Optimal Contextual Pricing under Agnostic Non-Lipschitz Demand

Jianyu Xu
Department of Machine Learning, UNC Charlotte
Charlotte, NC, USA
jxu22@charlotte.edu

Yu-Xiang Wang
Halıcıoğlu Data Science Institute, UC San Diego
La Jolla, CA, USA
yuxiangw@ucsd.edu

Abstract

We study contextual dynamic pricing with linear valuations and bounded-support agnostic noise, whose induced demand curve may be non-Lipschitz with arbitrary jumps and atoms. Such discontinuities break the cross-context interpolation arguments used by smooth-demand pricing algorithms, while the best previous method achieved only $\widetilde{O}(T^{3/4})$ regret. We propose Conservative-Markdown Redirect-UCB Pricing, a polynomial-time algorithm that combines randomized parameter estimation, conservative residual-grid probing, and confidence-based one-step redirection. Our algorithm achieves $\widetilde{O}(T^{2/3})$ optimal regret, matching the known lower bounds of Kleinberg and Leighton [2003] up to logarithmic factors and improving over the previous upper bound of Xu and Wang [2022]. Under stochastic well-conditioned contexts, this closes the longstanding open regret gap in linear-valuation contextual pricing under agnostic non-Lipschitz noise distributions.

1 Introduction

Learning from censored feedback is a recurring challenge across machine learning: the learner makes a decision, but observes only a partial outcome rather than the full reward landscape. Contextual dynamic pricing is a canonical instance of this challenge. A seller observes a feature vector describing the current buyer or product, posts a price in real time, and observes only whether the buyer purchases. The seller never observes the buyer's valuation itself. The goal is to learn a pricing policy whose cumulative revenue is close to that of a clairvoyant policy that knows both the linear valuation parameter and the noise distribution.

A standard model writes the valuation as a linear contextual component plus an idiosyncratic noise:

Contextual pricing. For $t = 1, 2, \ldots, T$:
1. A context $x_t \in \mathbb{R}^d$ is observed.
2. The buyer has private valuation $y_t = x_t^\top \theta^\star + \xi_t$.
3. The seller posts a price $p_t \in [0, B]$.
4. The seller observes only $o_t = \mathbf{1}\{p_t \le y_t\}$.
5. The seller receives revenue $p_t o_t$.

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

Table 1: Comparison of dynamic pricing regret bounds under different noise assumptions.

Noise distribution assumptions	Best known regret	Reference
$O(1/T)$ standard deviation	$O(\log T)$	Cohen et al. [2020]
known and log-concave	$O(\log T)$	Javanmard and Nazerzadeh [2019]
Parametric Family	$\widetilde{O}(\sqrt{T})$	Ban and Keskin [2021]
Unknown, second-order smooth	$\widetilde{O}(T^{2/3})$	Luo et al. [2022]
Unknown, kth-order smooth	$\widetilde{O}(T^{\frac{2k+1}{4k-1}})$	Fan et al. [2024]
Unknown, Lipschitz	$\widetilde{O}(T^{2/3})$	Tullii et al. [2024]
Unknown, Lipschitz	$\widetilde{\Omega}(T^{2/3})$	Xu and Wang [2022]
Agnostic, bounded	$\widetilde{O}(T^{3/4})$	Xu and Wang [2022]
Agnostic, non-contextual	$\Omega(T^{2/3})$	Kleinberg and Leighton [2003]
Agnostic, bounded	$\widetilde{\mathbf{O}}(\mathbf{T^{2/3}})$	Theorem 1, **this work**

The statistical difficulty of the problem depends strongly on the regularity of the noise distribution. If the noise distribution is known, parametric, log-concave, Lipschitz, or sufficiently smooth, existing algorithms can exploit this structure to transfer information across nearby residual prices. In many pricing environments, however, such regularity is not realistic. Buyers may cluster around psychological thresholds, reservation values may contain mass points, and heterogeneous subpopulations may create sharp jumps in the demand curve. In these cases, the survival function may be discontinuous and non-Lipschitz. A small residual perturbation can then change the purchase probability by a constant amount, invalidating interpolation arguments based on smoothness.

This paper studies contextual dynamic pricing in this fully agnostic non-Lipschitz regime. We impose no density, Lipschitz, smoothness, log-concavity, single-crossing, or parametric condition on the noise distribution beyond mild boundedness and i.i.d. assumptions. The noise may have arbitrary atoms and jumps. The formal setup and assumptions are given in Section 3.

Main contribution. We introduce *Conservative-Markdown Redirect-UCB Pricing*, a polynomial-time algorithm that achieves $\widetilde{O}(T^{2/3})$ regret for linear-valuation contextual pricing with fully agnostic bounded noise. This rate is optimal in its T-dependence up to logarithmic factors. A $T^{2/3}$ lower bound already holds for an interior non-contextual subclass [Kleinberg and Leighton, 2003] that can be embedded into our stochastic-context model, and Xu and Wang [2022] establish the same barrier for contextual pricing under Lipschitz demand. Therefore, our result closes the $T^{3/4}$ versus $T^{2/3}$ gap for agnostic linear-valuation pricing under the stochastic contextual setting studied here.

Scope of agnosticism. Our agnosticism is with respect to the residual demand curve: we impose no Lipschitz continuity, smoothness, density, log-concavity, or parametric structure on the valuation noise distribution, but only boundedness to avoid boundary clipping.

Algorithmic idea. The algorithm reduces contextual pricing to confidence-bound learning on a one-dimensional residual grid. A uniformly randomized first stage estimates the linear valuation component; subsequent prices use a conservative markdown so that every binary observation becomes a one-sided probe of two adjacent survival levels, a relation that remains valid even when the demand curve has jumps. In the adaptive stage, an upper-confidence rule probes uncertain grid points directly and redirects to the left once their confidence radius reaches the grid scale, so discontinuity costs are paid only during a limited exploration period.

Technical contribution. The analysis compares the oracle value directly with the optimistic score of a maintained residual-grid index, rather than estimating demand at oracle-selected points. Direct probes pay grid, confidence, and adjacent-jump costs, while redirected probes pay only a grid-scale price loss; summing these terms with $\Delta \asymp T^{-1/3}$ yields $\widetilde{O}(T^{2/3})$ regret. The same residual grid has size $\widetilde{O}(T^{1/3})$, giving a direct polynomial-time implementation.

Table 1 summarizes the regret landscape for linear-valuation contextual pricing under different noise assumptions. The rest of the paper is organized as follows. Section 2 reviews related work. Section 3

formalizes the model and assumptions. Section 4 presents the algorithm. Section 5 gives the main regret theorem and proof roadmap. Section 6 reports numerical experiments, and Section 7 discusses limitations, ethical considerations, and conclusions. Full proofs appear in the appendix.

2 Related Work

We review the closest literature on contextual dynamic pricing with unknown demand. Additional background on parametric pricing, non-contextual posted-price learning, confidence-bound methods, and adjacent contextual pricing models appears in Appendix A.

Contextual pricing with smooth or Lipschitz noise. A large body of work obtains strong regret guarantees by imposing regularity on the noise distribution. Javanmard and Nazerzadeh [2019] study dynamic pricing with known log-concave noise, where the revenue function is well behaved enough for greedy pricing after parameter learning. Ban and Keskin [2021] obtain $\widetilde{O}(\sqrt{T})$ regret under parametric demand assumptions. Luo et al. [2022] achieve $\widetilde{O}(T^{2/3})$ regret under second-order smoothness of the noise distribution through an explore-then-UCB strategy. Fan et al. [2024] exploit higher-order smoothness and introduce a semiparametric parameter-estimation method based on uniformly randomized prices, which we also use in Stage 1. Tullii et al. [2024] obtain $\widetilde{O}(T^{2/3})$ regret under Lipschitz noise via cross-context learning. These works use smoothness to transfer demand information between nearby residuals. Such transfer is unavailable in the agnostic setting considered here, where a residual perturbation of arbitrarily small size may cross an atom and change demand by a constant amount.

Fully agnostic contextual pricing. The closest work is Xu and Wang [2022], who study linear-valuation contextual pricing with an unknown noise distribution and no smoothness assumptions. Their D2-EXP4 algorithm discretizes a large policy class and applies an adversarial bandit method, obtaining $\widetilde{O}(T^{3/4})$ regret for adversarial context sequences. They also prove a $\widetilde{\Omega}(T^{2/3})$ lower bound for contextual pricing under a Lipschitz subclass, which applies to the larger agnostic model. Our algorithm works directly on a one-dimensional residual grid rather than enumerating a high-dimensional policy class. The conservative markdown creates a stable one-sided observation model on this grid, and the redirect-UCB rule controls discontinuity-induced bias through confidence-radius accounting. This yields the optimal $\widetilde{O}(T^{2/3})$ regret rate in polynomial time.

Lower bounds. The $\widetilde{O}(T^{2/3})$ rate is tight in its dependence on T. Xu and Wang [Xu and Wang, 2022] prove a $\widetilde{\Omega}(T^{2/3})$ lower bound for contextual pricing under Lipschitz noise, showing that the $T^{2/3}$ barrier already appears in a regular subclass of demand functions. In addition, the classical non-contextual lower bound of Kleinberg and Leighton [2003] can be embedded into our model by using a constant linear valuation together with independent dummy covariates to satisfy the full-rank stochastic-context condition. The hard valuation distributions can be placed in an interior price interval, so bounded support, zero-mean normalization, and the price-buffer condition are satisfied after the usual recentering of the noise into the intercept. Appendix C.10 gives the formal embedding.

3 Problem Setup

We study contextual dynamic pricing with linear valuations and binary purchase feedback. At each round $t = 1, \ldots, T$, the seller observes a context $x_t \in \mathbb{R}^d$. The customer's valuation is $y_t = \langle x_t, \theta^\star \rangle + \xi_t$, where $\theta^\star \in \mathbb{R}^d$ is unknown and ξ_t is an unobserved noise. The seller posts a price $p_t \in [0, B]$ based on past observations and observes only $o_t = \mathbf{1}\{p_t \le y_t\}$.

We write the demand curve in residual coordinates through the survival function $S(w) := \mathbb{P}(\xi_t \ge w)$. This convention handles atoms without additional notation: if $F(w) = \mathbb{P}(\xi_t \le w)$ is the usual right-continuous CDF, then $S(w) = 1 - F(w^-)$, so $\mathbb{E}[o_t \mid x_t, p_t] = S(p_t - \langle x_t, \theta^\star \rangle)$ for arbitrary noise distributions. We extend S outside the support by setting $S(w) = 1$ for $w \le -c$ and $S(w) = 0$ for $w > c$.

The expected revenue at context x and price p is $\pi(x,p) := pS(p - \langle x, \theta^\star \rangle)$. Let

$$\mathcal{R}_T := \sum_{t=1}^{T} \left\{ \max_{p \in [0,B]} \pi(x_t, p) - \pi(x_t, p_t) \right\}, \qquad R_T := \mathbb{E}[\mathcal{R}_T]. \tag{1}$$

For compactness, write $u_t := \langle x_t, \theta^\star \rangle$ and $V_t(w) := (u_t + w)S(w)$. The oracle value in round t is $V_t^\star := \max_{w \in [-c,c]} V_t(w)$; residuals outside $[-c, c]$ cannot improve revenue under the bounded-support model.

