OPTIMAL NONPARAMETRIC DYNAMIC PRICING WITH CENSORED DEMAND AND ADVERSARIAL INVENTORY

Anonymous authors
Paper under double-blind review

ABSTRACT

We study online dynamic pricing with censored demand, where an arbitrary inventory level is revealed before pricing and may adapt to past observations, while demand follows an unknown, price-dependent distribution that is stationary over time. For a horizon of T rounds, Xu et al. (2026) achieved $\widetilde{\mathcal{O}}(\sqrt{T})$ regret under restrictive structural assumptions including linear demand, price-independent additive noise, and conditions relating inventory levels to the noise support. Our first contribution is to extend this framework to a substantially more general and statistically harder nonparametric setting, requiring only the natural assumption that expected sales are nonincreasing in price and allowing nonlinear demand curves and price-dependent noise. For this model, we first propose a simple baseline, Double-Grid-UCB, which discretizes both price and inventory and achieves $\widetilde{\mathcal{O}}(T^{3/4})$ expected regret using separate revenue estimates for each price-inventory grid pair. Then, we develop Threshold-UCB, which improves the expected regret to $\widetilde{\mathcal{O}}(T^{2/3})$. Unlike Double-Grid-UCB, Threshold-UCB reuses sales observations across inventory levels through shared estimates of demand-tail probabilities, allowing the same data to support revenue upper bounds for multiple inventories rather than a single inventory bin. We also complement this upper bound with an $\Omega(T^{2/3})$ lower bound via a reduction from stochastic posted pricing, establishing its minimax optimality. Finally, extensive experiments across inventory processes, demand functions, and noise models demonstrate consistently superior performance of Threshold-UCB over benchmark algorithms.

1 INTRODUCTION

Dynamic pricing is a central tool in revenue management because price controls both the revenue earned per unit and the quantity demanded. Classical models characterize optimal prices when the demand law is known (Gallego & van Ryzin, 1994). Subsequent work has studied how a seller can learn an unknown demand curve while simultaneously earning revenue (Kleinberg & Leighton, 2003; Besbes & Zeevi, 2009; Broder & Rusmevichientong, 2012; Keskin & Zeevi, 2014), including contextual formulations in which demand depends on observed customer or product characteristics (Javanmard & Nazerzadeh, 2019; Xu & Wang, 2021; Luo et al., 2022; Xu & Wang, 2022). In these models, inventory is typically unlimited, fixed in advance, or controlled by the seller.

In many applications, however, available inventory is limited, varies across selling periods, and is outside the seller's control. Consider a retailer receiving a highly perishable crop from a nearby farm. Daily supply depends on when the crop ripens and can vary sharply, while unsold units spoil before the next period. From the retailer's perspective, this motivates treating the inventory sequence as arbitrary: in each period, the retailer observes the available inventory and then chooses a price. Demand, however, is censored by inventory: when a stockout occurs, the retailer observes only the number of units sold, but not the underlying demand. Beyond censoring demand, the varying inventory also affects optimal pricing: the revenue-maximizing price depends on the available inventory, so the relevant benchmark selects a different price for each observed inventory level.

Motivated by such applications, Xu et al. (2026) consider an adversarial-inventory model and propose an algorithm called C20CB, which achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret after T rounds and is minimax optimal

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

for their structured model up to logarithmic factors.[1] Their result, however, relies heavily on several restrictive assumptions. First, potential demand must be linear in price with fixed coefficients, plus additive noise. Second, the noise must be bounded, zero-mean, and i.i.d., with a common price-independent law and a Lipschitz CDF. Third, inventory must satisfy support and interiority conditions that ensure complete censoring at the lowest price and uncensored demand at the highest price. The latter guarantees that true demand, rather than only sales truncated by inventory, is observed at the highest price. Together, these assumptions enable estimation of the linear demand parameters and common noise distribution. However, the resulting algorithm and analysis do not directly extend to nonlinear demand or price-dependent noise, which arise naturally in practice. This leaves the following question open.

Can one obtain sublinear regret for adversarial-inventory dynamic pricing under nonparametric monotone demand in general? If so, what is the minimax-optimal regret rate?

