \in \mathcal{E}_{\text{IV}}(M) \subseteq \mathcal{H}(M)$, Lemma 9 implies that

$$q(\lambda) := \mathbb{P}(B \geq \lambda) = \overline{F}_\nu(\lambda), \qquad s(\lambda) := \mathbb{P}(S \leq \lambda) = F_\nu(\lambda)$$

are both M-Lipschitz on $[0, 1]$. Hence the deterministic Riemann sums

$$\widetilde{J}_k^B := \frac{1}{K_n} \sum_{i=k}^{K_n-1} q_i, \qquad \widetilde{J}_k^S := \frac{1}{K_n} \sum_{i=1}^{k} s_i$$

satisfy

$$|\widetilde{J}_k^B - J_k^B| \leq \frac{M}{K_n}, \qquad |\widetilde{J}_k^S - J_k^S| \leq \frac{M}{K_n} \qquad \forall k \in [K_n].$$

For the stochastic part, fix k. The variable $\widehat{J}_k^B$ is a weighted sum of the independent Bernoulli family $\{U_{i,r} : i \in \{k, \ldots, K_n - 1\},\ r \in [H_n]\}$, each with weight $(H_nK_n)^{-1}$. Therefore

$$\mathbb{P}\Big(|\widehat{J}_k^B - \widetilde{J}_k^B| > u\Big) \leq 2\exp(-2H_nK_nu^2).$$

Taking $u = \sqrt{L_n/(2H_nK_n)}$ and union-bounding over $k \in [K_n]$ gives failure probability at most $2K_ne^{-L_n} \leq \delta_{\text{ph}}/128$. The same argument for the family $\{V_{i,r} : i \in \{1, \ldots, k\},\ r \in [H_n]\}$ yields the analogous bound for $\widehat{J}_k^S$. Combining deterministic and stochastic errors proves the lemma. □

Lemma 22 (Deterministic scale relations)**.** *There exists a constant $C_M \geq 1$, depending only on M, such that for every $n \geq 16$ the following hold.*

(i) $H_n \leq 2n^{1/3}$, $K_n \leq 2n^{1/3}$, *and* $w_n = H_n K_n \leq 4n^{2/3}$.

(ii) For every arm $k \in [K_n]$ and every phase round $t \geq w_n$, one has

$$H_n \leq T_{k,t} \leq n.$$

(iii) The exploration radius satisfies

$$\sqrt{\frac{L_n}{2H_n K_n}} + \frac{M}{K_n} \leq C_M L_n n^{-1/3}.$$

Proof. Since $n \geq 16$, one has $n^{1/3} \geq 2$, hence

$$H_n = K_n = \left\lceil n^{1/3} \right\rceil \leq n^{1/3} + 1 \leq 2n^{1/3},$$

which proves part (i), and then

$$w_n = H_n K_n \leq 4n^{2/3}.$$

Part (ii) is immediate from the deterministic exploration schedule, which plays each arm exactly H_n times before the exploitation phase starts. For part (iii), use $H_n K_n \geq n^{2/3}$ and $K_n \geq n^{1/3}$ to obtain

$$\sqrt{\frac{L_n}{2H_n K_n}} + \frac{M}{K_n} \leq \sqrt{\frac{L_n}{2}}\, n^{-1/3} + M n^{-1/3} \leq (1+M) L_n n^{-1/3},$$

because $L_n \geq 1$. This proves the lemma. □

Proposition 3 (Phase confidence bound)**.** *There exists a constant $C_{\mathrm{IV}} \geq 1$, depending only on M, such that the following event has probability at least $1 - \delta_{\mathrm{ph}}/4$: for every phase round $t \geq w_n$ and every arm $k \in [K_n]$,*

$$0 \leq \mathrm{UCB}_{k,t} - g_\nu(p_k) \leq C_{\mathrm{IV}} L_n \left(n^{-1/3} + T_{k,t}^{-1/2} \right).$$

Proof. Work on the intersection of the events from Lemmas 20 and 21; by a union bound this event has probability at least $1 - \delta_{\mathrm{ph}}/4$. Fix $t \geq w_n$ and $k \in [K_n]$. Set

$$a_{k,t} := \sqrt{\frac{L_n}{2T_{k,t}}}, \qquad b_n := \sqrt{\frac{L_n}{2H_n K_n}} + \frac{M}{K_n}.$$

By Lemmas 20 and 21,

$$s_k \leq \widehat{\mu}_{k,t} + a_{k,t} \leq s_k + 2a_{k,t}, \qquad q_k \leq \widehat{\nu}_{k,t} + a_{k,t} \leq q_k + 2a_{k,t},$$
$$J_k^B \leq \widehat{J}_k^B + b_n \leq J_k^B + 2b_n, \qquad J_k^S \leq \widehat{J}_k^S + b_n \leq J_k^S + 2b_n.$$

Because all optimistic factors are clipped at 1, this implies

$$s_k \leq \overline{\mu}_{k,t} \leq s_k + 2a_{k,t}, \qquad q_k \leq \overline{\nu}_{k,t} \leq q_k + 2a_{k,t},$$
$$J_k^B \leq \overline{J}_k^B \leq J_k^B + 2b_n, \qquad J_k^S \leq \overline{J}_k^S \leq J_k^S + 2b_n.$$

Hence optimism is immediate:

$$\mathrm{UCB}_{k,t} = \overline{\mu}_{k,t}\, \overline{J}_k^B + \overline{\nu}_{k,t}\, \overline{J}_k^S \geq s_k J_k^B + q_k J_k^S = g_\nu(p_k).$$