3.1 Assumptions

We impose boundedness and identifiability conditions on the linear contextual component, while leaving the demand curve otherwise unrestricted: S may be discontinuous, non-Lipschitz, and may contain arbitrary jumps and atoms.

Assumption 1 (Bounded support, exogeneity, and normalization)**.** *There exist known constants $B_x, B_\theta, c, B < \infty$ such that, almost surely, $\|x_t\|_2 \le B_x$, $\|\theta^\star\|_2 \le B_\theta$, $\xi_t \in [-c, c]$, and $0 \le y_t \le B$. The noises $\{\xi_t\}_{t=1}^T$ are i.i.d. from an unknown distribution F and independent of the context sequence. We use the location normalization $\mathbb{E}[\xi_t] = 0$.*

The valuation bound $0 \le y_t \le B$ is used only in Stage 1 to convert uniformly randomized prices into an unbiased linear signal. The zero-mean normalization is without loss of generality when the context contains an intercept, since a nonzero noise mean can be absorbed into the intercept coefficient. The independence condition can be weakened to $\mathbb{E}[\xi_t \mid x_t] = 0$ for the Stage 1 argument.

Assumption 2 (Stochastic contexts)**.** *The contexts $\{x_t\}_{t=1}^T$ are i.i.d. from a distribution $\mathcal{D}_x$ supported on $\{x : \|x\|_2 \le B_x\}$, and $\lambda_{\min}(\mathbb{E}[x_t x_t^\top]) \ge \lambda_0 > 0$.*

Assumption 2 ensures that the linear parameter is identifiable from the randomized Stage 1 observations.

Assumption 3 (Buffered admissible prices)**.** *There exists a constant $\kappa > 0$ such that $\langle x_t, \theta^\star \rangle - c \ge \kappa$ and $\langle x_t, \theta^\star \rangle + c \le B$ almost surely. We assume the horizon is large enough that $2\Delta \le \kappa$, where Δ is the Stage 1 target accuracy in Algorithm 1; smaller horizons are absorbed into the problem-dependent constant.*

Scope of the assumptions. Our theorem assumes bounded noise and valuations, i.i.d. well-conditioned contexts, and a fixed interior price buffer $\kappa > 0$. The context condition enables Stage 1 estimation, while the buffer prevents clipping of the markdown probes; neither condition imposes continuity or smoothness on the noise distribution. The $\widetilde{O}(T^{2/3})$ optimality claim concerns the horizon dependence for fixed dimension and problem parameters.

3.2 Notation

Let $\hat{u}_t := \langle x_t, \widehat{\theta} \rangle$ be the predicted linear value after Stage 1, and define the Stage 1 good event

$$\mathcal{E}_\theta := \left\{ \max_{1 \le t \le T} |\hat{u}_t - u_t| \le \Delta \right\}. \tag{2}$$

Given Δ, define a residual grid with spacing 2Δ by $M := \lceil c/\Delta \rceil$ and $w_i := -c + 2i\Delta$ for $i = 0, 1, \ldots, M$. The queried index set is $\mathcal{I} := \{0, 1, \ldots, M-1\}$. We use the boundary convention $w_{-1} := w_0 - 2\Delta$ and $S(w_{-1}) := 1$.

For each queried index $j \in \mathcal{I}$, define the adjacent jump height $\alpha_j := S(w_{j-1}) - S(w_j) \ge 0$. Since S is non-increasing and bounded in $[0, 1]$, $\sum_{j=0}^{M-1} \alpha_j \le 1$. This quantity captures the maximum discontinuity cost associated with probing the interval $[w_{j-1}, w_j]$. Let $\mathcal{H}_t$ denote the sigma-algebra generated by observations through round t. We use $\widetilde{O}(\cdot)$ to hide factors polylogarithmic in T and polynomial in $(d, B_x, B_\theta, B, c, 1/\lambda_0, 1/\kappa)$. We focus on the horizon dependence and treat d and other problem constants as fixed.

4 Algorithm

We present Conservative-Markdown Redirect-UCB Pricing (CMRUP) in Algorithm 1. The algorithm has three stages. Stage 1 estimates the linear valuation parameter. Stage 2 collects initial demand observations on a residual grid. Stage 3 uses upper confidence bounds on the residual grid, together with a one-step redirect rule, to choose prices adaptively.

Markdown probes. For a queried index $j \in \mathcal{I} = \{0, \dots, M-1\}$, define the probe price

$$p_{t,j}^{+} := \mathrm{clip}_{[0,B]}\big(\hat{u}_t + w_{j+1} - 3\Delta\big). \tag{3}$$

On $\mathcal{E}_\theta$ and under Assumption 3, clipping does not occur in Stages 2–3. Since $|\hat{u}_t - u_t| \le \Delta$, the effective residual of this probe satisfies

$$p_{t,j}^{+} - u_t = w_{j+1} - 3\Delta + (\hat{u}_t - u_t) \in [w_{j-1}, w_j].$$

Thus the conditional purchase probability of probe j obeys

$$S(w_j) \le \mathbb{E}[o_t \mid \mathcal{H}_{t-1}, x_t, \text{probe } j] \le S(w_{j-1}). \tag{4}$$

This deterministic sandwich is the basic observation model used by the algorithm on the residual grid.

Empirical means and confidence radii. Let $\mathcal{T}_j(t)$ contain the earlier rounds on which probe j was *actually played*, including redirects from $j+1$. Write $n_j(t) := |\mathcal{T}_j(t)|$, $\hat{m}_j(t) := n_j(t)^{-1} \sum_{s \in \mathcal{T}_j(t)} o_s$, and $\bar{m}_j(t) := n_j(t)^{-1} \sum_{s \in \mathcal{T}_j(t)} \mathbb{E}[o_s \mid \mathcal{H}_{s-1}, x_s, \text{probe } j]$, setting both means to zero when $n_j(t) = 0$. When probe a_t is played, update only this index by $n_{a_t} \leftarrow n_{a_t} + 1$ and $\hat{m}_{a_t} \leftarrow ((n_{a_t} - 1)\hat{m}_{a_t} + o_t)/n_{a_t}$. Define

$$b_j(t) := C_{\mathrm{ucb}} \sqrt{\frac{\log T}{\max\{1, n_j(t)\}}}. \tag{5}$$

The constants C_θ and C_{ucb} are sufficiently large universal constants ensuring the respective Stage 1 and uniform-confidence events; the analysis does not optimize their smallest valid values. The experiments use $C_{\mathrm{ucb}} = 4$.

Optimistic scores. At a Stage 3 round t, each queried index receives the score

$$U_{t,j} := \big(\hat{u}_t + w_{j+1} + \Delta\big) \min\{1, \hat{m}_j(t) + b_j(t)\}. \tag{6}$$

The first factor is an optimistic price multiplier for residuals in $[w_j, w_{j+1}]$, and the second factor is an upper confidence estimate for $S(w_j)$.

Algorithm 1 Conservative-Markdown Redirect-UCB Pricing (CMRUP)

Input: horizon T; bounds B, B_x, B_θ, c; constants $C_\theta, C_{\mathrm{ucb}}, C_\Sigma$.
$T_1 \leftarrow \lceil T^{2/3} \rceil$, $T_w \leftarrow \lceil T^{2/3} \rceil$.

Stage 1: parameter estimation.
for $t = 1, \ldots, T_1$ **do**
    Post $p_t \sim \mathrm{Unif}[0, B]$ independently; observe o_t.
end for
Set $z_t \leftarrow B o_t$ for $t \le T_1$.
$\widehat{\theta} \leftarrow \left(\sum_{t \le T_1} x_t x_t^\top \right)^\dagger \sum_{t \le T_1} z_t x_t$.
(†: Moore–Penrose pseudoinverse; inverse on the covariance-good event.)
$\widehat{\Sigma}_1 \leftarrow T_1^{-1} \sum_{t=1}^{T_1} x_t x_t^\top$; $r_1 \leftarrow C_\Sigma B_x^2 \sqrt{d \log(T)/T_1}$.
$\underline{\lambda}_1 \leftarrow [\lambda_{\min}(\widehat{\Sigma}_1) - r_1]_+$.
if $\underline{\lambda}_1 = 0$ **then**
    Post uniform prices for the remaining rounds; **stop**.
end if
$\Delta \leftarrow C_\theta B B_x^2 \underline{\lambda}_1^{-1} \sqrt{d \log(T)/T_1}$.
Construct the grid $w_i = -c + 2i\Delta$, $i = 0, \ldots, M$, where $M = \lceil c/\Delta \rceil$.
Initialize $n_j \leftarrow 0$ and $\hat{m}_j \leftarrow 0$ for all $j \in \mathcal{I} = \{0, \ldots, M-1\}$.

Stage 2: warmup probes.
for $t = T_1 + 1, \ldots, T_1 + T_w$ **do**
    Observe x_t and set $\hat{u}_t \leftarrow \langle x_t, \widehat{\theta} \rangle$.
    Sample $J_t \sim \mathrm{Unif}(\mathcal{I})$.
    Post $p_t \leftarrow \mathrm{clip}_{[0,B]}(\hat{u}_t + w_{J_t+1} - 3\Delta)$; observe o_t.
    Update n_{J_t} and $\hat{m}_{J_t}$ using o_t.
end for

Stage 3: adaptive pricing.
for $t = T_1 + T_w + 1, \ldots, T$ **do**
    Observe x_t and set $\hat{u}_t \leftarrow \langle x_t, \widehat{\theta} \rangle$.
    for $j \in \mathcal{I}$ **do**
        $b_j \leftarrow C_{\mathrm{ucb}} \sqrt{\log T / \max\{1, n_j\}}$.
        $U_{t,j} \leftarrow (\hat{u}_t + w_{j+1} + \Delta) \min\{1, \hat{m}_j + b_j\}$.
    end for
    $j_t \in \arg\max_{j \in \mathcal{I}} U_{t,j}$.
    if $b_{j_t} > \Delta$ **then**
        $a_t \leftarrow j_t$. ▷ direct probe
    else if $j_t = 0$ **then**
        $a_t \leftarrow 0$. ▷ boundary case
    else
        $a_t \leftarrow j_t - 1$. ▷ one-step redirect
    end if
    Post $p_t \leftarrow \mathrm{clip}_{[0,B]}(\hat{u}_t + w_{a_t+1} - 3\Delta)$; observe o_t.
    Update n_{a_t} and $\hat{m}_{a_t}$ using o_t.
end for

4.1 Mechanics of the Algorithm

Residual-grid observations. The 3Δ markdown combines one grid step of length 2Δ with the prediction-error allowance Δ. As a result, a probe indexed by j generates an observation whose conditional mean lies between $S(w_j)$ and $S(w_{j-1})$. The possible mismatch between these two values is exactly the adjacent jump height α_j.