Contributions. We provide a complete answer to this question. Our first contribution is to introduce a substantially more general, nonparametric adversarial-inventory model, which requires only the natural assumption that expected sales are nonincreasing in price, without imposing an additive-noise representation or other parametric structure. We then propose two algorithms for this model: the first, simpler algorithm achieves a suboptimal regret of $\widetilde{\mathcal{O}}(T^{3/4})$, while the second improves the rate to $\widetilde{\mathcal{O}}(T^{2/3})$. Both algorithms follow the upper confidence bound (UCB) principle, choosing the price with the largest optimistic estimate of its revenue under the current inventory. Specifically,

- In Section 3.1, we introduce Double-Grid-UCB, which discretizes both price and inventory and uses separate revenue estimates for each price-inventory pair to select prices optimistically. It achieves $\widetilde{\mathcal{O}}(PDT^{3/4})$ expected regret (Theorem 3.1), where P and D are the largest possible price and inventory, respectively. However, learning each inventory bin separately leaves useful information from other inventory levels unused.
- To improve this rate, in Section 3.2, we exploit the common demand distribution across inventory levels at each price. Specifically, we first consider the classical Kaplan-Meier estimator and show that its confidence guarantee *does not* directly apply to our setting with adaptive censoring. We therefore develop Threshold-UCB, which shares each sales observation across multiple demand thresholds and achieves $\widetilde{\mathcal{O}}(PDT^{2/3})$ expected regret (Theorem 3.2). A matching $\Omega(T^{2/3})$ lower bound for $P = D = 1$ establishes minimax optimality up to logarithmic factors (Theorem C.1).
- Finally, in Section 4, we compare Threshold-UCB with C20CB (Xu et al., 2026) and other benchmarks across linear and nonlinear demand, varying inventories, and additive and multiplicative noise. Threshold-UCB is competitive with C20CB in the regime covered by C20CB's theory and achieves lower mean regret in most tested regimes outside that regime.

1.1 Related Work

Dynamic pricing and posted-price learning. Dynamic pricing is a classical problem in revenue management, with foundational work on finite inventories under a known stochastic demand law (Gallego & van Ryzin, 1994). When demand is unknown, Kleinberg & Leighton (2003) study online posted pricing and establish regret upper and lower bounds under identical, i.i.d., or adversarial buyer valuations. Our lower bound builds directly on their $\Omega(T^{2/3})$ construction under i.i.d. valuations, which we embed into our model through constant inventory and binary demand. Research on pricing and learning also considers different demand structures, ranging from linear models (Keskin & Zeevi, 2014) to general parametric choice models (Broder & Rusmevichientong, 2012) and nonparametric demand (Besbes & Zeevi, 2009). Contextual pricing further incorporates customer and product features into pricing decisions, including work on linear valuation models under known noise distributions (Javanmard & Nazerzadeh, 2019; Xu & Wang, 2021) and extensions to unknown noise distributions under weaker assumptions (Luo et al., 2022; Xu & Wang, 2022).

Learning from lost sales. Censoring also plays a central role in data-driven inventory control, where sales reveal demand only up to the available inventory. Its implications for learning have been

[1] We use $\widetilde{\mathcal{O}}(\cdot)$ to suppress logarithmic factors.

studied in the newsvendor problem (Besbes & Muharremoglu, 2013). Bayesian approaches update demand beliefs from censored sales (Chen, 2010; Bisi et al., 2011), while nonparametric approaches use stochastic gradient methods or Kaplan–Meier estimation to construct adaptive base-stock policies without observing unmet demand (Huh et al., 2009; 2011). When prices are also decision variables, joint pricing and inventory control methods learn nonparametric demand while choosing both prices and stocking levels (Chen et al., 2021; 2024). Within this joint-decision setting, Chen et al. (2020) additionally consider a constraint on the number of price changes under a parametric demand model. In contrast to this literature, our learner does not choose inventory or use inventory decisions to control censoring. Instead, inventory is revealed as a context before pricing and may adapt to past sales, while the learner chooses a price based on the current inventory.

Adversarially censored dynamic pricing. Closest to our work is Xu et al. (2026), which studies the same within-period order of inventory, price, and censored sales and permits an adversarial inventory sequence. Their algorithm achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret under a more restrictive model that assumes linear mean demand, a common bounded price-independent additive-noise CDF, regularity of that CDF, and inventory support conditions. The shared additive shock makes potential demand automatically nonincreasing in price and enables observations at different prices to be pooled to estimate common demand parameters and a common noise law. Our work removes this parametric structure by considering a broader nonparametric model, for which the worst-case regret rate is necessarily $T^{2/3}$. Thus, their $\Omega(\sqrt{T})$ lower bound and our $\Omega(T^{2/3})$ lower bound apply to different model classes and do not conflict. See Section 2.1 for detailed comparisons between the two models.

2 Problem Formulation

Notation. For a positive integer n, let $[n] \triangleq \{1, \dots, n\}$. We write $\mathbf{1}\{A\}$ for the indicator of an event A, and use $\mathbb{E}[\cdot]$ and $\mathbb{P}(\cdot)$ for expectation and probability. For $x \in \mathbb{R}$, let $x_+ \triangleq \max\{x, 0\}$.