For the upper bound, use $\overline{J}_k^B, \overline{J}_k^S, s_k, q_k \in [0,1]$ to write

$$\overline{\mu}_{k,t}\, \overline{J}_k^B - s_k J_k^B = (\overline{\mu}_{k,t} - s_k)\overline{J}_k^B + s_k(\overline{J}_k^B - J_k^B) \leq 2a_{k,t} + 2b_n,$$
$$\overline{\nu}_{k,t}\, \overline{J}_k^S - q_k J_k^S = (\overline{\nu}_{k,t} - q_k)\overline{J}_k^S + q_k(\overline{J}_k^S - J_k^S) \leq 2a_{k,t} + 2b_n.$$

Therefore

$$\mathrm{UCB}_{k,t} - g_\nu(p_k) \le 4a_{k,t} + 4b_n.$$

By Lemma 22(iii),

$$b_n \le C_M L_n n^{-1/3}.$$

Also $a_{k,t} \le L_n T_{k,t}^{-1/2}$ because $L_n \ge 1$. Therefore

$$\mathrm{UCB}_{k,t} - g_\nu(p_k) \le C'_{\mathrm{IV}} L_n \left(n^{-1/3} + T_{k,t}^{-1/2} \right)$$

for a constant $C'_{\mathrm{IV}} = C'_{\mathrm{IV}}(M)$. After renaming the constant, this proves the proposition. □

Proposition 4 (Phasewise IV prefix bound)**.** *There exists a constant $A_{\mathrm{IV}} \ge 1$, depending only on M, such that the following holds. Fix a phase horizon $n \ge 16$ and a target phase failure probability $\delta_{\mathrm{ph}} \in (0, 1/4)$. Run Algorithm 7, and let $X_1, \dots, X_n$ be its same prices. Then with probability at least $1 - \delta_{\mathrm{ph}}$, simultaneously for every suffix-prefix length*

$$m \in \{1, \dots, n - w_n\},$$

one has

$$\sum_{t=w_n+1}^{w_n+m} \big(V - g_\nu(X_t)\big) \le A_{\mathrm{IV}} L_n \Big(m n^{-1/3} + n^{1/6}\sqrt{m} \Big). \tag{12}$$

Consequently,

$$\sum_{t=1}^{n} \big(V - g_\nu(X_t)\big) \le A_{\mathrm{IV}} L_n n^{2/3}. \tag{13}$$

Proof. By Proposition 3, there exists an event of probability at least $1 - \delta_{\mathrm{ph}}$ on which, for every arm $k \in [K_n]$ and every post-exploration phase round t,

$$0 \le \mathrm{UCB}_{k,t-1} - g_\nu(p_k) \le C_{\mathrm{IV}} L_n \left(n^{-1/3} + T_{k,t-1}^{-1/2} \right).$$

Work on this event.

Let $k^\star \in \arg\max_{k \in [K_n]} g_\nu(p_k)$ be a best grid arm. Since the grid spacing is $\Delta_n = 1/K_n$ and g_ν is $2M$-Lipschitz,

$$0 \le V - g_\nu(p_{k^\star}) \le 2M\Delta_n \le 2M n^{-1/3}. \tag{14}$$

Fix a post-exploration round $t > w_n$ and let k_t be the arm played at that round. Since the subroutine chooses a maximizer of (11) over all $k \in [K_n]$,

$$g_\nu(p_{k^\star}) \le \mathrm{UCB}_{k^\star,t-1} \le \mathrm{UCB}_{k_t,t-1}.$$

Applying Proposition 3 to the played arm k_t yields

$$g_\nu(p_{k^\star}) \le g_\nu(p_{k_t}) + C_{\mathrm{IV}} L_n \left(n^{-1/3} + T_{k_t,t-1}^{-1/2} \right).$$

Combining with (14) gives

$$V - g_\nu(X_t) \le A_1 L_n \left(n^{-1/3} + T_{k_t,t-1}^{-1/2} \right)$$

for a constant $A_1 = A_1(M)$. Summing over $t = w_n + 1, \dots, w_n + m$,

$$\sum_{t=w_n+1}^{w_n+m} (V - g_\nu(X_t)) \le A_1 L_n \left(m n^{-1/3} + \sum_{k=1}^{K_n} \sum_{r=1}^{N_k(m)} (H_n + r - 1)^{-1/2} \right),$$

where $N_k(m)$ is the number of post-exploration plays of arm k inside this prefix. Since

$$\sum_{r=1}^{N} (H_n + r - 1)^{-1/2} \le 2\sqrt{N},$$

we obtain

$$\sum_{k=1}^{K_n}\sum_{r=1}^{N_k(m)}(H_n+r-1)^{-1/2}\le 2\sum_{k=1}^{K_n}\sqrt{N_k(m)}\le 2\sqrt{K_n m}\le 3n^{1/6}\sqrt{m},$$

where the last step uses Lemma 22(i). This proves (12) after enlarging the constant.

Finally, (13) follows by taking $m=n-w_n$ and adding the exploration prefix cost. By Lemma 22(i),

$$w_n\le 4n^{2/3}.$$

□

C.3 The Value-Auditing Meta-Strategy

C.3.1 Phase budgets and thresholds

Fix the total horizon $T\ge 1$ and the global failure level

$$\delta:=T^{-12}.$$

Define the logarithmic factor

$$L:=\log(C_0T^{20}),$$

where C_0 is a sufficiently large universal constant. By enlarging C_0 if necessary we may and do assume $L\ge 1$ for every $T\ge 1$. Fix once and for all three deterministic design constants

$$C_{\mathrm{base}}=C_{\mathrm{base}}(M),\qquad C_\Theta=C_\Theta(M),\qquad C_j=C_j(M),$$

depending only on the known density bound M, large enough that

$$C_\Theta\ge 32A_{\mathrm{IV}}\qquad\text{and}\qquad C_j\ge 64A_{\mathrm{IV}}^2,$$

and that the feasibility argument below goes through. Then define the base epoch scale

$$B_\star:=2^{\lceil\log_2(C_{\mathrm{base}}L^6)\rceil}. \tag{15}$$

By enlarging C_{base} if necessary we may and do assume $B_\star\ge 32$. If $T<B_\star$, define $\mathsf{VAM}_T(V)$ to play any fixed same price for all T rounds. Lemma 24 disposes of this regime uniformly. From this point on we describe only the nontrivial case $T\ge B_\star$.