Upper confidence search. The score $U_{t,j}$ is an upper confidence estimate of the revenue associated with residual interval $[w_j, w_{j+1}]$. Lemma 5 shows that, on the joint good event, the maximum score

upper bounds the oracle revenue in every Stage 3 round. The score uses only observations collected from probes that were actually played.

One-step redirect. Directly probing index j may incur a discontinuity cost proportional to α_j. The algorithm directly probes j only while $b_j > \Delta$, which limits the number of direct probes at each index. Once $b_j \le \Delta$, the selected index is executed by probing $j-1$ instead. This redirected probe has at least the demand level needed to validate the selected score, and the incurred price loss is only of order Δ.

Role of the warmup stage. The warmup stage initializes the residual-grid observations and gives every grid point a chance to be sampled before the adaptive stage. It is not the source of the $T^{2/3}$ rate: the Stage 3 confidence analysis depends only on counts of actually played probes and would also work with infinite initial radii and forced first visits. We use $T_w = \lceil T^{2/3} \rceil$ because this cost is of the same order as the target regret and keeps the algorithmic description simple. A shorter initialization or a Stage 3-only initialization rule can be analyzed with the same counting argument.

5 Regret Analysis

We state the regret guarantee for Algorithm 1 and summarize the proof. Full proofs are in Appendix C.

Theorem 1 (Cumulative regret)**.** *Under Assumptions 1–3, Algorithm 1 satisfies*

$$R_T \le \widetilde{O}(T^{2/3}).$$

The guarantee allows arbitrary bounded-support noise distributions, including distributions whose CDF or survival function has atoms, jumps, and no Lipschitz or smoothness regularity.

The constants hidden in $\widetilde{O}(\cdot)$ may depend polynomially on $(d, B_x, B_\theta, B, c, 1/\lambda_0, 1/\kappa)$; details appear in Appendix C.9.

Proof architecture. The proof has four ingredients. First, uniformly randomized prices in Stage 1 produce $\widehat{\theta}$ with $\max_t |\hat{u}_t - u_t| \le \Delta = \widetilde{O}(T^{-1/3})$ with high probability. Second, conservative markdown converts this prediction guarantee into the deterministic sandwich $S(w_j) \le m_{t,j} \le S(w_{j-1})$ whenever probe index j is played. Third, empirical confidence bounds on adaptively collected probe means imply optimism: the largest score $\max_j U_{t,j}$ upper bounds the oracle revenue in each Stage 3 round. Finally, direct probes and redirected probes are charged separately. Direct probes may pay the adjacent jump cost α_j, but each index is directly probed only until its confidence radius reaches the grid scale; redirected probes avoid this jump cost and lose only $O(\Delta)$ in price.

Key lemmas. The first lemma gives the prediction guarantee for the linear component.

Lemma 2 (Parameter estimation)**.** *Under Assumptions 1–2, if $T_1 = \lceil T^{2/3} \rceil$ and Δ is chosen as in Algorithm 1, then with probability at least $1 - T^{-2}$, we have $\max_{1 \le t \le T} |\hat{u}_t - u_t| \le \Delta$.*

Proof sketch. For $p_t \sim \mathrm{Unif}[0, B]$ and $y_t \in [0, B]$, the signal $z_t := Bo_t$ satisfies $\mathbb{E}[z_t \mid x_t] = \mathbb{E}[y_t \mid x_t] = \langle x_t, \theta^\star \rangle$, using exogeneity and $\mathbb{E}[\xi_t] = 0$. Standard bounded-design least-squares concentration, together with the empirical covariance lower bound implied by Assumption 2, gives $\|\widehat{\theta} - \theta^\star\|_2 \lesssim (BB_x/\lambda_0)\sqrt{d \log(T)/T_1}$ with high probability. Multiplying by B_x and using the definition of Δ yields the claim. See Appendix C.3 for a detailed proof. □

Lemma 3 (Markdown sandwich)**.** *On $\mathcal{E}_\theta$, for every Stage 2–3 round in which probe index j is played, its conditional purchase probability $m_{t,j} := \mathbb{E}[o_t \mid \mathcal{H}_{t-1}, x_t, \text{probe } j]$ satisfies*

$$S(w_j) \le m_{t,j} \le S(w_{j-1}).$$

Proof sketch. The un-clipped probe price is $\hat{u}_t + w_{j+1} - 3\Delta$. Assumption 3 prevents clipping, and $|\hat{u}_t - u_t| \le \Delta$ places the realized residual in $[w_{j-1}, w_j]$. Monotonicity of S gives the sandwich. See Appendix C.4 for more details. □

Lemma 4 (Uniform confidence for adaptive probes)**.** *There exists a sufficiently large constant C_{ucb} such that, with probability at least $1 - T^{-2}$, simultaneously for all Stage 3 rounds t and all $j \in \mathcal{I}$,*

$$|\hat{m}_j(t) - \bar{m}_j(t)| \le \frac{1}{2} b_j(t), \qquad b_j(t) := C_{\mathrm{ucb}} \sqrt{\frac{\log T}{\max\{1, n_j(t)\}}}.$$

On the same event, $S(w_j) \le \hat{m}_j(t) + b_j(t) \le S(w_{j-1}) + 2b_j(t)$.

Proof sketch. For each queried index, enumerate the stopping times at which that index is actually probed. The centered observations at these times form a bounded martingale-difference sequence under the stopped filtration, even though the probing times are adaptive and may include redirected plays. A Hoeffding–Azuma bound for each index and sample size, followed by a union bound over j and $m \le T$, yields the uniform event. Combining it with Lemma 3 gives the two one-sided inequalities. See Appendix C.5 for a complete proof. □

Lemma 5 (Optimism of the score)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$, every Stage 3 round t satisfies*

$$V_t^\star \le \max_{j \in \mathcal{I}} U_{t,j}.$$

Proof sketch. Let $w_t^\star$ be an optimal residual and choose q such that $w_t^\star \in [w_q, w_{q+1}]$. Such a $q \le M - 1$ exists because $w_0 = -c$ and $w_M \ge c$; if $w_t^\star = c$, we take $q = M - 1$. On $\mathcal{E}_\theta$, the price multiplier in $U_{t,q}$ dominates $u_t + w_t^\star$, while monotonicity gives $S(w_q) \ge S(w_t^\star)$. Lemma 4 gives $\hat{m}_q + b_q \ge S(w_q)$, so the score of q upper bounds $V_t^\star$. By Assumption 3 and $2\Delta \le \kappa$, all candidate price levels appearing in the optimistic scores are nonnegative. See Appendix C.6. □

Lemma 6 (One-round regret)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$, let j_t be the score maximizer in Stage 3.*

(i) If Algorithm 1 is in direct-probe mode and plays j_t, then

$$V_t^\star - \pi(x_t, p_t) \le C\Delta + B\alpha_{j_t} + CBb_{j_t}(t).$$

(ii) If Algorithm 1 is in redirect mode, then

$$V_t^\star - \pi(x_t, p_t) \le C_B \Delta,$$

where C_B depends only on B.

Proof sketch. By optimism, the oracle value is at most the selected score. In direct-probe mode, the selected score can exceed the revenue of the played probe only through the grid-scale price slack, the adjacent jump α_{j_t}, and the confidence radius b_{j_t}. The truncation $\min\{1, \hat{m}_j + b_j\}$ can only reduce the score. In redirect mode, the algorithm probes $j_t - 1$ when $j_t \ge 1$; this action has conditional demand at least $S(w_{j_t - 1})$, which covers the selected score's demand component up to the confidence radius. Since redirect mode occurs only when $b_{j_t} \le \Delta$, the remaining gap is $O(\Delta)$. The boundary case $j_t = 0$ uses $S(w_{-1}) = S(w_0) = 1$. See Appendix C.7 for more details. □

Lemma 7 (Cumulative Stage 3 regret on the good event)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$, the cumulative regret in Stage 3 is $\widetilde{O}(T^{2/3})$.*

Proof sketch. Redirect rounds contribute at most $O(\Delta)$ each by Lemma 6, hence $O(T\Delta) = \widetilde{O}(T^{2/3})$. Direct-probe rounds contribute Δ, $b_{j_t}(t)$, and α_{j_t}. The first term is again $O(T\Delta)$. The confidence terms satisfy the standard counting bound $\sum_{\mathrm{probe}} b_{j_t}(t) \le \widetilde{O}(\sqrt{MT}) = \widetilde{O}(T^{2/3})$ since $M = O(\Delta^{-1})$. For the jump terms, index j is directly probed only while $b_j > \Delta$, so it is directly probed at most $O(\log T / \Delta^2)$ times; therefore $\sum_{\mathrm{probe}} \alpha_{j_t} \le O(\log T/\Delta^2) \sum_j \alpha_j = \widetilde{O}(T^{2/3})$. See Appendix C.8. □

Proof sketch of Theorem 1. Stages 1 and 2 each last $O(T^{2/3})$ rounds and have per-round regret at most B. Lemma 7 controls Stage 3 on the joint good event, whose complement has probability $O(T^{-2})$ and contributes at most $BT \cdot O(T^{-2}) = O(1/T)$. Thus $R_T \le \widetilde{O}(T^{2/3})$. Full details are in Appendix C.9.

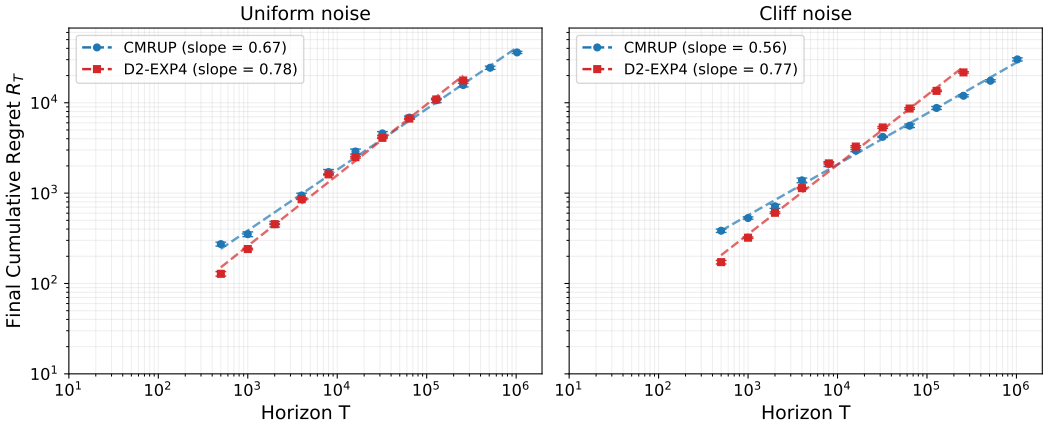

Figure 1: Final cumulative pseudo-regret R_T on synthetic contextual-pricing instances with smooth uniform noise and discontinuous cliff noise. CMRUP is averaged over 10 independent seeds, while D2-EXP4 is averaged over 5 seeds due to its substantially higher computational cost. Error bars show standard errors across seeds. Dashed lines are all-point least-squares fits on the log-log scale, using all available horizons for each algorithm.