Problem setting. The learner and the environment interact for $T \geq 2$ rounds. Let $\mathcal{F}_{t-1}$ denote the sigma-algebra generated by the learner's internal random seed and all inventories, prices, and sales observed through round $t-1$. At the beginning of round t, the environment reveals an inventory level $\gamma_t \in [0, D]$, where $D > 0$ is the known maximum inventory. The conditional law of γ_t may depend on $\mathcal{F}_{t-1}$, allowing inventory to adapt to the observed past. The learner then selects a price $p_t \in [0, P]$ using $\mathcal{F}_{t-1}$ and γ_t, where $P > 0$ is the maximum allowed price.

After the price is posted, a nonnegative potential demand $Y_t(p_t)$ is realized, where $Y_t(p)$ denotes demand at price p before inventory censoring. The learner observes the sale $X_t \triangleq \min\{Y_t(p_t), \gamma_t\}$ and earns realized revenue $r_t \triangleq p_t X_t$. For the analysis, define the pre-demand sigma-algebra $\mathcal{G}_t \triangleq \sigma(\mathcal{F}_{t-1}, \gamma_t, p_t)$. The random functions $Y_t(\cdot) : [0, P] \mapsto \mathbb{R}^+$ have the same law in every round, and $Y_t(\cdot)$ is independent of $\mathcal{G}_t$. Thus, inventory and price may depend on the observed past, but neither can anticipate the current demand shock. We impose no restriction on dependence across prices within a round, as sales feedback is available only at the posted price.

For $p \in [0, P]$, define the demand CDF and survival function by $F_p(u) \triangleq \mathbb{P}(Y_t(p) \leq u)$ and $S_p(u) \triangleq 1 - F_p(u) = \mathbb{P}(Y_t(p) > u)$, respectively. At inventory γ, we define the expected sales and revenue by $s(p, \gamma) \triangleq \mathbb{E}[\min\{Y_t(p), \gamma\}] = \int_0^\gamma S_p(u)\,\mathrm{d}u$ and $R(p, \gamma) \triangleq p \cdot s(p, \gamma)$. The integral representation follows from the layer-cake identity for a nonnegative random variable.

The only assumption we make in our model is the following natural cross-price monotonicity condition on expected sales.

Assumption 2.1 (Monotone expected sales)**.** *For every inventory level $\gamma \in [0, D]$, the expected sales $s(p, \gamma)$ is nonincreasing in p on $[0, P]$. Equivalently, $s(p, \gamma) \geq s(p', \gamma)$ for every $0 \leq p \leq p' \leq P$.*

Assumption 2.1 is implied by first-order stochastic dominance of potential demand across prices. Specifically, if $p \leq p'$ implies $S_p(u) \geq S_{p'}(u)$ for every $u \geq 0$, then integrating this inequality over $[0, \gamma]$ gives the required monotonicity. In particular, the assumption holds whenever potential demands at different prices can be constructed using the same underlying randomness so that $p \mapsto Y_t(p)$ is nonincreasing almost surely. For example, if $Y_t(p) = \mu(p) + \epsilon_t$ where $\mu(p)$ is nonincreasing in p, then using the same realization of ϵ_t for all prices ensures that $Y_t(p) \geq Y_t(p')$ whenever $p \leq p'$.

Assumption 2.1 is used only for the later continuous-to-discrete approximation, and our model requires no smoothness, parametric form, or additional distributional relation for demand.

Goal. Our goal is to minimize expected regret relative to an oracle that knows the demand law and optimizes expected revenue separately for each realized inventory. More concretely, we define regret as the oracle's cumulative expected revenue minus that of the learner:

$$\mathrm{Reg}_T \triangleq \sum_{t=1}^{T} \max_{p\in[0,P]} R(p,\gamma_t) - \sum_{t=1}^{T} R(p_t,\gamma_t). \tag{1}$$

All our algorithms use the uniform price grid defined by $\mathcal{P}_K \triangleq \{p_m = (m-1)P/K : m \in [K+1]\}$, which contains $K+1$ prices. The integer $K \geq 1$ may differ across algorithms. Since $\gamma_t \leq D$ for all t, we have $\sum_{t=1}^{T} \gamma_t \leq DT$. By Assumption 2.1, restricting the oracle to $\mathcal{P}_K$ incurs an additive approximation error of at most PDT/K. Specifically, for $p \in [0,P]$, let $\underline{p}$ be the largest grid price no greater than p. By Assumption 2.1, $s(\underline{p},\gamma) \geq s(p,\gamma)$. Moreover, $p - \underline{p} \leq P/K$ and $s(p,\gamma) \leq \gamma$, so $R(\underline{p},\gamma) \geq \underline{p}\, s(p,\gamma) \geq R(p,\gamma) - \frac{P\gamma}{K}$. Taking the maximum over all prices in each round gives

$$\mathrm{Reg}_T \leq \frac{PDT}{K} + \sum_{t=1}^{T}\left[\max_{m\in[K+1]} R(p_m,\gamma_t) - R(p_t,\gamma_t)\right] \triangleq \frac{PDT}{K} + \mathrm{Reg}_T^{\mathcal{P}_K}. \tag{2}$$