For $T\ge B_\star$, set

$$K_T:=1+\lfloor\log_2(T/B_\star)\rfloor \tag{16}$$

and define the deterministic base-scale phase budgets by

$$B_k:=2^{k-1}B_\star,\qquad k=1,\dots,K_T-1,\qquad B_{K_T}:=T-\sum_{i=1}^{K_T-1}2^{i-1}B_\star. \tag{17}$$

Then

$$\sum_{k=1}^{K_T}B_k=T,\qquad B_k=2^{k-1}B_\star\text{ for every }k<K_T,$$
$$2^{K_T-1}B_\star\le T<2^{K_T}B_\star,\qquad B_\star\le B_{K_T}\le T.$$

and every phase budget is deterministic.

For every phase $k\in[K_T]$, the meta-strategy chooses the audit clock as aggressively as the phase budget permits while still leaving a clean horizon for the call to $\mathsf{PhaseScout}$. This is the reason for the next definition: J_k is the largest checkpoint count compatible with a burn-in segment, J_k^2 monitored same-price rounds, and the remaining tail of the phase. Define the feasible audit count by

$$J_k:=\max\left\{j\in\mathbb{Z}_{\ge 0}:\ \left\lceil(B_k-j)^{1/3}\right\rceil^2+j^2\le B_k-j\right\}. \tag{18}$$

Equivalently, J_k is the largest integer $j\ge 0$ such that, writing

$$n:=B_k-j,\qquad w(n):=\left\lceil n^{1/3}\right\rceil^2,$$

one has

$$w(n) + j^2 \le n.$$

Such a value exists because $j = 0$ is feasible whenever $w(B_k) = \left\lceil B_k^{1/3} \right\rceil^2 \le B_k$, and this certainly holds for every $B_k \ge B_\star$ once C_{base} is chosen large enough. Then define the phase parameters

$$n_k := B_k - J_k, \qquad \eta_k := \frac{\delta}{16k^2}, \qquad w_k := \left\lceil n_k^{1/3} \right\rceil^2, \qquad r_k := n_k - w_k - J_k^2, \tag{19}$$

the rejection thresholds

$$\Theta_k := C_\Theta L\, n_k^{-1/3}, \qquad j_{0,k} := \left\lceil C_j L\, n_k^{1/3} \right\rceil, \tag{20}$$

and the checkpoint-deviation radius

$$\beta_{k,j} := c_0 \sqrt{L/j}, \qquad j \in [J_k]. \tag{21}$$

Thus phase k uses exactly n_k same-price rounds of a fresh stateful copy of $\mathsf{PhaseScout}_{n_k,\eta_k}$ and exactly J_k audit rounds, so its total real length is exactly

$$n_k + J_k = B_k.$$

The first w_k same-price rounds are burn-in, the next J_k^2 same-price rounds are partitioned into audited blocks of lengths $1, 3, \dots, 2J_k - 1$, and the remaining tail of length r_k is played without further audits whenever the phase completes without rejection. There is no additional bookkeeping prefix and no extra padding beyond this final unaudited tail.

By feasibility,

$$0 \le r_k.$$

Because J_k is the *largest* feasible audit count, the value $J_k + 1$ is infeasible. Writing $n_k = B_k - J_k$, this means

$$\left\lceil (n_k - 1)^{1/3} \right\rceil^2 + (J_k + 1)^2 > n_k - 1.$$

Since $\left\lceil (n_k - 1)^{1/3} \right\rceil^2 \le \left\lceil n_k^{1/3} \right\rceil^2 = w_k$, it follows that

$$(J_k + 1)^2 > n_k - 1 - w_k.$$

Set

$$A_k := n_k - w_k.$$

If $A_k \le 1$, then $r_k = A_k - J_k^2 \le 1 \le 2\sqrt{n_k}$. If $A_k > 1$, then $J_k + 1 > \sqrt{A_k - 1}$, hence

$$J_k > \sqrt{A_k - 1} - 1$$

and therefore

$$r_k = A_k - J_k^2 < A_k - \left(\sqrt{A_k - 1} - 1\right)^2 = 2\sqrt{A_k - 1} \le 2\sqrt{n_k}.$$

Thus in all cases

$$r_k \le 2\sqrt{n_k}.$$

Also $J_k \le \sqrt{n_k}$, so

$$n_k = B_k - J_k \ge B_k - \sqrt{B_k} \ge \frac{B_k}{2},$$

because every phase satisfies $B_k \ge B_\star \ge 4$. In particular, since $B_\star \ge 32$, every displayed phase satisfies

$$n_k \ge \frac{B_k}{2} \ge 16,$$

so Algorithm 7 is always called within its stated horizon regime.

By the choice of $B_\star$, every phase satisfies

$$j_{0,k} \le J_k. \tag{22}$$

Indeed, for every phase one has $n_k \geq B_k/2 \geq B_\star/2 \geq \frac{1}{2} C_{\mathrm{base}} L^6$. Hence, for C_{base} sufficiently large,

$$w_k = \left\lceil n_k^{1/3} \right\rceil^2 \leq 2 n_k^{2/3}$$

and

$$r_k \leq 2\sqrt{n_k} \leq 2 n_k^{2/3}.$$

Therefore

$$J_k^2 = n_k - w_k - r_k \geq n_k - 4 n_k^{2/3} \geq \frac{1}{2} n_k,$$

so

$$J_k \geq \frac{1}{\sqrt{2}} n_k^{1/2}.$$

On the other hand, because $j_{0,k} = \left\lceil C_j L \, n_k^{1/3} \right\rceil$ and $C_j L \, n_k^{1/3} \geq 1$, one has

$$j_{0,k} \leq 2 C_j L \, n_k^{1/3} \leq \frac{1}{4} n_k^{1/2}$$

for C_{base} sufficiently large. This proves (22).