6 Numerical Experiments

We report synthetic experiments designed to verify the final-regret scaling predicted by Theorem 1. The experiments compare CMRUP with the agnostic D2-EXP4 baseline of Xu and Wang [2022] on both a smooth Lipschitz noise distribution and a discontinuous non-Lipschitz distribution containing an atom. The primary quantity of interest is the final cumulative pseudo-regret R_T as a function of the horizon T.

Setup. We generate i.i.d. contexts $x_t \in \mathbb{R}^5$ with an intercept coordinate $x_{t,0} = 1$ and independent features $x_{t,j} \sim \mathrm{Unif}(0,1)$ for $j = 1, \ldots, 4$. The parameter is $\theta^\star = (2, 0.125, 0.125, 0.125, 0.125)^\top$, so that $u_t = \langle x_t, \theta^\star \rangle \in (2, 2.5)$. We set the noise support to $[-1, 1]$ and consider two noise distributions: uniform noise, $\xi_t \sim \mathrm{Unif}[-1, 1]$, whose survival function is Lipschitz; and cliff noise, where $\xi_t = 0$ with probability 0.3 and otherwise $\xi_t \sim \mathrm{Unif}[-1, 1]$, creating a jump discontinuity of height 0.3 in the survival function at $w = 0$.

For CMRUP, we run horizons $T \in \{500, 1000, 2000, \ldots, 1{,}024{,}000\}$ with 10 independent seeds per horizon. In the experiments, the default residual-grid multiplier is $s = 0.35$ and $C_{\mathrm{ucb}} = 4$; Appendix C.11 specifies the implementation normalization of Δ. The conservative markdown coefficient remains fixed at 3, so Stage-2/3 probes use a 3Δ markdown. For D2-EXP4, we use the same synthetic streams and pseudo-regret metric, running horizons $T \in \{500, 1000, 2000, \ldots, 256{,}000\}$ with 5 independent seeds per horizon. Since exact enumeration of the D2-EXP4 policy class is computationally prohibitive at these horizons, we use the sampled-policy implementation described in Appendix C.11. All experiments use CPU-only local computations.

Final-regret scaling. Figure 1 plots R_T against T on log-log axes. For each algorithm and noise distribution, we fit a power law $R_T \approx CT^\alpha$ by ordinary least squares in log-log space using all available horizons for that algorithm.

On the **uniform noise** instance, CMRUP has fitted exponent 0.67, while D2-EXP4 has fitted exponent 0.78. Although the two algorithms have comparable regret at small horizons, D2-EXP4 exhibits a visibly steeper growth rate, whereas CMRUP tracks the predicted $T^{2/3}$ scaling. On the **cliff noise** instance, the separation is more pronounced: CMRUP has fitted exponent 0.56, while D2-EXP4 has fitted exponent 0.77. This is consistent with the role of the conservative markdown and redirect rule, which prevent the atom-induced jump from causing a worse asymptotic scaling. The fitted exponents are finite-horizon slopes: the smaller cliff slope may reflect a larger fraction of low-regret redirected rounds in this particular instance, rather than a stronger asymptotic guarantee.

Table 2: Paired mechanism diagnostics (10 matched seeds, 12 horizons). All non-exponent entries use $T = 1{,}024{,}000$ ($\pm$: paired standard error). Grid scales $s \in \{0.20, 0.35, 0.50, 0.70\}$ were tested without retuning; 0.50 had the lowest mean final regret. The no-markdown cliff result is inconclusive.

Measure	Uniform noise	Cliff noise
CMRUP Stage 3 fitted exponent	0.678	0.512
Stage 3 share of total regret	72.2%	54.9%
No redirect: excess final regret	$4{,}139 \pm 1{,}527$	$5{,}763 \pm 569$
No redirect: final-regret ratio	1.115	1.191
No redirect: direct-jump proxy ratio	≈ 2.4	≈ 3.2
No markdown ($a = 0$): excess regret	$7{,}059 \pm 1{,}887$	$1{,}851 \pm 3{,}902$

Scope. These simulations are rate-verification experiments rather than a comprehensive empirical benchmark. The D2-EXP4 baseline is a sampled-policy implementation of the prior agnostic reduction, included to illustrate the empirical scaling difference under the same synthetic environment; it is not an exhaustive implementation of the exponentially large policy class. The paired mechanism diagnostics, summarized in Table 2, isolate the redirect and markdown components and report Stage 3-only slopes. All variants were pre-specified and used matched seeds without retuning. These finite-horizon findings are mechanism diagnostics, not asymptotic rate claims.

7 Discussion and Conclusion

We close with limitations, ethical considerations, and a summary of the main result. Additional related discussions appear in Appendix B.

Limitations and extensions. Compared with the $\widetilde{O}(T^{3/4})$ guarantee of Xu and Wang [2022], which permits adversarial contexts, our $\widetilde{O}(T^{2/3})$ rate improves the horizon exponent on the stochastic-context subclass. The bound is optimal in T for fixed problem parameters, not in d: the displayed analysis yields an envelope $\widetilde{O}((1 + \sqrt{d}/\lambda_0)T^{2/3})$ when other parameters are fixed, versus $\widetilde{O}(T^{3/4} + \sqrt{d}\,T^{5/8})$ in the earlier adversarial-context analysis, so neither bound uniformly dominates in growing dimension. The fixed buffer $\kappa > 0$ keeps markdown probes away from the price boundaries; $2\Delta \le \kappa$ follows at sufficiently large horizons, and Algorithm 1 estimates the covariance eigenvalue lower bound rather than requiring its numerical value beforehand. Appendix B discusses bounded-support and clipping extensions, horizon-free operation, and richer context processes.

Computation. The direct implementation is polynomial time: Stage 1 solves a d-dimensional least-squares problem, and Stage 3 scans $M = \widetilde{O}(T^{1/3})$ residual-grid points per round. This gives total time $\widetilde{O}(T^{4/3} + T^{2/3}d^2 + d^3)$ and memory $\widetilde{O}(T^{1/3} + d^2)$. Faster score maximization may improve constants in large-scale deployments, but is orthogonal to the regret analysis.

Ethical considerations. This paper studies a theoretical online pricing model. The algorithm prices one item or service instance at a time and does not simultaneously offer the identical good to different buyers at different prices. Nevertheless, contextual pricing can create disparate-impact concerns if contexts contain protected attributes or proxies. Responsible deployment should audit features, monitor realized prices and acceptance rates across groups, and enforce policy or fairness constraints when appropriate. The proposed method is relatively transparent—a linear predictor, a residual grid, and explicit confidence bounds—which can facilitate such audits.

Conclusion. We presented a polynomial-time algorithm for contextual dynamic pricing with linear valuations and fully agnostic non-Lipschitz noise. The algorithm combines randomized estimation of the linear component, conservative markdown probes on a residual grid, and an optimistic redirect rule that controls discontinuity-induced bias without assuming smoothness of the demand curve. It achieves $\widetilde{O}(T^{2/3})$ expected regret, matching known lower bounds up to logarithmic factors. The result shows that atoms and arbitrary jumps in the noise distribution do not preclude optimal regret in contextual pricing, provided the algorithm uses a pricing rule that respects the one-sided structure of censored feedback.

Acknowledgments and Disclosure of Funding

This work received no specific external funding.

References

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Gah-Yi Ban and N Bora Keskin. Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science*, 67(9):5549–5568, 2021.

Hamsa Bastani, David Simchi-Levi, and Ruihao Zhu. Meta dynamic pricing: Transfer learning across experiments. *Management Science*, 68(3):1865–1881, 2022.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

Omar Besbes and Assaf Zeevi. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739, 2015.

Daniele Bracale, Moulinath Banerjee, Yuekai Sun, Salam Turki, and Kevin Stoll. Dynamic pricing in the linear valuation model using shape constraints. *Transactions on Machine Learning Research*, 2025.

Jinzhi Bu, David Simchi-Levi, and Yunzong Xu. Online pricing with offline data: Phase transition and inverse square law. In *international conference on machine learning*, pages 1202–1210. PMLR, 2020.

Nicolo Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, 61(1):549–564, 2014.

Nicolo Cesa-Bianchi, Tommaso Cesari, and Vianney Perchet. Dynamic pricing with finitely many unknown valuations. In *Algorithmic Learning Theory*, pages 247–273. PMLR, 2019.

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.

Xi Chen, Xuan Zhang, and Yuan Zhou. Fairness-aware online price discrimination with nonparametric demand models. *arXiv preprint arXiv:2111.08221*, 2021.

Xi Chen, David Simchi-Levi, and Yining Wang. Privacy-preserving dynamic personalized pricing with demand learning. *Management Science*, 68(7):4878–4898, 2022.

Wang Chi Cheung, David Simchi-Levi, and He Wang. Dynamic pricing and demand learning with limited price experimentation. *Operations Research*, 65(6):1722–1731, 2017.

Maxime C Cohen, Ruben Lobel, and Georgia Perakis. Dynamic pricing through data sampling. *Production and Operations Management*, 27(6):1074–1088, 2018.

Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *Management Science*, 66(11):4921–4943, 2020.

Maxime C Cohen, Sentao Miao, and Yining Wang. Dynamic pricing with fairness constraints. *Available at SSRN 3930622*, 2021.

Maxime C Cohen, Adam N Elmachtoub, and Xiao Lei. Price discrimination with fairness constraints. *Management Science*, 2022.

Jianqing Fan, Yongyi Guo, and Mengxin Yu. Policy optimization using semiparametric models for dynamic pricing. *Journal of the American Statistical Association*, 119(545):552–564, 2024.

Negin Golrezaei, Adel Javanmard, and Vahab Mirrokni. Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Advances in Neural Information Processing Systems*, 32, 2019.

Xueping Gong, Wei You, and Jiheng Zhang. Minimax optimality in contextual dynamic pricing with general valuation models. *Operations Research*, 74(2):879–897, 2026.

Adel Javanmard. Perishability of data: dynamic pricing under varying-coefficient models. *Journal of Machine Learning Research*, 18(53):1–31, 2017.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *The Journal of Machine Learning Research*, 20(1):315–363, 2019.

Adel Javanmard, Hamid Nazerzadeh, and Simeng Shao. Multi-product dynamic pricing in high-dimensions with heterogeneous price sensitivity. In *International Symposium on Information Theory (ISIT-20)*, pages 2652–2657. IEEE, 2020.

N Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

N Bora Keskin and Assaf Zeevi. Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research*, 42(2):277–307, 2017.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *IEEE Symposium on Foundations of Computer Science (FOCS-03)*, pages 594–605. IEEE, 2003.

Akshay Krishnamurthy, Thodoris Lykouris, Chara Podimata, and Robert Schapire. Contextual search in the presence of irrational agents. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC-21)*, pages 910–918, 2021.

Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS-18)*, pages 268–282. IEEE, 2018.