The first term is the price-grid approximation error, while $\mathrm{Reg}_T^{\mathcal{P}_K}$ is the regret relative to the best grid price at each round. We emphasize again that this continuous-to-discrete approximation is the only place where Assumption 2.1 is utilized, and the rest of the analysis does not require it.

2.1 Relation to the linear-additive model of Xu et al. (2026).

In this section, we discuss how our model is substantially more general and challenging than that of Xu et al. (2026). In our notation, their model assumes $Y_t(p) = a - bp + N_t$ for $p \in [0,P]$, with unknown parameters $0 < a \leq a_{\max}$ and $0 < b_{\min} \leq b \leq b_{\max}$. The additive noise variables N_t are i.i.d. with zero mean and follow a common, price-independent distribution supported on $[-c,c]$ with a Lipschitz CDF. The constants $a_{\max}, b_{\min}, b_{\max}, c, P$ and additionally an inventory lower bound $\gamma_{\min}$ are assumed to be known. Their inventory conditions require, for every round t,

$$2c < \gamma_t < a - c, \qquad \gamma_t \geq \gamma_{\min} > a_{\max} - b_{\min}P + c. \tag{3}$$

They additionally require $a - bP - c > 0$ and $P \geq a/(2b)$. Thus, inventory exceeds the width of the noise support but remains below the smallest possible demand at price zero. At the maximum price P, by contrast, even the largest possible demand lies strictly below every admissible inventory. Note that the condition $b > 0$ also ensures that Assumption 2.1 holds. Their C20CB algorithm achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret under these assumptions when the horizon is sufficiently large.

These conditions are restrictive, and they also make learning much easier. The known gap in Eq. (3) guarantees a nonempty price interval ending at P on which demand is never censored, whatever the inventory. Uncensored observations at two prices in this interval identify a, b, and the noise distribution. A linear shift of these observations then gives an unbiased revenue estimate at every price and inventory level, so censoring never has to be handled (see Appendix A for details). Our model, by contrast, assumes only monotone expected sales and imposes no additive-noise representation, allowing nonlinear, heteroskedastic, and price-dependent demand distributions. The $\Omega(T^{2/3})$ lower bound in Appendix C shows the statistical cost of this broader class, even when inventory is constant and censoring disappears.

3 Algorithm

In this section, we introduce two upper confidence bound (UCB) algorithms for our problem. As a warm-up, in Section 3.1, we introduce Double-Grid-UCB, a simple baseline that discretizes both price and inventory, maintains separate revenue estimates for each price-inventory grid pair, and achieves $\widetilde{\mathcal{O}}(PDT^{3/4})$ expected regret. In Section 3.2, we develop Threshold-UCB, which improves this rate to $\widetilde{\mathcal{O}}(PDT^{2/3})$ by reusing sales observations across inventory levels through shared estimates of

demand-tail probabilities. Below, we first present a general UCB framework and its regret analysis that are shared by the two algorithms.

Fix the price grid $\mathcal{P}_K = \{p_m = (m-1)P/K : m \in [K+1]\}$. After observing γ_t at round t, the algorithm computes an index $\mathsf{U}_t(m, \gamma_t)$ for each price $p_m \in \mathcal{P}_K$. Suppose that we can construct a uniform confidence event $\mathcal{E}$ with $\Pr(\mathcal{E}^c) \le 1/T$, on which

$$R(p_m, \gamma_t) \le \mathsf{U}_t(m, \gamma_t), \tag{4}$$

for every $m \in [K+1]$ and $t \in [T]$. Thus, $\mathsf{U}_t(m, \gamma_t)$ is an upper confidence bound on the expected revenue at price p_m under the current inventory. The algorithm simply selects $m_t \in \arg\max_{m \in [K+1]} \mathsf{U}_t(m, \gamma_t)$, posts price $p_t = p_{m_t}$, and updates its estimates after seeing the sale X_t.