The final summations over phases use the following deterministic bookkeeping fact.

Lemma 23 (Summability of phase scales and audit overheads)**.** *With the deterministic phase decomposition above, the following hold.*

(i)

$$\sum_{k=1}^{K_T} B_k = T, \qquad B_k = 2^{k-1} B_\star \text{ for every } k < K_T,$$
$$2^{K_T - 1} B_\star \leq T < 2^{K_T} B_\star, \qquad B_\star \leq B_{K_T} \leq T.$$

(ii) For every phase $k \in [K_T]$,

$$n_k \leq B_k, \qquad J_k \leq \sqrt{n_k}, \qquad 0 \leq r_k \leq 2\sqrt{n_k}, \qquad n_k \geq \frac{B_k}{2}.$$

(iii) For every $\alpha \in (0, 1]$,

$$\sum_{k=1}^{K_T} n_k^\alpha \leq C_\alpha T^\alpha$$

for a constant C_α depending only on α.

(iv)

$$\sum_{k=1}^{K_T} \sqrt{n_k} = O(\sqrt{T}), \qquad \sum_{k=1}^{K_T} J_k = O(\sqrt{T}).$$

In particular,

$$\sum_{k=1}^{K_T} n_k^{2/3} = O(T^{2/3}), \qquad \sum_{k=1}^{K_T} n_k^{3/4} = O(T^{3/4}).$$

Proof. Part (i) is immediate from the definition of the budgets and the choice

$$K_T = 1 + \lfloor \log_2(T/B_\star) \rfloor.$$

Indeed,

$$2^{K_T - 1} B_\star \leq T < 2^{K_T} B_\star,$$

and therefore

$$B_{K_T} = T - \left(2^{K_T - 1} - 1\right) B_\star \geq B_\star.$$

Part (ii) records the deterministic phase geometry proved above.

For part (iii), note that $n_k \le B_k$ for every phase, hence

$$\sum_{k=1}^{K_T} n_k^\alpha \le \sum_{k=1}^{K_T} B_k^\alpha = \sum_{k=1}^{K_T-1} (2^{k-1}B_\star)^\alpha + B_{K_T}^\alpha.$$

By part (i),

$$2^{K_T-1}B_\star \le T \qquad \text{and} \qquad B_{K_T} \le T,$$

so

$$\sum_{k=1}^{K_T} n_k^\alpha \le C'_\alpha (2^{K_T-1}B_\star)^\alpha + T^\alpha \le C_\alpha T^\alpha.$$

For part (iv), taking $\alpha = 1/2$ in part (iii) yields

$$\sum_{k=1}^{K_T} \sqrt{n_k} = O(\sqrt{T}).$$

Also $J_k \le \sqrt{n_k}$ for every phase, so

$$\sum_{k=1}^{K_T} J_k = O(\sqrt{T}).$$

The displayed special cases follow from part (iii) with $\alpha \in \{2/3, 3/4\}$. □

C.3.2 Algorithm

A *reserved same-price slot* of state st means: query the next price from PhaseScout, play that same price on the real market, reveal the two same-price bits, and update the hidden IV state with those bits. Audit rounds are separate fresh market rounds and are never fed back to PhaseScout.

For the proofs it is convenient to adopt a *latent continuation convention*. Independently for each deterministic phase $k \in [K_T]$, attach an auxiliary i.i.d. stream of environment draws. If the meta-strategy ever stops before exhausting some same-price or audit slots of phase k, or if it never reaches phase k at all, then those unused slots are filled from this auxiliary stream and are used only for analysis. Because the phase learner never sees audit outcomes, this continuation extends every phase to a full latent same-price trajectory of length n_k and a full latent audit family

$$Y_{k,1}, \dots, Y_{k,J_k},$$

without altering the realized algorithm or its regret.

For every phase $k \in [K_T]$ and checkpoint $j \in [J_k]$, let $\bar{g}_{k,j}$ be the true average of g_ν over the j-th latent audited same-price block of phase k. The statistic $\widehat{G}_{k,j}$ continues to denote the running checkpoint estimator, evaluated on the latent continuation of phase k, and define

$$G_{k,j} := \frac{1}{j^2} \sum_{\ell=1}^{j} (2\ell - 1)\bar{g}_{k,\ell}. \tag{23}$$

Let $\mathcal{F}_{k,j}^{\mathrm{play}}$ be the sigma-field generated by the latent same-price play in the first j audited blocks of phase k. We also introduce the global latent-audit event

$$\mathcal{E}_{\mathrm{aud}} := \bigcap_{k=1}^{K_T} \bigcap_{j=1}^{J_k} \left\{ |\widehat{G}_{k,j} - G_{k,j}| \le \beta_{k,j} \right\}.$$

This event will be used in both the generic-side and IV-side proofs.

Lemma 24 (Small-horizon reduction)**.** *If $T < B_\star$, then the default policy used in the definition of $\mathsf{VAM}_T(V)$ already satisfies both target bounds from Theorem 3. More precisely, for every $\nu \in \mathcal{H}(M)$,*

$$R_T(\mathsf{VAM}_T(V); \nu) \le C_M (\log(T+2))^2 T^{2/3} \quad \textit{and} \quad R_T(\mathsf{VAM}_T(V); \nu) \le C'_M (\log(T+2))^2 T^{3/4}$$

for constants C_M, C'_M depending only on M.