Renato Paes Leme, Chara Podimata, and Jon Schneider. Corruption-robust contextual search through density updates. In *Conference on Learning Theory*, pages 3504–3505. PMLR, 2022.

Allen Liu, Renato Paes Leme, and Jon Schneider. Optimal contextual pricing and extensions. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA-21)*, pages 1059–1078. SIAM, 2021.

Ilan Lobel, Renato Paes Leme, and Adrian Vladu. Multidimensional binary search for contextual decision-making. *Operations Research*, 66(5):1346–1361, 2018.

Yiyun Luo, Will Wei Sun, et al. Distribution-free contextual dynamic pricing. *arXiv preprint arXiv:2109.07340*, 2021.

Yiyun Luo, Will Wei Sun, and Yufeng Liu. Contextual dynamic pricing with unknown noise: Explore-then-ucb strategy and improved regrets. In *Advances in Neural Information Processing Systems*, 2022.

Jieming Mao, Renato Paes Leme, and Jon Schneider. Contextual pricing for lipschitz buyers. In *NeurIPS*, pages 5648–5656, 2018.

Mila Nambiar, David Simchi-Levi, and He Wang. Dynamic learning and pricing with model misspecification. *Management Science*, 65(11):4980–5000, 2019.

Virag Shah, Ramesh Johari, and Jose Blanchet. Semi-parametric dynamic contextual pricing. *Advances in Neural Information Processing Systems*, 32, 2019.

Matilde Tullii, Solenne Gaucher, Nadav Merlis, and Vianney Perchet. Improved algorithms for contextual dynamic pricing. *arXiv preprint arXiv:2406.11316*, 2024.

Yining Wang, Xi Chen, Xiangyu Chang, and Dongdong Ge. Uncertainty quantification for demand prediction in contextual dynamic pricing. *Production and Operations Management*, 30(6):1703–1717, 2021.

Jianyu Xu and Yu-Xiang Wang. Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Zongsen Yang, Xingyu Fu, Pin Gao, and Ying-Ju Chen. Fairness regulation of prices in competitive markets. *Available at SSRN*, 2022.

A More Related Works

This appendix complements Section 2 with additional background on dynamic pricing, bandit methods, and fairness-aware pricing.

Semiparametric and distribution-free contextual pricing. Several papers study contextual pricing with unknown residual distributions under different structural assumptions. Shah et al. [2019] consider semiparametric dynamic contextual pricing and jointly learn regression parameters and residual distributions. Luo et al. [2021] study contextual dynamic pricing with known $\theta^\star$ and unknown market noise, while Liu et al. [2021] and Mao et al. [2018] analyze contextual pricing/search models under different assumptions on buyers and feedback. These works are conceptually related because they reduce aspects of pricing to learning residual or threshold structure from censored feedback. Our setting focuses specifically on the linear-valuation model with bounded but otherwise arbitrary noise, and the main challenge is to obtain the minimax $T^{2/3}$ rate without continuity of the residual demand curve.

Classical non-contextual dynamic pricing. The non-contextual version of the problem has a single unknown demand curve and no covariates. Kleinberg and Leighton [2003] established the classical $\Omega(T^{2/3})$ lower bound for online posted-price auctions. Besbes and Zeevi [2009] study dynamic pricing without knowing the demand function and develop near-optimal policies in nonparametric settings. This literature motivates the $T^{2/3}$ benchmark that also appears in contextual pricing. Contextual pricing is harder because the oracle price shifts with x_t and the learner observes only binary feedback, so the linear component and the residual demand curve must be learned simultaneously.

Parametric and high-dimensional pricing. A separate line of work studies dynamic pricing under parametric demand models or high-dimensional contextual structure. Examples include parametric personalized pricing [Ban and Keskin, 2021], high-dimensional dynamic pricing [Javanmard et al., 2020], and related feature-based pricing models [Cohen et al., 2020]. These results obtain strong guarantees by imposing parametric structure, sparsity, or known regularity conditions. Our model is nonparametric in the residual noise distribution and allows discontinuities in demand.

Online pricing with demand learning: A broader view. A substantial body of work studies dynamic pricing when the demand function must be learned online under various structural assumptions. Besbes and Zeevi [2015] show that even misspecified linear demand models can yield near-optimal pricing policies, highlighting the surprising sufficiency of simple parametric models. Keskin and Zeevi [2014] design semi-myopic policies for single-product pricing under unknown parametric demand and prove asymptotic optimality up to logarithmic factors, while Keskin and Zeevi [2017] extend the analysis to non-stationary environments and characterize the additional regret cost of tracking a drifting demand curve. On the algorithmic side, Cheung et al. [2017] study pricing under a strict cap on the number of price changes and obtain matching iterated-logarithm regret bounds. Nambiar et al. [2019] address model misspecification directly and propose a random price shock policy with provable robustness guarantees. Bu et al. [2020] characterize the value of offline data for online pricing through a phase-transition phenomenon and an inverse-square law governing how sample size, location, and dispersion affect the optimal regret. Bastani et al. [2022] take a multi-task perspective and show that transfer learning across related pricing experiments via a shared Thompson-sampling prior can substantially accelerate learning. Cohen et al. [2018] develop data-driven pricing methods using sample-average and max-min ratio formulations under limited demand information. The joint pricing-and-inventory variant is studied by Javanmard [2017], who introduces a projected stochastic gradient pricing policy for high-dimensional contextual demand with time-varying parameters. These works complement the present paper by studying the learning-and-earning trade-off from different angles—misspecification robustness, limited experimentation, offline warm-starting, and meta-learning—whereas our focus is on removing regularity assumptions on the noise distribution altogether.

Posted-price mechanisms and bandit-based pricing. Contextual search is closely related to dynamic pricing: both involve learning an unknown threshold from binary feedback. Lobel et al. [2018] introduce the Projected Volume algorithm for d-dimensional contextual binary search and

obtain $O(d\log(d/\varepsilon))$ regret, with direct applications to feature-based pricing. Leme and Schneider [2018] develop intrinsic-volume potential arguments that yield $O_d(1)$ total loss for symmetric loss and doubly logarithmic regret for the pricing variant of contextual search. Robustness to model misspecification in contextual search is studied by Krishnamurthy et al. [2021], who design corruption-robust algorithms whose regret degrades gracefully with the number of irrational agents, and by Leme et al. [2022], who prove tight regret bounds via density-update methods. On the privacy front, Chen et al. [2022] propose differentially private contextual pricing policies with near-optimal regret. Wang et al. [2021] develop confidence-interval-based uncertainty quantification for demand parameters in contextual pricing. In incentive-aware settings, Golrezaei et al. [2019] design contextual robust pricing policies that remain effective when buyers behave strategically. For non-contextual posted-price auctions, Cesa-Bianchi et al. [2019] study revenue maximization when valuations are supported on finitely many unknown points and establish tight $\Theta(\sqrt{KT})$ regret, while Cesa-Bianchi et al. [2014] give efficient $\widetilde{O}(\sqrt{T})$-regret algorithms for reserve-price optimization in second-price auctions. The closely related bilateral-trade setting is studied by Cesa-Bianchi et al. [2024], who provide a complete characterization of regret rates across different feedback and valuation models. These works collectively show that the interplay between censored feedback, context dimensionality, strategic behavior, and distributional assumptions governs the statistical complexity of online pricing; our contribution focuses specifically on removing distributional regularity in the residual demand curve while retaining the optimal $\widetilde{O}(T^{2/3})$ rate.

Contextual search and threshold feedback. Contextual search and pricing with threshold feedback are closely related to dynamic pricing because the learner observes binary information about an unknown threshold. Mao et al. [2018] study contextual pricing for Lipschitz buyers, Liu et al. [2021] analyze optimal contextual pricing and extensions, and Leme et al. [2022] study model misspecification in contextual search and prove a tight bound. These works differ in model details and regularity assumptions as their non-continuity assumptions are mainly made on the observations (the "demand") instead of distributions (the "demand curve"). That said, they share the challenge of learning from one-bit threshold observations.

Feature-based and policy-class approaches. Feature-based pricing algorithms often reduce the problem to learning over a class of policies. This approach is flexible because it can handle weak assumptions on the noise distribution, but it may require a large policy class. Xu and Wang [2022] use this perspective for agnostic linear-valuation pricing and obtain $\widetilde{O}(T^{3/4})$ regret via a discretized policy class and an adversarial bandit algorithm. Our algorithm instead exploits the additive linear-valuation structure more directly: once the linear component is estimated, the remaining uncertainty is one-dimensional in the residual coordinate.

Semiparametric residual learning. Shah et al. [2019] study semiparametric dynamic contextual pricing, where both regression parameters and the residual distribution must be learned from censored feedback. Fan et al. [2024] develop a uniformly randomized pricing estimator that yields a semiparametric route to learning the linear component; we use the same identity in Stage 1. The novelty of our analysis lies in how the residual demand curve is handled after parameter learning: conservative markdown produces one-sided grid observations, and redirect-UCB controls discontinuities without imposing smoothness.

Shape constraints and discontinuous demand. The survival function $S(w) = \mathbb{P}(\xi \geq w)$ is monotone even when it is discontinuous. Shape-constrained ideas are therefore natural for agnostic pricing. Bracale et al. [2025] use isotonic regression in the linear-valuation model under Hölder-type regularity assumptions, while Gong et al. [2026] study contextual dynamic pricing with general valuation models but require a Lipschitz noise CDF and bounded valuation functions. Our analysis uses only monotonicity and bounded support. The markdown sandwich ensures that binary observations can be interpreted as one-sided probes of adjacent survival levels, and the confidence-bound analysis does not require continuity or differentiability of S.

Confidence-bound methods. Upper-confidence-bound algorithms are a standard tool for balancing exploration and exploitation in multi-armed bandits [Auer et al., 2002, Auer, 2002]. Our Stage 3 analysis uses the same counting principle: if an arm with confidence radius b_j is selected repeatedly, its count increases and the radius shrinks, and the cumulative confidence cost is bounded by $\widetilde{O}(\sqrt{MT})$

over M grid indices. The pricing-specific difficulty is that probing a residual index near a discontinuity may introduce a one-sided jump bias. The redirect rule separates this jump cost from the confidence cost and bounds it using the monotonicity identity $\sum_j \alpha_j \leq 1$.

Fairness-aware pricing. Personalized and contextual pricing can raise fairness concerns, especially when contexts contain sensitive information or proxies. A growing literature studies dynamic pricing under fairness constraints or fairness regulation [Cohen et al., 2021, 2022, Chen et al., 2021, Yang et al., 2022]. Our regret analysis does not impose fairness constraints, but the algorithmic structure is compatible with them: one can restrict the admissible price set, constrain the context features, or post-process the selected price to satisfy externally specified fairness rules. Appendix B discusses ethical considerations in more detail.

B More Discussions

This appendix expands on the limitations and implementation considerations discussed in Section 7.