Let $m_t^\star \in \arg\max_{m \in [K+1]} R(p_m, \gamma_t)$ denote the index of a best grid price for the current inventory. On $\mathcal{E}$, the finite-grid regret in round t satisfies

$$\begin{aligned} &R(p_{m_t^\star}, \gamma_t) - R(p_t, \gamma_t) \\ &= \big[R(p_{m_t^\star}, \gamma_t) - \mathsf{U}_t(m_t^\star, \gamma_t)\big] + \big[\mathsf{U}_t(m_t^\star, \gamma_t) - \mathsf{U}_t(m_t, \gamma_t)\big] + \big[\mathsf{U}_t(m_t, \gamma_t) - R(p_t, \gamma_t)\big] \\ &\le \mathsf{U}_t(m_t, \gamma_t) - R(p_t, \gamma_t), \end{aligned} \tag{5}$$

where the inequality holds because the first bracketed term is nonpositive by Eq. (4), and the second is nonpositive because m_t maximizes the index. It therefore remains to bound the overestimation error at the selected price. For nonnegative widths satisfying $\mathsf{U}_t(m, \gamma_t) - R(p_m, \gamma_t) \le w_t(m, \gamma_t)$ on $\mathcal{E}$, the finite-grid regret in round t is thus at most $w_t(m_t, \gamma_t)$. Moreover, on every sample path, we have $0 \le \mathrm{Reg}_T \le PDT$, since $\gamma_t \le D$ and revenue is at most $P\gamma_t$ in each round. The failure event therefore contributes at most $PDT \Pr(\mathcal{E}^c) \le PD$ to expected regret. Combining this observation with the price-grid approximation gives

$$\mathbb{E}[\mathrm{Reg}_T] \le \frac{PDT}{K} + \mathbb{E}[\mathrm{Reg}_T^{\mathcal{P}_K}] \le \frac{PDT}{K} + \mathbb{E}\left[\sum_{t=1}^{T} w_t(m_t, \gamma_t)\right] + PD. \tag{6}$$

It remains to construct UCB indices satisfying Eq. (4) and obtain a tight bound on the expected sum of the width $w_t(m, \gamma_t)$.

3.1 Warm-Up: Double-Grid-UCB with $\widetilde{\mathcal{O}}(T^{3/4})$ expected regret

As a warm-up, we first instantiate the preceding UCB framework with Double-Grid-UCB, which maintains a separate revenue estimate for each price and inventory bin, as summarized in Algorithm 2 in Appendix B.1. Specifically, in addition to the price grid, we divide the inventory interval $[0, D]$ into J bins of width $\Delta \triangleq D/J$. Define the associated grid points by $u_j \triangleq j\Delta$ for $j \in \{0, \dots, J\}$ and the bin assignment by $\ell(\gamma) \triangleq \min\{J, 1 + \lfloor \gamma/\Delta \rfloor\}$ for $\gamma \in [0, D]$. The lower representative of bin $\ell \in [J]$ is $u_{\ell-1}$, so we have $u_{\ell(\gamma)-1} \le \gamma$ and $\gamma - u_{\ell(\gamma)-1} \le \Delta$. After observing γ_t, the algorithm rounds it down to $u_{\ell_t - 1}$, where $\ell_t \triangleq \ell(\gamma_t)$. Since $0 \le R(p, \gamma_t) - R(p, u_{\ell_t - 1}) \le P\Delta$ for every $p \in [0, P]$, the finite-grid regret satisfies

$$\mathrm{Reg}_T^{\mathcal{P}_K} \le \frac{PDT}{J} + \sum_{t=1}^{T} \left[\max_{m \in [K+1]} R(p_m, u_{\ell_t - 1}) - R(p_t, u_{\ell_t - 1})\right]. \tag{7}$$

Thus, after paying the inventory-rounding cost PDT/J, it suffices to apply the preceding UCB argument at the grid representatives. To estimate the revenue at the representative inventory level $u_{\ell_t - 1}$, we exploit the fact that $u_{\ell_t - 1} \le \gamma_t$. Although demand $Y_t(p_t)$ is only observed through the censored sales $X_t = \min\{Y_t(p_t), \gamma_t\}$, censoring at γ_t does not affect its truncation at $u_{\ell_t - 1}$. Hence, we can construct the observable clipped revenue $p_t \min\{X_t, u_{\ell_t - 1}\} = p_t \min\{Y_t(p_t), u_{\ell_t - 1}\}$, where the equality follows from $u_{\ell_t - 1} \le \gamma_t$. This quantity therefore provides an uncensored observation of the revenue associated with truncating demand at $u_{\ell_t - 1}$. For each price-inventory pair (m, ℓ), denote by $N_{t,m,\ell}^{\mathrm{dg}}$ its number of "visits" (that is, event of posting price p_m at rounded inventory $u_{\ell-1}$) before round t, and by $Q_{t,m,\ell}^{\mathrm{dg}}$ the sum of the corresponding clipped revenues. Define $\widehat{R}_{t,m,\ell}^{\mathrm{dg}} \triangleq Q_{t,m,\ell}^{\mathrm{dg}} / \max\{1, N_{t,m,\ell}^{\mathrm{dg}}\}$ as the empirical estimate of $R(p_m, u_{\ell-1})$. Define the confidence radius by

$$C_{t,m,\ell}^{\mathrm{dg}} \triangleq PD \sqrt{\frac{\log(2(K+1)JT^2)}{2\max\{1, N_{t,m,\ell}^{\mathrm{dg}}\}}}. \tag{8}$$