Proof. In this regime the algorithm plays a fixed same price on all T rounds, so $R_T(\mathsf{VAM}_T(V);\nu) \le T$. Since $T < B_\star \le C_M(\log(T+2))^6$, we obtain

$$T = T^{2/3}T^{1/3} \le C_M(\log(T+2))^2 T^{2/3}$$

and similarly

$$T = T^{3/4}T^{1/4} \le C'_M(\log(T+2))^2 T^{3/4}.$$

Thus both target bounds hold after adjusting constants. □

C.4 Generic-side analysis

Theorem 25 (Generic-side guarantee)**.** *Fix $M > 0$. There exists a constant $C_{\mathrm{all}}(M)$ such that for every $\nu \in \mathcal{H}(M)$,*

$$R_T(\mathsf{VAM}_T(V);\nu) \le C_{\mathrm{all}}(M)\,\mathrm{polylog}(T)\,T^{3/4}.$$

The generic-side proof follows the meta-strategy logic. On the global audit event, every phase that finishes without rejection must end with a small average value deficit at its last checkpoint, hence it contributes only its target phase cost. If a phase rejects, the same checkpoint bookkeeping bounds the regret already spent before the switch, and the remainder is delegated to the robust fallback. Summing over the geometric phase budgets then yields the theorem.

Proof. By Lemma 24, it suffices to treat the regime $T \ge B_\star$, so from now on we work with the phase decomposition of Section C.3.

For every deterministic phase $k \in [K_T]$, the latent continuation convention identifies a full latent sequence of phase-k block averages and block audits

$$(\bar g_{k,1}, Y_{k,1}), \dots, (\bar g_{k,J_k}, Y_{k,J_k}).$$

By Lemma 19, for every phase k and every checkpoint $j \in [J_k]$,

$$\mathbb{P}\big(|\widehat{G}_{k,j} - G_{k,j}| > \beta_{k,j} \mid \mathcal{F}^{\mathrm{play}}_{k,j}\big) \le 2e^{-L}.$$

By Lemma 23,

$$\sum_{k=1}^{K_T} J_k = O(\sqrt{T}),$$

so a union bound yields

$$\mathbb{P}(\mathcal{E}_{\mathrm{aud}}) \ge 1 - C\sqrt{T}\,e^{-L} \ge 1 - \delta/2$$

for C_0 sufficiently large. On $\mathcal{E}^c_{\mathrm{aud}}$ we use the trivial pathwise bound

$$\mathfrak{B}_T(\mathsf{VAM}_T(V);\nu) \le T,$$

so its expected contribution to $R_T(\mathsf{VAM}_T(V);\nu) = \mathbb{E}[\mathfrak{B}_T(\mathsf{VAM}_T(V);\nu)]$ is negligible.

Work on $\mathcal{E}_{\mathrm{aud}}$.

Step 1: passed phases. Fix a phase k with $n := n_k$ that completes without rejection. Write

$$J := J_k, \qquad r := r_k.$$

The final checkpoint J is then reached. Because it did not reject,

$$\widehat{\delta}_{k,J} \le \Theta_k + 8\beta_{k,J}.$$

On $\mathcal{E}_{\mathrm{aud}}$ this implies

$$\delta_{k,J} := V - G_{k,J} \le \widehat{\delta}_{k,J} + \beta_{k,J} \le \Theta_k + 9\beta_{k,J}.$$

Therefore the cumulative regret of the J^2 audited suffix same-price rounds is at most

$$J^2\delta_{k,J} \le J^2\Theta_k + 9J^2\beta_{k,J} \le C\big(Ln^{2/3} + \sqrt{L}\,n^{3/4}\big).$$

The burn-in contributes at most

$$w_k \le 4n^{2/3}.$$

Also,

$$r \le 2\sqrt{n}, \qquad J \le \sqrt{n}.$$

So every passed phase contributes at most

$$C\big(Ln^{2/3} + \sqrt{L}\, n^{3/4} + \sqrt{n}\big).$$

Because $n = n_k \ge B_k/2 \ge B_\star/2 \ge \frac{1}{2} C_{\mathrm{base}} L^6$ and C_{base} is chosen large, one has $\sqrt{n} \le Ln^{2/3}$. Thus every passed phase contributes at most

$$C\big(Ln^{2/3} + \sqrt{L}\, n^{3/4}\big). \tag{24}$$

Step 2: the rejected phase, if any. Suppose phase k, with $n := n_k$, rejects for the first time at checkpoint $j^* \in \{j_{0,k}, \dots, J_k\}$.

If $j^* = j_{0,k}$, then before rejection the phase has spent at most

$$w_k + j_{0,k}^2 + j_{0,k} \le C\big(n^{2/3} + L^2 n^{2/3}\big) \le CL^2 n^{2/3}$$

regret, since each real round costs at most 1.

If $j^* > j_{0,k}$, then checkpoint $j^* - 1$ was reached and did not reject. Thus

$$\widehat{\delta}_{k,j^*-1} \le \Theta_k + 8\beta_{k,j^*-1}.$$

On $\mathcal{E}_{\mathrm{aud}}$,

$$\delta_{k,j^*-1} := V - G_{k,j^*-1} \le \Theta_k + 9\beta_{k,j^*-1}.$$

Hence the cumulative regret of the first $(j^*-1)^2$ audited suffix same-price rounds is at most

$$(j^*-1)^2 \delta_{k,j^*-1} \le C\big(Ln^{2/3} + \sqrt{L}\,(j^*)^{3/2}\big) \le C\big(Ln^{2/3} + \sqrt{L}\, n^{3/4}\big).$$

The burn-in, the already incurred audits from the first $j^* - 1$ blocks, the current block, and its audit add at most

$$w_k + (j^*-1) + (2j^*-1) + 1 \le 4n^{2/3} + 3\sqrt{n},$$

hence every rejected phase contributes at most

$$C\big(L^2 n^{2/3} + \sqrt{L}\, n^{3/4}\big) \tag{25}$$

before the switch.