Overall: Limitations and extensions. The analysis assumes i.i.d. contexts with a well-conditioned covariance matrix. This condition is used to learn the linear parameter in Stage 1. Extending the result to richer context processes is an important direction. For example, with obliviously chosen contexts satisfying an empirical full-rank condition, one may still expect the randomized-price estimator to recover $\theta^\star$; handling fully adaptive or adversarial contexts would require a different parameter-learning component.

We also assume bounded noise and a price buffer ensuring that the markdown probes remain in the admissible price interval. Bounded support is standard in worst-case regret analyses for posted-price learning and can often be approximated by truncation under sub-Gaussian or sub-exponential tails. The buffer condition is a boundary condition for the price interval rather than a smoothness condition on demand. Removing it would require explicitly treating clipped markdown probes; this appears technically feasible but would add boundary cases to the sandwich argument.

Computational complexity. The algorithm is polynomial time. Stage 1 requires forming and inverting a $d \times d$ covariance matrix, with cost $O(T_1 d^2 + d^3)$ using a direct implementation. Stage 2 performs $T_w = \widetilde{O}(T^{2/3})$ constant-time grid updates. Stage 3 scans $M = \widetilde{O}(T^{1/3})$ grid indices per round to compute the optimistic score, giving total time $\widetilde{O}(TM) = \widetilde{O}(T^{4/3})$, plus $O(1)$ empirical-mean updates per round. The overall direct implementation cost is therefore

$$\widetilde{O}(T^{4/3} + T^{2/3} d^2 + d^3),$$

with memory $\widetilde{O}(T^{1/3} + d^2)$. This is polynomial in both T and d and avoids enumerating a high-dimensional policy class. For very large-scale deployments, faster score maximization would be useful; this computational optimization is orthogonal to the regret analysis.

Bounded support and truncation. The bounded-support assumption on ξ_t allows the algorithm to work on a finite residual grid. If the noise has light tails, a truncated version of the algorithm can be applied on an interval whose length grows logarithmically with T. The regret contribution from tail events can then be controlled by the tail probability, while the grid size and markdown scale acquire only logarithmic factors. A complete treatment would require replacing the bounded valuation assumption in Stage 1 by a corresponding concentration condition for the randomized-price regression signal.

Price-buffer and clipping. The price-buffer assumption ensures that the conservative markdown probes are not clipped in Stages 2–3 on the Stage 1 good event. This keeps the sandwich proof simple: the effective residual of probe j lies exactly in $[w_{j-1}, w_j]$. The condition is not a regularity assumption on F or S; it only keeps the probed prices away from the boundary of the admissible price interval. For the asymptotic theorem, it is enough that the buffer dominates the grid scale, i.e., $\kappa \gtrsim \Delta$. A fixed positive buffer is a clean sufficient condition.

Boundary clipping can be handled by explicit conventions. If a lower-clipped price is zero and valuations are nonnegative, purchase occurs with probability one, which is consistent with the lower

boundary survival value. Upper clipping is ruled out by the bounded valuation model. One can alternatively discard lower-clipped observations from the residual-grid dataset and charge the clipped rounds directly as boundary rounds. We do not include this extension in the main theorem in order to keep the presentation focused on the core discontinuity issue rather than boundary bookkeeping.

Horizon dependence. The grid scale Δ and the lengths of Stages 1–2 depend on the horizon T. This is common in regret-optimal pricing algorithms. A horizon-free version can be obtained by a doubling trick: run the algorithm in episodes of lengths $1, 2, 4, \ldots$ and restart the grid at the beginning of each episode. If the regret in an episode of length 2^r is at most $C(2^r)^{2/3}\,\mathrm{polylog}(2^r)$, then over all completed episodes up to horizon T,

$$\sum_{r \le \log_2 T} C(2^r)^{2/3}\,\mathrm{polylog}(2^r) \le C' T^{2/3}\,\mathrm{polylog}(T).$$

Thus the doubling trick preserves the $\widetilde{O}(T^{2/3})$ rate.

Context assumptions. The i.i.d. context assumption is used to obtain a high-probability least-squares estimate of $\theta^\star$ from uniformly randomized prices. If the learner is given an exogenous exploration design with a well-conditioned empirical covariance matrix, the same proof can be adapted. Removing both stochasticity and empirical full-rank conditions is more difficult, because the binary feedback may not contain enough information to identify the linear component in directions that are rarely explored by the contexts.

Choice of the price range in Stage 1. The randomized-price regression identity in Stage 1 uses prices uniformly distributed on an interval containing the valuation support. The variance of the regression signal $z_t = Bo_t$ scales with B^2, so a tighter valid price range can improve finite-sample performance. The theory uses a known upper bound B for simplicity. In practice, one may use the smallest reliable admissible range, or combine the method with a preliminary range-estimation step when conservative bounds are available.

Grid size and dependence on the support radius. The confidence cost in Stage 3 scales as $\widetilde{O}(\sqrt{MT})$, where $M = O(c/\Delta)$. Thus the support radius c enters the problem-dependent constants. This dependence is natural for a uniform residual grid over the entire bounded support. Adaptive grid refinement or focusing on promising residual regions may improve constants in benign instances, but such refinements would need additional bookkeeping to preserve optimism under discontinuities.

Ethical considerations. The algorithm prices one item or service instance at a time. It does not simultaneously offer the same identical good to different buyers at different prices, and therefore does not instantiate simultaneous same-good price discrimination. The purpose of the context is to model heterogeneity across sequential pricing instances, such as differentiated products, changing service conditions, or buyer-product pairs.

At the same time, contextual pricing systems should be deployed with explicit governance. If the context vector includes protected attributes or strong proxies, any revenue-maximizing algorithm may produce undesirable disparities over time. Practical deployments should therefore audit the feature set, monitor realized prices and acceptance rates across relevant groups, and enforce policy constraints when required. The transparency of the proposed algorithm—a linear predictor, an explicit residual grid, and confidence-bound decisions—can support such audits more readily than opaque black-box pricing models.

C Proofs for the Redirect-UCB Analysis

This appendix gives the full proof of Theorem 1. We use the notation of Sections 3–4. Throughout, C, C', C_B denote positive problem-dependent numerical constants whose values may change from line to line but do not depend on T.

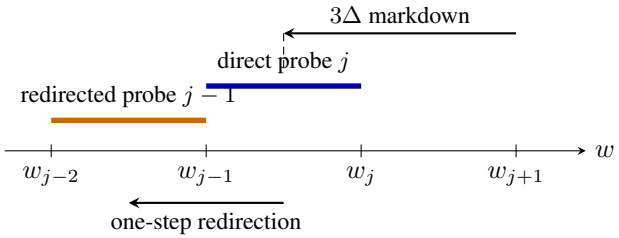

Figure 2: Residual-grid schematic. With 2Δ grid spacing and $\pm\Delta$ prediction error, a 3Δ markdown places the direct j-probe in $[w_{j-1}, w_j]$. Once $b_j \le \Delta$, the algorithm instead probes the interval one step to the left.

C.1 Notation reference

Symbol	Meaning
T_1, T_w	Stage 1 estimation and Stage 2 warmup durations
$u_t, \hat{u}_t, \theta^\star, \hat{\theta}$	True and estimated linear valuations and parameters
$S(w), \alpha_j$	Residual survival function and adjacent jump height
$\widehat{\Sigma}_1, \underline{\lambda}_1$	Empirical covariance and observable eigenvalue lower bound
$\kappa, \Delta, w_j, \mathcal{I}$	Price buffer, grid scale, grid points, query indices
$n_j(t), \hat{m}_j(t), \bar{m}_j(t)$	Played-probe count, empirical mean, predictable mean
$b_j(t), U_{t,j}$	Confidence radius and optimistic score
$\mathcal{E}_\theta, \mathcal{E}_{\text{conf}}$	Parameter-estimation and confidence events

C.2 Good events

Recall the Stage 1 prediction event

$$\mathcal{E}_\theta := \left\{ \max_{1 \le t \le T} |\hat{u}_t - u_t| \le \Delta \right\}.$$

For each queried index j, let $\mathcal{T}_j(t)$ be the set of all rounds before t in which probe action j was actually played. This includes Stage 2 warmup probes, Stage 3 direct probes, and Stage 3 redirected probes from $j+1$ to j. Let

$$n_j(t) := |\mathcal{T}_j(t)|, \qquad \hat{m}_j(t) := \begin{cases} n_j(t)^{-1} \sum_{s \in \mathcal{T}_j(t)} o_s, & n_j(t) > 0, \\ 0, & n_j(t) = 0. \end{cases}$$

For $n_j(t) > 0$, define the predictable average

$$\bar{m}_j(t) := \frac{1}{n_j(t)} \sum_{s \in \mathcal{T}_j(t)} \mathbb{E}[o_s \mid \mathcal{H}_{s-1}, x_s, \text{probe } j].$$

If $n_j(t) = 0$, set $\bar{m}_j(t) = 0$. Define

$$b_j(t) := C_{\text{ucb}} \sqrt{\frac{\log T}{\max\{1, n_j(t)\}}}.$$

Let $\mathcal{E}_{\text{conf}}$ denote the event that, simultaneously for every Stage 3 round t and every $j \in \mathcal{I}$,

$$|\hat{m}_j(t) - \bar{m}_j(t)| \le \frac{1}{2} b_j(t).$$

C.3 Proof of Lemma 2

Proof. In Stage 1, prices are sampled independently as $p_t \sim \text{Unif}[0, B]$. Since $y_t \in [0, B]$ almost surely under Assumption 1,

$$\mathbb{E}[o_t \mid x_t, \xi_t] = \mathbb{P}(p_t \le y_t \mid x_t, \xi_t) = \frac{y_t}{B}.$$

Thus, with $z_t := Bo_t$,

$$\mathbb{E}[z_t \mid x_t] = \mathbb{E}[y_t \mid x_t] = \langle x_t, \theta^\star \rangle + \mathbb{E}[\xi_t \mid x_t] = \langle x_t, \theta^\star \rangle,$$

where the last equality uses Assumption 1. Therefore the Stage 1 regression model is

$$z_t = x_t^\top \theta^\star + \eta_t, \qquad \mathbb{E}[\eta_t \mid x_t] = 0,$$

with bounded noise $|\eta_t| \le 2B$.