Then a standard Hoeffding bound shows that, with probability at least $1 - 1/T$,

$$\left|\widehat{R}^{\mathrm{dg}}_{t,m,\ell} - R(p_m, u_{\ell-1})\right| \le C^{\mathrm{dg}}_{t,m,\ell}, \tag{9}$$

for every $m \in [K+1]$, $\ell \in [J]$, and $t \in [T]$. Denote this confidence event by $\mathcal{E}$. This confidence bound allows us to construct a revenue UCB for each price p_m at inventory representative $u_{\ell-1}$ as $\mathsf{U}^{\mathrm{dg}}_t(m, u_{\ell-1}) \triangleq \widehat{R}^{\mathrm{dg}}_{t,m,\ell} + C^{\mathrm{dg}}_{t,m,\ell}$. The algorithm then selects the price maximizing $\mathsf{U}^{\mathrm{dg}}_t(m, u_{\ell_t-1})$ and updates only the selected pair (m_t, ℓ_t) after receiving X_t. Moreover, Eq. (9) implies that, on $\mathcal{E}$, we have $0 \le \mathsf{U}^{\mathrm{dg}}_t(m, u_{\ell-1}) - R(p_m, u_{\ell-1}) \le 2C^{\mathrm{dg}}_{t,m,\ell}$, for every $m \in [K+1]$, $\ell \in [J]$, and $t \in [T]$. Thus, we can take $w_t(m, u_{\ell-1}) \triangleq 2C^{\mathrm{dg}}_{t,m,\ell}$ in the preceding UCB argument. By Eq. (5), the regret at the rounded inventory in round t is at most $w_t(m_t, u_{\ell_t-1})$. Summing these widths over visits to the $(K+1)J$ pairs and adding the inventory-rounding cost from Eq. (7), together with the price-grid and failure-event terms from Eq. (6), yields the following guarantee.

Theorem 3.1. *Suppose Assumption 2.1 holds. Algorithm 2 with* $K = J = \lceil T^{1/4} \log^{-1/4} T \rceil$ *guarantees that* $\mathbb{E}[\mathrm{Reg}_T] = \mathcal{O}(PDT^{3/4} \log^{1/4} T)$.

The complete proof is deferred to Appendix B.1. While Theorem 3.1 establishes the first sublinear regret guarantee for this problem, the resulting $T^{3/4}$ rate still leaves room for improvement. The loss arises from treating inventory bins separately: a finer grid reduces the inventory-rounding error PDT/J, but increases the estimation term by a factor of $\sqrt{J}$. In particular, each observation updates only a single inventory bin, even though revenues across different inventory levels are governed by the same underlying demand distribution. This suggests constructing revenue UCBs jointly across inventory levels, allowing observations to be shared rather than estimating each bin in isolation.

3.2 Threshold-UCB with $\widetilde{\mathcal{O}}(T^{2/3})$ expected regret

To share observations across inventory levels, we exploit the fact that inventory changes how demand is censored, but not the demand distribution at a fixed price. Specifically, since $R(p_m, \gamma) = p_m \int_0^\gamma S_{p_m}(u)\,\mathrm{d}u$, an estimate of the survival function S_{p_m} yields revenue estimates across inventory levels. The difficulty is that we observe only the censored sales $X_t = \min\{Y_t(p_t), \gamma_t\}$ rather than the demand realization itself.

Why Kaplan–Meier does not directly apply. A natural candidate is the Kaplan–Meier (KM) estimator (Kaplan & Meier, 1958), which is designed for right-censored data and has also been used for inventory learning (Huh et al., 2011). In the classical setting, one observes $W = \min\{Y, C\}$ together with the censoring indicator $\delta = \mathbf{1}\{Y \le C\}$, where Y is the quantity of interest and C is the censoring level. Given a collection of such observations, the KM estimator pools information across different censoring levels through the product-limit construction $\widehat{S}_{\mathrm{KM}}(y) = \prod_{x \le y} \left(1 - \frac{d(x)}{n(x)}\right)$, where $d(x)$ counts observations with $W = x$ and $\delta = 1$, and $n(x)$ counts observations with $W \ge x$.For i.i.d. observations with independent Y and C, results such as Bitouzé et al. (1999) provide uniform error bounds for the KM estimator.