After rejection, the algorithm invokes RF_H on the remaining real horizon H. Because the environment is i.i.d. across real rounds and every unreached latent variable used to define $\mathcal{E}_{\mathrm{aud}}$ comes from an auxiliary stream independent of the future real interaction, the future real rounds seen by the fallback are conditionally i.i.d. from ν given the realized pre-switch history. Hence, if $\mathfrak{R}_H^{\mathrm{fb}}$ denotes the realized regret incurred by the fallback on those remaining H real rounds, then conditional on the switch history,

$$\mathbb{E}\big[\mathfrak{R}_H^{\mathrm{fb}} \mid \text{switch history}\big] = R_H(\mathsf{RF}_H; \nu) \le C_{\mathrm{gen}}(M)\, \mathrm{polylog}(T)\, T^{3/4},$$

by Proposition 2, because $H \le T$ and in the present regime $T \ge B_\star \ge 16$. The same bound therefore remains valid after additionally conditioning on $\mathcal{E}_{\mathrm{aud}}$, since on the latent continuation convention the event $\mathcal{E}_{\mathrm{aud}}$ is measurable with respect to the switch history together with auxiliary variables independent of those future real rounds.

Step 3: sum over phases. There is at most one rejected phase. Summing (24) over all passed phases and, if present, (25) over the single rejected phase, and using Lemma 23(iii), gives

$$\sum_{k=1}^{K_T} \big(L^2 n_k^{2/3} + \sqrt{L}\, n_k^{3/4}\big) \le C' \mathrm{polylog}(T)\, T^{3/4}.$$

On the event of no rejection, the pathwise pre-switch bound already equals the full regret. On the event of a rejection, we add the conditional expected fallback regret bound above. Therefore

$$\mathbb{E}_\nu[\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu) \mid \mathcal{E}_{\mathrm{aud}}] \le C'' \mathrm{polylog}(T)\, T^{3/4}.$$

Since $R_T(\mathsf{VAM}_T(V); \nu) = \mathbb{E}_\nu[\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu)]$, adding the negligible contribution of $\mathcal{E}_{\mathrm{aud}}^c$ proves the theorem. □

C.5 Phase-law coupling inside the meta-strategy

The phase analysis above concerns a fresh run of PhaseScout over n_k same-price rounds. To use Proposition 4 inside the meta-strategy, we must verify that the reserved same-price rounds of phase k have exactly the same phase-history law.

Fix a product environment $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$ and a phase k. Inside the deterministic phase-k block of length B_k, the meta-strategy reserves exactly n_k same-price slots for the phase-k call to PhaseScout and exactly J_k audit slots interleaved after the odd blocks. If the meta-strategy would reject phase k early, then the unused same-price and audit slots of that phase are filled, for analysis only, from the auxiliary continuation stream of phase k introduced above. Thus the phase-k call to PhaseScout is analytically continued through all n_k same-price slots and all J_k audit slots without feeding any of those latent future outcomes back into the actual algorithm.

Let

$$r_{k,1} < r_{k,2} < \cdots < r_{k,n_k}$$

be the deterministic sequence of reserved same-price real rounds of phase k, listed in increasing order. For each phase index $s \le n_k$, let $P_{k,s}^{\mathrm{emb}}$ be the same price chosen by the phase-k learner at that latent phase round. Let

$$(\widetilde{S}_{k,s}, \widetilde{B}_{k,s})$$

denote the latent environment sample attached to that same-price slot: it equals the real environment draw

$$(S_{r_{k,s}}, B_{r_{k,s}})$$

when the slot occurs before any actual switch, and otherwise it equals the next unused draw from the auxiliary continuation stream of phase k. Define the associated phase feedback

$$Y_{k,s}^{B,\mathrm{emb}} := \mathbf{1}\{\widetilde{B}_{k,s} \ge P_{k,s}^{\mathrm{emb}}\}, \qquad Y_{k,s}^{S,\mathrm{emb}} := \mathbf{1}\{\widetilde{S}_{k,s} \le P_{k,s}^{\mathrm{emb}}\}.$$

Let

$$\mathsf{H}_{k,s}^{\mathrm{emb}} := \left(P_{k,1}^{\mathrm{emb}}, Y_{k,1}^{B,\mathrm{emb}}, Y_{k,1}^{S,\mathrm{emb}}, \ldots, P_{k,s}^{\mathrm{emb}}, Y_{k,s}^{B,\mathrm{emb}}, Y_{k,s}^{S,\mathrm{emb}}\right)$$

be the phase history generated inside the meta-strategy up to phase round s.

Let

$$\mathsf{H}_{k,s}^{\mathrm{ref}} := \left(P_{k,1}^{\mathrm{ref}}, Y_{k,1}^{B,\mathrm{ref}}, Y_{k,1}^{S,\mathrm{ref}}, \ldots, P_{k,s}^{\mathrm{ref}}, Y_{k,s}^{B,\mathrm{ref}}, Y_{k,s}^{S,\mathrm{ref}}\right)$$

be the phase history of a fresh run of $\mathsf{PhaseScout}_{n_k,\eta_k}$ under the same product environment.

Lemma 26 (Phase-law identity)**.** *Fix a phase k and condition on the meta-strategy history up to the start of phase k. Then for every latent phase round $s \le n_k$,*

$$\mathsf{H}_{k,s}^{\mathrm{emb}} \overset{d}{=} \mathsf{H}_{k,s}^{\mathrm{ref}}.$$

Consequently, every measurable event or random variable depending only on the latent phase-k history up to round s has the same law in the actual phase execution and in the fresh phase run. In particular, Proposition 4 applies to the actual same-price trajectory used by Algorithm 2 on every event on which phase k reaches the relevant phase round.