Let

$$\widehat{\Sigma}_1 := \frac{1}{T_1} \sum_{t=1}^{T_1} x_t x_t^\top, \qquad \Sigma := \mathbb{E}[x_t x_t^\top].$$

Write $\lambda_\star := \lambda_{\min}(\Sigma) \ge \lambda_0$. By matrix Bernstein for bounded random-design covariates, for $T_1 \gtrsim (B_x^4/\lambda_0^2) d \log T$,

$$\lambda_{\min}(\widehat{\Sigma}_1) \ge \lambda_\star/2$$

with probability at least $1 - T^{-3}$. More specifically, for a sufficiently large universal C_Σ, the matrix-concentration event implies $\|\widehat{\Sigma}_1 - \Sigma\|_{\mathrm{op}} \le r_1$. It follows that $\underline{\lambda}_1 \le \lambda_\star$; when $r_1 \le \lambda_\star/4$, also $\underline{\lambda}_1 \ge \lambda_\star - 2r_1 \ge \lambda_\star/2$. On this event, the Moore–Penrose estimator in Algorithm 1 equals the usual inverse estimator, and

$$\widehat{\theta} - \theta^\star = \widehat{\Sigma}_1^{-1} \left(\frac{1}{T_1} \sum_{t=1}^{T_1} x_t \eta_t \right).$$

A vector Bernstein or coordinate-wise Hoeffding argument gives, with probability at least $1 - T^{-3}$,

$$\left\| \frac{1}{T_1} \sum_{t=1}^{T_1} x_t \eta_t \right\|_2 \le CBB_x \sqrt{\frac{d \log T}{T_1}}.$$

Combining the two displays,

$$\|\widehat{\theta} - \theta^\star\|_2 \le \frac{CBB_x}{\lambda_\star} \sqrt{\frac{d \log T}{T_1}}.$$

Thus

$$\max_{1 \le t \le T} |\hat{u}_t - u_t| \le B_x \|\widehat{\theta} - \theta^\star\|_2 \le C \frac{BB_x^2}{\lambda_\star} \sqrt{\frac{d \log T}{T_1}} \le \Delta$$

by the choice of C_θ, since $\underline{\lambda}_1 \le \lambda_\star$ and the chosen grid scale is conservative. For sufficiently large horizons the same covariance event gives $\underline{\lambda}_1 \ge \lambda_\star/2$, so Δ grows by at most a factor two relative to the population-eigenvalue construction; the fallback is not invoked. Smaller horizons are absorbed into the problem-dependent regret constant. A union bound over the covariance and noise events proves the lemma. □

C.4 Proof of Lemma 3

Proof. On $\mathcal{E}_\theta$, $|\hat{u}_t - u_t| \le \Delta$. When probe index j is played in Stages 2–3, Assumption 3 ensures that clipping does not occur, so

$$p_t = \hat{u}_t + w_{j+1} - 3\Delta.$$

Since $w_{j+1} = w_j + 2\Delta$ and $w_{j-1} = w_j - 2\Delta$,

$$p_t - u_t = w_{j+1} - 3\Delta + (\hat{u}_t - u_t) \in [w_j - 2\Delta, w_j] = [w_{j-1}, w_j].$$

Because S is non-increasing,

$$S(w_j) \le S(p_t - u_t) \le S(w_{j-1}).$$

The middle term equals $\mathbb{E}[o_t \mid \mathcal{H}_{t-1}, x_t, \text{probe } j]$ by the definition of S, proving the claim. □

C.5 Proof of Lemma 4

Proof. Fix a queried index j. Let $\tau_{j,m}$ be the stopping time of the m-th play of probe action j, where a play may arise from Stage 2, a direct Stage 3 probe, or a redirected Stage 3 probe from $j+1$ to j. Define

$$Y_{j,m} := o_{\tau_{j,m}}, \qquad \mu_{j,m} := \mathbb{E}[o_{\tau_{j,m}} \mid \mathcal{H}_{\tau_{j,m}-1}, x_{\tau_{j,m}}, \text{probe } j].$$

Then $Y_{j,m} - \mu_{j,m}$ is a bounded martingale-difference sequence with respect to the stopped filtration generated by the probe times. The adaptivity of the probe times is harmless because the action is chosen before observing $o_{\tau_{j,m}}$.

For any fixed j and m, Hoeffding–Azuma gives

$$\mathbb{P}\left(\left|\sum_{r=1}^{m}(Y_{j,r} - \mu_{j,r})\right| \geq C\sqrt{m \log T}\right) \leq 2T^{-5}$$

for a sufficiently large universal constant C. Taking a union bound over at most $M \leq T$ queried indices and $m \leq T$ sample sizes yields, with probability at least $1 - T^{-2}$,

$$\left|\frac{1}{m}\sum_{r=1}^{m} Y_{j,r} - \frac{1}{m}\sum_{r=1}^{m}\mu_{j,r}\right| \leq \frac{1}{2}C_{\mathrm{ucb}}\sqrt{\frac{\log T}{m}}$$

for all j and $m \geq 1$, after increasing C_{ucb}. This is exactly the event $\mathcal{E}_{\mathrm{conf}}$.

On $\mathcal{E}_\theta$, Lemma 3 gives $S(w_j) \leq \mu_{j,m} \leq S(w_{j-1})$ for every probe of index j. Hence $S(w_j) \leq \bar{m}_j(t) \leq S(w_{j-1})$. Combining this with $|\hat{m}_j(t) - \bar{m}_j(t)| \leq b_j(t)/2$ gives

$$\hat{m}_j(t) + b_j(t) \geq \bar{m}_j(t) + \frac{1}{2}b_j(t) \geq S(w_j),$$

and

$$\hat{m}_j(t) + b_j(t) \leq \bar{m}_j(t) + \frac{3}{2}b_j(t) \leq S(w_{j-1}) + 2b_j(t),$$

where the constant 2 is used for notational simplicity. This proves the lemma. □

C.6 Proof of Lemma 5

Proof. Fix a Stage 3 round t, and let $w_t^\star \in [-c, c]$ be an optimal residual. Choose $q \in \mathcal{I}$ such that

$$w_t^\star \in [w_q, w_{q+1}].$$

Such an index exists because $w_0 = -c$ and $w_M \geq c$. On $\mathcal{E}_\theta$,

$$\hat{u}_t + w_{q+1} + \Delta \geq u_t + w_{q+1} \geq u_t + w_t^\star.$$

Because S is non-increasing and $w_q \leq w_t^\star$,

$$S(w_q) \geq S(w_t^\star).$$

By Lemma 4,

$$\hat{m}_q(t) + b_q(t) \geq S(w_q) \geq S(w_t^\star).$$

Since $S(w_t^\star) \leq 1$, the truncation does not hurt this lower bound:

$$\min\{1, \hat{m}_q(t) + b_q(t)\} \geq S(w_t^\star).$$

Therefore

$$U_{t,q} = (\hat{u}_t + w_{q+1} + \Delta)\min\{1, \hat{m}_q(t) + b_q(t)\} \geq (u_t + w_t^\star)S(w_t^\star) = V_t^\star.$$

Taking the maximum over j proves the claim. □

C.7 Proof of Lemma 6

Proof. By Lemma 5 and the choice of j_t,

$$V_t^\star \leq U_{t,j_t}.$$

We compare this selected score with the revenue of the action actually played.

Direct-probe mode. Let $j = j_t$. The algorithm plays probe j, so

$$p_t = \hat{u}_t + w_{j+1} - 3\Delta, \qquad a_{t,j} := \hat{u}_t + w_{j+1} + \Delta = p_t + 4\Delta.$$

Let $m_{t,j}$ be the conditional purchase probability of this played action. Since $\min\{1, x\} \le x$ for all x,

$$U_{t,j} \le (p_t + 4\Delta)(\hat{m}_j + b_j).$$

On $\mathcal{E}_{\text{conf}}$ and by Lemma 3,

$$\hat{m}_j + b_j \le S(w_{j-1}) + 2b_j.$$

Moreover $m_{t,j} \ge S(w_j)$, so

$$S(w_{j-1}) = S(w_j) + \alpha_j \le m_{t,j} + \alpha_j.$$

Thus

$$\hat{m}_j + b_j \le m_{t,j} + \alpha_j + 2b_j.$$

Therefore

$$\begin{aligned} V_t^\star - \pi(x_t, p_t) &\le U_{t,j} - p_t m_{t,j} \\ &\le (p_t + 4\Delta)(m_{t,j} + \alpha_j + 2b_j) - p_t m_{t,j} \\ &\le 4\Delta + (B + 4\Delta)\alpha_j + 2(B + 4\Delta)b_j. \end{aligned}$$

For all sufficiently large horizons $\Delta \le 1$; constants are absorbed into C and C_B. This proves

$$V_t^\star - \pi(x_t, p_t) \le C\Delta + B\alpha_j + C_B b_j.$$

Redirect mode, $j_t \ge 1$. Let $j = j_t \ge 1$. The algorithm plays probe $j - 1$, so

$$p_t = \hat{u}_t + w_j - 3\Delta, \qquad a_{t,j} := \hat{u}_t + w_{j+1} + \Delta = p_t + 6\Delta.$$

Let $m^-_{t,j}$ be the conditional purchase probability of the played probe $j - 1$. By Lemma 3,

$$m^-_{t,j} \ge S(w_{j-1}).$$

For the selected score at index j, Lemma 4 gives

$$\hat{m}_j + b_j \le S(w_{j-1}) + 2b_j \le m^-_{t,j} + 2b_j.$$

Again using $\min\{1, x\} \le x$,

$$\begin{aligned} V_t^\star - \pi(x_t, p_t) &\le U_{t,j} - p_t m^-_{t,j} \\ &\le (p_t + 6\Delta)(m^-_{t,j} + 2b_j) - p_t m^-_{t,j} \\ &\le 6\Delta + 2(B + 6\Delta)b_j. \end{aligned}$$

Redirect mode occurs only when $b_j \le \Delta$, so the last display is at most $C_B\Delta$.

Boundary case $j_t = 0$. When $j_t = 0$, the algorithm plays probe 0. This is the direct-probe argument with

$$\alpha_0 = S(w_{-1}) - S(w_0) = 1 - 1 = 0,$$

because $w_0 = -c$ and $S(w_0) = 1$ under the support convention. In redirect mode, $b_0 \le \Delta$, so the regret is again at most $C_B\Delta$. □

C.8 Proof of Lemma 7

Proof. Let $\mathcal{P}$ be the set of Stage 3 rounds in direct-probe mode and $\mathcal{R}$ the set of Stage 3 rounds in redirect mode or the boundary case. By Lemma 6, the redirect rounds contribute at most

$$\sum_{t \in \mathcal{R}} C_B\Delta \le C_B T\Delta = \widetilde{O}(T^{2/3}).$$

For direct-probe rounds, Lemma 6 gives three terms. The grid-scale term is

$$\sum_{t \in \mathcal{P}} C\Delta \le CT\Delta = \widetilde{O}(T^{2/3}).$$

For the confidence-radius term, each direct probe of index j increments n_j. Hence

$$\begin{aligned}\sum_{t\in\mathcal{P}} b_{j_t}(t) &\le C_{\mathrm{ucb}}\sqrt{\log T}\sum_{j=0}^{M-1}\sum_{m=1}^{n_j(T)}\frac{1}{\sqrt{m}}\\ &\le 2C_{\mathrm{ucb}}\sqrt{\log T}\sum_{j=0}^{M-1}\sqrt{n_j(T)}\\ &\le 2C_{\mathrm{ucb}}\sqrt{\log T}\sqrt{M\sum_j n_j(T)}\\ &\le \widetilde{O}(\sqrt{MT}) = \widetilde{O}(T^{2/3}),\end{aligned}$$

where $M = O(\Delta^{-1}) = \widetilde{O}(T^{1/3})$.