However, the required i.i.d. censoring structure is not guaranteed in our setting. Indeed, the inventory γ_t is observed before the price p_t is selected, and both decisions may depend on past sales. Thus, the rule selecting censoring levels for a fixed price can change over time, and its retained observations need not form an i.i.d. sample. Current demand still remains independent of the pre-demand information, but this does not ensure an i.i.d. censoring history. Therefore, the classical KM confidence guarantee does not directly apply. We also evaluate a UCB-style baseline based on the KM estimator in Section 4, which indeed incurs substantially higher regret than our algorithm empirically.

Our algorithm: Threshold-UCB. We are now ready to present our key idea. As in Algorithm 2, we still keep the price grid $\mathcal{P}_K$ and optimistic price selection. However, instead of estimating revenue separately for each inventory bin, the key idea is to approximate $R(p_m, \gamma) = p_m \int_0^\gamma S_{p_m}(u)\,\mathrm{d}u$ by a sum of survival probabilities, whose estimates can be shared across inventory levels at each price. Specifically, set $\Delta = D/J$ and $u_j \triangleq j\Delta$ for $j \in \{0, 1, \ldots, J\}$, where $J \ge 2$, and define the

discretized revenue as

$$R_m^{(J)}(\gamma) \triangleq p_m \Delta \sum_{j \in [J]: u_{j-1} < \gamma} S_{p_m}(u_{j-1}). \tag{10}$$

We then estimate each $S_{p_m}(u_{j-1}) = 1 - F_{p_m}(u_{j-1})$ by $1 - \widehat{F}_{t,m,j}$, where $\widehat{F}_{t,m,j}$ estimates the CDF value $F_{p_m}(\cdot)$ at threshold u_{j-1}. These probabilities depend on the price and threshold, but not on inventory. Therefore, observations collected at different inventories can estimate the same probability, and changing γ only changes which estimates enter the revenue sum in Eq. (10).

To construct $\widehat{F}_{t,m,j}$, we use the fact that whenever $u_{j-1} < \gamma_t$, the sale reveals the indicator $\mathbf{1}\{Y_t(p_t) \le u_{j-1}\}$ exactly as $\mathbf{1}\{X_t \le u_{j-1}\}$. Denote by $N_{t,m,j}$ the number of earlier rounds $s < t$ with posted price index $m_s = m$ and inventory $\gamma_s > u_{j-1}$, and by $Z_{t,m,j}$ the number of those rounds with sales $X_s \le u_{j-1}$ (see the updates in Line 8 of Algorithm 1). We then define $\widehat{F}_{t,m,j} \triangleq Z_{t,m,j} / \max\{1, N_{t,m,j}\}$. Therefore, unlike Algorithm 2's revenue estimate for a single bin, this estimate pools all observations at price p_m whose inventories exceed u_{j-1}.

To see how accurate this estimate is, the same Hoeffding argument gives $|\widehat{F}_{t,m,j} - F_{p_m}(u_{j-1})| \le C^{\mathrm{th}}_{t,m,j}$ simultaneously for all $m \in [K+1]$, $j \in [J]$, and $t \in [T]$, with probability at least $1 - 1/T$, where $C^{\mathrm{th}}_{t,m,j}$ is initialized and updated in Lines 3 and 9 of Algorithm 1. For a fixed (m, j), whether round s contributes to the estimate depends only on the conditions $m_s = m$ and $\gamma_s > u_{j-1}$, both determined before current demand is realized. Each retained indicator therefore has conditional mean $F_{p_m}(u_{j-1})$, which is sufficient for the Hoeffding argument under adaptive sampling. In contrast, the classical KM confidence bound requires i.i.d. demand-censoring pairs.

With these estimators, we construct a UCB for the discretized revenue $R_m^{(J)}(\gamma)$ by combining the survival estimates $1 - \widehat{F}_{t,m,j}$ with their confidence radii $C^{\mathrm{th}}_{t,m,j}$. Specifically, we define $\mathsf{U}^{\mathrm{th}}_t(m, \gamma) \triangleq p_m \Delta \sum_{j \in [J]: u_{j-1} < \gamma} (1 - \widehat{F}_{t,m,j}) + B_{t,m}(\gamma)$, where $B_{t,m}(\gamma) \triangleq p_m \Delta \sum_{j \in [J]: u_{j-1} < \gamma} C^{\mathrm{th}}_{t,m,j}$ and empty sums equal zero. On this uniform confidence event, the preceding concentration bound gives $0 \le \mathsf{U}^{\mathrm{th}}_t(m, \gamma) - R_m^{(J)}(\gamma) \le 2B_{t,m}(\gamma)$, so we can take $w_t(m, \gamma) \triangleq 2B_{t,m}(\gamma)$ in the UCB argument for the discretized revenues. It therefore remains to bound the sum of the selected bonuses.