Proof. We argue by induction on s. For $s = 0$ both histories are empty. Assume the claim holds up to round $s-1$. By construction, the phase-k learner is the same measurable map in the actual phase execution and in the fresh phase run; therefore its next price is a deterministic function of the previous phase history:

$$P_{k,s}^{\mathrm{emb}} = \Phi_s(\mathsf{H}_{k,s-1}^{\mathrm{emb}}), \qquad P_{k,s}^{\mathrm{ref}} = \Phi_s(\mathsf{H}_{k,s-1}^{\mathrm{ref}}).$$

By the induction hypothesis, these next prices have the same conditional law.

Now condition on the common next price value p. In the actual phase execution, the designated same-price slot at phase time s uses a fresh i.i.d. sample: it is the real environment sample $(S_{r_{k,s}}, B_{r_{k,s}})$ when that slot occurs before any actual switch, and otherwise it is the next draw from the auxiliary continuation stream of phase k. In either case this sample is independent of the prior phase history because the environment is i.i.d., the continuation stream is independent, and the audit slots are disjoint from the same-price feedback fed to PhaseScout. Hence

$$\left(Y_{k,s}^{B,\mathrm{emb}}, Y_{k,s}^{S,\mathrm{emb}}\right) \mid \{P_{k,s}^{\mathrm{emb}} = p\}$$

has the phase feedback law

$$Y^B = \mathbf{1}\{B \geq p\}, \qquad Y^S = \mathbf{1}\{S \leq p\}.$$

This is exactly the feedback kernel used by the fresh phase run at price p. Therefore the conditional transition kernel from phase round $s-1$ to s is identical in the actual phase execution and in the fresh phase run. The induction closes, proving the equality in distribution for every $s \leq n_k$. □

C.6 IV-side analysis

Theorem 27 (IV-side guarantee)**.** *Fix $M > 0$. There exists a constant $C_{\mathrm{IV,all}}(M)$ such that for every $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$,*

$$R_T(\mathsf{VAM}_T(V); \nu) \leq C_{\mathrm{IV,all}}(M)\,\mathrm{polylog}(T)\,T^{2/3}.$$

The IV-side proof uses the same decomposition but a different invariant. We combine the audit-concentration event with the all-prefix guarantee of PhaseScout. On that intersection, every reached checkpoint stays below threshold, so the meta-strategy never switches away from the IV routine. The total regret is then just the sum of the phasewise $T^{2/3}$ costs of PhaseScout plus the lower-order audit overhead.

Proof. Fix $\nu \in \mathcal{E}_{\mathrm{IV}}(M)$. By Lemma 24, it suffices to treat the regime $T \geq B_\star$, so from now on we work with the phase decomposition of Section C.3.

For each deterministic phase $k \in [K_T]$, let L_{n_k} be the logarithmic factor of Proposition 4. By the construction of n_k, the bound $n_k \leq T$, and the choice $\eta_k = \delta/(16k^2)$, one has

$$L_{n_k} \leq L$$

after enlarging the constant C_0 if necessary.

For every deterministic phase $k \in [K_T]$, let $\mathcal{E}_k^{\mathrm{IV}}$ be the event that the *latent* phase-k same-price trajectory from the continuation convention satisfies the phase-prefix bound of Proposition 4. If phase k is actually reached, then Lemma 26 shows that, conditional on the meta-strategy history up to the start of phase k, this latent phase history has the same law as a fresh run of $\mathsf{PhaseScout}_{n_k,\eta_k}$. If phase k is never reached, then by construction its latent continuation is generated entirely from an independent auxiliary i.i.d. stream and therefore is itself a fresh copy of $\mathsf{PhaseScout}_{n_k,\eta_k}$. Hence in all cases

$$\mathbb{P}\big((\mathcal{E}_k^{\mathrm{IV}})^c\big) \leq \eta_k = \frac{\delta}{16k^2}.$$

Define

$$\mathcal{E}_{\mathrm{good}} := \mathcal{E}_{\mathrm{aud}} \cap \bigcap_{k=1}^{K_T} \mathcal{E}_k^{\mathrm{IV}}.$$

Since

$$\sum_{k=1}^{\infty} \eta_k = \frac{\delta}{16} \sum_{k=1}^{\infty} \frac{1}{k^2} < \frac{\delta}{8},$$

and $\mathbb{P}(\mathcal{E}_{\mathrm{aud}}^c) \leq \delta/2$, we have

$$\mathbb{P}(\mathcal{E}_{\mathrm{good}}) \geq 1 - \delta$$

after enlarging C_0 if necessary. On $\mathcal{E}_{\mathrm{good}}^c$ we use the trivial pathwise bound

$$\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu) \leq T,$$

whose contribution to $R_T(\mathsf{VAM}_T(V); \nu) = \mathbb{E}[\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu)]$ is negligible.

Work on $\mathcal{E}_{\mathrm{good}}$.

Step 1: phases never reject. Fix a phase k and let

$$j \in \{j_{0,k}, \ldots, J_k\}$$

be a reached checkpoint. By Proposition 4, applied to the embedded phase-k subroutine with prefix length

$$m = j^2,$$

the true average target gap over the first j^2 audited suffix same-price rounds satisfies

$$\delta_{k,j} := \frac{1}{j^2} \sum_{t=w_k+1}^{w_k+j^2} \big(V - g_\nu(X_t)\big) \le A_{\mathrm{IV}} L_{n_k} \left(n_k^{-1/3} + n_k^{1/6} j^{-1} \right).$$

Since $j \ge j_{0,k} = \left\lceil C_j L\, n_k^{1/3} \right\rceil$, one has

$$\sqrt{\frac{L}{j}} \le C_j^{-1/2} n_k^{-1/6}$$

and therefore, because $C_j \ge 64 A_{\mathrm{IV}}^2$,

$$A_{\mathrm{IV}} L_{n_k} n_k^{1/6} j^{-1} \le A_{\mathrm{IV}} \frac{L}{j} n_k^{1/6} \le A_{\mathrm{IV}} C_j^{-1/2} \sqrt{\frac{L}{j}} \le \frac{1}{8} \sqrt{\frac{L}{j}} \le \beta_{k,j},$$

after enlarging the universal constant c_0 if necessary. Therefore, because $C_\Theta \ge 32 A_{\mathrm{IV}}$,

$$\delta_{k,j} \le A_{\mathrm{IV}} L n_k^{-1/3} + \beta_{k,j} \le \frac{\Theta_k}{32} + \beta_{k,j}.$$

On $\mathcal{E}_{\mathrm{aud}}$,

$$\widehat{\delta}_{k,j} \le \delta_{k,j} + \beta_{k,j} \le \Theta_k + 8\beta_{k,j}.$$

Hence the rejection condition of Algorithm 2 never triggers at any reached checkpoint of any phase. So on $\mathcal{E}_{\mathrm{good}}$ the robust fallback is never called.