For the adjacent-jump term, index j is directly probed only while

$$b_j(t) = C_{\mathrm{ucb}}\sqrt{\frac{\log T}{\max\{1, n_j(t)\}}} > \Delta.$$

Thus the number of direct probes of any fixed index j is at most $C_{\mathrm{ucb}}^2 \log T/\Delta^2 + 1$. Therefore

$$\sum_{t\in\mathcal{P}} \alpha_{j_t} \le \left(C\frac{\log T}{\Delta^2} + 1\right)\sum_{j=0}^{M-1}\alpha_j \le C\frac{\log T}{\Delta^2} + 1 = \widetilde{O}(T^{2/3}),$$

using $\sum_j \alpha_j \le 1$. Combining the three direct-probe terms and the redirect contribution proves the lemma. □

C.9 Proof of Theorem 1

Proof. Stages 1 and 2 each last $O(T^{2/3})$ rounds and incur per-round regret at most B, so their expected contribution is $O(T^{2/3})$.

Let

$$\mathcal{E} := \mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}.$$

By Lemmas 2 and 4, $\mathbb{P}(\mathcal{E}^c) \le O(T^{-2})$. On $\mathcal{E}$, Lemma 7 gives Stage 3 regret $\widetilde{O}(T^{2/3})$. On $\mathcal{E}^c$, total regret is at most BT, so the failure-event contribution is at most $BT \cdot O(T^{-2}) = O(1/T)$. Therefore

$$R_T = \mathbb{E}[\mathcal{R}_T] \le \widetilde{O}(T^{2/3}).$$

□

Remark 8. The proof yields the following informal decomposition, suppressing logarithmic factors and problem-dependent constants:

$$R_T \;\lesssim\; B(T_1 + T_w) + T\Delta + \sqrt{MT\log T} + \frac{\log T}{\Delta^2},$$

where $M = O(c/\Delta)$ is the number of queried residual-grid indices and

$$\Delta = C_\theta \frac{BB_x^2}{\underline{\lambda}_1}\sqrt{\frac{d\log T}{T_1}}, \quad \underline{\lambda}_1 \ge \lambda_\star/2 \text{ on the covariance-good event.}$$

The four terms correspond respectively to the two initial stages, grid-scale loss, UCB confidence accumulation, and the cumulative adjacent-jump cost. Taking $T_1 = T_w = \lceil T^{2/3}\rceil$ and $M = O(\Delta^{-1})$ gives $\widetilde{O}(T^{2/3})$.

C.10 Lower-bound alignment under the assumptions of this paper

We briefly justify that the assumptions used in Theorem 1 do not remove the standard $T^{2/3}$ lower-bound barrier. The point is that a non-contextual posted-price hard instance can be embedded into our contextual linear-valuation model while satisfying bounded support, zero-mean noise, stochastic well-conditioned contexts, and the price-buffer condition.

Consider any one-dimensional posted-price lower-bound family supported on an interior valuation interval $[\underline{v}, \bar{v}] \subset (0, B)$ for which every pricing algorithm incurs $\Omega(T^{2/3})$ regret; the construction of Kleinberg and Leighton [2003] can be placed on such an interval by an affine rescaling of prices and valuations. Let V denote the valuation in such a hard instance and set $u_0 := \mathbb{E}[V]$, $\xi := V - u_0$. Then $\mathbb{E}[\xi] = 0$ and $\xi \in [-c, c]$ for $c := \max\{u_0 - \underline{v}, \bar{v} - u_0\}$.

Now construct a contextual instance in dimension $d \geq 2$ as follows. Let

$$x_t = (1, Z_{t,2}, \ldots, Z_{t,d}),$$

where $Z_{t,2}, \ldots, Z_{t,d}$ are independent Rademacher random variables, independent across t and independent of ξ_t. Let

$$\theta^\star = (u_0, 0, \ldots, 0).$$

Then $x_t^\top \theta^\star = u_0$, so the buyer valuation is $y_t = u_0 + \xi_t = V_t$, exactly the original non-contextual valuation. The context distribution is i.i.d. and well-conditioned, since $\mathbb{E}[x_t x_t^\top] = I_d$. The dummy covariates carry no valuation information, but they ensure that the full-rank context assumption is satisfied.

The bounded-support and normalization assumptions hold by construction. The price-buffer condition also holds whenever the hard instance is placed in an interior interval: $x_t^\top \theta^\star - c \geq \underline{v} > 0$ and $x_t^\top \theta^\star + c \leq \bar{v} \leq B$. Therefore this contextual instance satisfies Assumptions 1–3. Any contextual pricing algorithm for our model, when run on this instance, induces a non-contextual posted-price algorithm with the same revenue process and the same regret. Hence an $o(T^{2/3})$ regret bound for our model would contradict the non-contextual lower bound.

The contextual lower bound of Xu and Wang [2022] provides a complementary comparison: it shows that the same $T^{2/3}$ barrier already arises under Lipschitz demand in a contextual linear-valuation setting. The embedding above is sufficient for the formal minimax lower bound under the assumptions of this paper, while the Xu–Wang lower bound places our rate in the broader contextual-pricing landscape.

C.11 Experimental implementation details

Grid-scale calibration. For the reported experiments, the effective residual-grid scale is $\Delta = s\Delta_0$, where $\Delta_0 := 2\sqrt{d \log(T)/T_1}$, with default $s = 0.35$ and $C_{\mathrm{ucb}} = 4$. The grid-scale sensitivity diagnostic uses $s \in \{0.20, 0.35, 0.50, 0.70\}$ with the same normalization and without retuning other parameters. The markdown coefficient is fixed at 3 throughout. This numerical calibration is separate from the problem-dependent constant C_θ in the regret guarantee.

All reported experiments use pseudo-regret computed from the known synthetic survival function. For a posted price p_t and linear value u_t, the algorithm's expected revenue is $p_t S(p_t - u_t)$, where throughout the paper $S(w) = \Pr(\xi \geq w)$ uses the inclusive convention matching the purchase event $p_t \leq y_t$. The oracle benchmark is $\max_{p \in [0,B]} pS(p - u_t)$. For the uniform-noise instance this maximum is computed from the piecewise-quadratic form of the revenue curve. For the cliff-noise instance, where $\Pr(\xi = 0) = 0.3$, the candidate set additionally includes the atom price $p = u_t$.

Exact D2-EXP4 requires enumerating a discretized family of valuation parameters and noise distributions, which is computationally prohibitive at the horizons in Figure 1. We therefore use a sampled-policy implementation. Each sampled policy is indexed by a candidate linear parameter and a candidate monotone survival curve on a residual grid. In our runs, the default ensemble contains $K_{\mathrm{policy}} = 2048$ policies, sampled from $K_\theta = 256$ parameter candidates and $K_F = 64$ survival-function candidates. The parameter pool samples the intercept from $[1, 3]$ and the remaining coefficients from $[0, 0.3]$, together with a few structured anchor vectors. The survival-function pool contains uniform-like, step-like, and cliff-like structured candidates, supplemented by randomly generated monotone survival curves. EXP4 is then run over this sampled ensemble using an action grid

with spacing $\gamma = \mathrm{clip}(T^{-1/4}, 0.02, 0.10)$, learning rate $\eta = \min\{0.5, \sqrt{\log(K_{\mathrm{policy}})/(TK_{\mathrm{act}})}\}$, and explicit exploration probability $\min\{0.2, \sqrt{K_{\mathrm{act}} \log(K_{\mathrm{policy}})/T}\}$, where K_{act} is the number of price-grid actions. The displayed D2-EXP4 curves should therefore be read as a computationally tractable sampled-policy implementation of the D2-EXP4 reduction, not as an exhaustive implementation of the full exponentially large policy class.

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 main claims are stated in the abstract, introduction, and Theorem 1. The paper clearly specifies the assumptions under which CMRUP achieves $\widetilde{O}(T^{2/3})$ expected regret, and Section 6 states the scope of the numerical experiments.

   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: Section 7 and Appendix B discuss the main limitations, including bounded-support noise, i.i.d. well-conditioned contexts, the interior price-buffer condition, horizon-dependent tuning, and the fact that the experiments are rate-verification studies rather than a comprehensive empirical benchmark.

   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 full assumptions are stated in Section 3. The main theorem and proof roadmap are given in Section 5, and complete proofs are provided in the Appendix. The Appendix also includes details on the lower-bound embedding and the experimental oracle convention.

   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: [Yes]

   Justification: Section 6 specifies the synthetic data-generating process, context distribution, valuation parameter, two noise distributions, horizon ranges, number of random seeds, pseudo-regret metric, and CMRUP grid-scale multiplier. Appendix C.11 provides additional implementation details for the oracle and sampled-policy D2-EXP4 baseline.

   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: [Yes]

   Justification: We provide anonymized supplementary code for reproducing the synthetic experiments. The experiments use only simulated data generated from the distributions described in Section 6, so no external dataset is required.

   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: [Yes]

   Justification: Section 6 specifies the context distribution, dimension, valuation parameter, price bound, noise distributions, horizon ranges, random seeds, CMRUP tuning, and the D2-EXP4 baseline setup. The Appendix further describes the pseudo-regret oracle and sampled-policy D2-EXP4 implementation.

   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: [Yes]

Justification: Figure 1 reports means over independent random seeds, with CMRUP averaged over 10 seeds and D2-EXP4 averaged over 5 seeds. Error bars show standard errors across seeds. The fitted exponents are obtained by ordinary least-squares fits in log-log space using all available horizons for each algorithm.

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: [Yes]

   Justification: All reported experiments use synthetic data and were locally run on CPU-only machines. No GPU was used.

   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: The research conforms to the NeurIPS Code of Ethics. The paper is primarily theoretical and uses only synthetic experiments. Section 7 and Appendix B discuss ethical considerations related to contextual pricing, including fairness risks, protected-attribute proxies, auditing, and price constraints.

   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: [Yes]

    Justification: Section 7 and Appendix B discuss broader impacts. On the positive side, CMRUP is transparent and auditable because it uses a linear valuation model, a one-dimensional residual grid, and explicit confidence bounds. The paper also acknowledges that contextual pricing can raise fairness and disparate-impact concerns if context variables contain sensitive attributes or proxies, and discusses mitigation through feature auditing, price monitoring, admissible-price restrictions, and fairness-aware post-processing.

    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 a deployed pricing system, dataset, or model with high misuse risk. The experiments use only synthetic data, and the contribution is a theoretical online learning algorithm.

    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 existing datasets, pretrained models, or third-party 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: [Yes]

    Justification: We include anonymized supplementary code as a research asset for review. The supplementary package documents the synthetic data generator, algorithms, and experimental configurations. No new dataset, pretrained model, or benchmark is introduced; all data are synthetically generated. The code is provided for review and is not publicly released at submission time.

    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 nor research with human subjects.

    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 subjects, crowdsourcing, or user studies, so IRB approval or equivalent review is not applicable.

    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: [N/A]

    Justification: LLMs were used only for writing, editing, and formatting purposes, and were not an important, original, or non-standard component of the research.

    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.