Step 2: total regret on $\mathcal{E}_{\mathrm{good}}$. Since fallback is never called, every deterministic phase block runs to completion. Thus, on $\mathcal{E}_{\mathrm{good}}$, the benchmark-gap process satisfies

$$\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu) \le \sum_{k=1}^{K_T} \Big(A_{\mathrm{IV}} L\, n_k^{2/3} + J_k \Big).$$

Indeed, for every phase, the n_k same-price rounds of PhaseScout contribute at most $A_{\mathrm{IV}} L\, n_k^{2/3}$ by Proposition 4 and Lemma 26, and the J_k audits each cost at most 1. Applying Lemma 23(iii) and (iv) gives

$$\mathfrak{B}_T(\mathsf{VAM}_T(V); \nu) \le A_{\mathrm{IV}} L\, O(T^{2/3}) + O(\sqrt{T}) \le C \mathrm{polylog}(T)\, T^{2/3}.$$

Taking expectations on $\mathcal{E}_{\mathrm{good}}$ and adding the negligible bad-event contribution proves the theorem. □

C.7 Completion of the proof of Theorem 3

Proof. The hard-side guarantee follows from Theorem 25, and the IV-side guarantee follows from Theorem 27. Combining these two bounds yields exactly the statement of Theorem 3. □

D A Quantitative View of Weak Dependence

A quantitative view of weak dependence. The independent-versus-dependent dichotomy can be refined using the absolute-regularity distance from independence

$$\beta(\nu) := \|\nu - \nu_S \otimes \nu_B\|_{\mathrm{TV}},$$

where ν_S and ν_B are the seller and buyer marginals. For any reward bounded in $[0,1]$, this distance controls the expectation gap; in particular, $\sup_{x \in [0,1]} |g_\nu(x) - g_{\nu_S \otimes \nu_B}(x)| \le \beta(\nu)$. The existing surrogate-consistency algorithm therefore satisfies, for $\alpha \in [2/3, 3/4]$,

$$R_T(\nu) \le \widetilde{O}\Big(T^\alpha + T \min\{\beta(\nu), T^{-\alpha/3}\} \Big).$$

Conversely, any algorithm with $O(T^{\alpha})$ regret on independent-values instances obeys, for each dependence budget $d \in (0, 1)$,

$$\sup_{\nu:\, \beta(\nu) \le d} R_T(\nu) \ge \Omega\Big(T \min\{d, T^{-\alpha/3}\}\Big) .$$

Thus the excess regret grows linearly with $T\beta(\nu)$ below the detection scale $T^{-\alpha/3}$ and saturates at the adaptation lower-bound scale above it (up to logarithmic factors). With the exact value hint, the value-auditing algorithm also admits the uniform interpolation

$$R_T(\nu) \le \widetilde{O}\Big(\min\{T^{2/3} + T\beta(\nu),\, T^{3/4}\}\Big) .$$

These bounds show that weak dependence does not force an abrupt transition to the general-class rate. The dependence coefficient is a distributional upper bound; the sharper, decision-specific quantity is the uniform discrepancy between the true and product-surrogate gain curves.

Declaration of LLM Usage

We acknowledge the use of large language models (LLMs) in the preparation of this work. Specifically, LLMs were employed to assist with proof elaboration, writing refinement, and literature survey. It is important to note that while certain proof details were elaborated with LLM assistance, the correctness verification and error correction of all proofs were independently conducted by the authors. Furthermore, all core research ideas, theoretical contributions, and algorithmic designs presented in this paper are entirely the original work of the authors.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 state the paper's three contributions: the no-go frontier, the constructive tradeoff algorithm, and the known-value best-of-both-worlds recovery; Sections 3 to 5 and Sections A to C give the corresponding formal statements and proofs.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The end of 7 contains a dedicated scope-and-limitations paragraph.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.

- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumptions are stated in Sections 2 to 5, and complete proofs are provided in Sections A to C with numbered lemmas, propositions, and theorems cross-referenced from the main text.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: The paper is purely theoretical and does not report experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.

- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A]

   Justification: The paper does not rely on experimental datasets or empirical code artifacts for its main claims.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [N/A]

Justification: The paper contains no training pipeline or empirical evaluation.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [N/A]

Justification: The paper contains no experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [N/A]

Justification: The paper contains no experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: This work is a theoretical analysis of online bilateral trade and does not involve human subjects, personal data, or unsafe deployment procedures; we are not aware of any aspect that conflicts with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper studies regret guarantees and impossibility tradeoffs in an abstract online-learning model and does not introduce a deployable system, dataset, or model release; any downstream societal impact is indirect.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release data or models of the kind contemplated by this question.

    Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use external code, datasets, or model checkpoints as research assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new datasets, code packages, or models as research assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or human-subject research.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human-subject research.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: LLMs were used to assist with proof elaboration, writing refinement, and literature survey. The correctness verification and error correction of all proofs, as well as all core research ideas, theoretical contributions, and algorithmic designs, are the sole work of the authors.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.