To bound this sum, we exploit the pooling of observations across inventory levels. For a fixed price p and threshold γ, observations from different inventory bins estimate the same CDF value and therefore increase a single shared count. The accumulated confidence radii then scale with the square root of the pooled observation count, rather than the sum of separate square-root costs across bins. This avoids the additional uncertainty cost incurred by the separate inventory-bin estimates in Algorithm 2. For each threshold, only $K+1$ pooled estimates remain, and the same confidence-radius summation gives a total contribution of $\mathcal{O}(\sqrt{KT \log(KJT)})$.

Moreover, every threshold contributing to the current bonus receives an observation on that round, so each contribution can be charged to an update of its shared count. Summing over the J thresholds introduces a factor of J, which is offset by their integration weight $\Delta = D/J$. Therefore, we obtain $\sum_{t=1}^{T} B_{t,m_t}(\gamma_t) = \mathcal{O}\left(P\Delta J\sqrt{KT \log(KJT)}\right) = \mathcal{O}\left(PD\sqrt{KT \log(KJT)}\right)$. The remaining dependence on J in this estimation bound comes only from the logarithmic confidence adjustment needed to cover all prices, thresholds, and rounds simultaneously. Thus, refining the threshold grid reduces the integration error without introducing an additional $\sqrt{J}$ estimation factor. Combining this bound with the selected widths, the price-grid error PDT/K, the integration error PDT/J, and the failure-event contribution PD yields the following guarantee.

Theorem 3.2. *Suppose Assumption 2.1 holds. Algorithm 1 with* $K = J = \lceil T^{1/3} \log^{-1/3} T \rceil$ *guarantees that* $\mathbb{E}[\mathrm{Reg}_T] = \mathcal{O}(PDT^{2/3} \log^{1/3} T)$.

The complete proof is deferred to Appendix B.2. Threshold-UCB therefore achieves $\widetilde{\mathcal{O}}(PDT^{2/3})$ expected regret by sharing observations across inventory levels. To complement this guarantee, we also establish an $\Omega(T^{2/3})$ lower bound with $P = D = 1$ through a reduction to classical stochastic posted pricing: constant inventory and binary demand recover the random-valuation model of Kleinberg & Leighton (2003). We defer the lower-bound argument to Appendix C. This reduction also explains why the parametric rate of Xu et al. (2026) does not extend to the present class. Their $\Omega(\sqrt{T})$ benchmark concerns a linear demand curve with shared additive noise, whereas

Algorithm 1: Threshold-UCB

Input: Price grid $\mathcal{P}_K = \{p_m = (m-1)P/K : m \in [K+1]\}$, maximum inventory D, integer $J \geq 2$, and horizon T

Set $\Delta = D/J$ and $u_j = j\Delta$ for $j \in \{0, 1, \ldots, J\}$;

For all $(m, j) \in [K+1] \times [J]$, initialize $N_{1,m,j} = Z_{1,m,j} = \widehat{F}_{1,m,j} = 0$;

Set $C^{\mathrm{th}}_{1,m,j} = \sqrt{\log(2(K+1)JT^2)/2}$ for all $(m, j) \in [K+1] \times [J]$;

for $t = 1, \ldots, T$ **do**

  Observe γ_t and choose $m_t \in \arg\max_{m \in [K+1]} p_m \Delta \sum_{j \in [J]: u_{j-1} < \gamma_t} \left(1 - \widehat{F}_{t,m,j} + C^{\mathrm{th}}_{t,m,j}\right)$;

  Decide $p_t = p_{m_t}$ and observe sales $X_t = \min\{Y_t(p_t), \gamma_t\}$.

  for $j \in [J]$ *such that* $u_{j-1} < \gamma_t$ **do**

    Set $N_{t+1,m_t,j} = N_{t,m_t,j} + 1$ and Set $Z_{t+1,m_t,j} = Z_{t,m_t,j} + \mathbf{1}\{X_t \leq u_{j-1}\}$;

    Set $\widehat{F}_{t+1,m_t,j} = Z_{t+1,m_t,j}/N_{t+1,m_t,j}$ and $C^{\mathrm{th}}_{t+1,m_t,j} = \sqrt{\frac{\log(2(K+1)JT^2)}{2N_{t+1,m_t,j}}}$;

  for $(m, j) \in [K+1] \times [J]$ *with* $m \neq m_t$ *or* $u_{j-1} \geq \gamma_t$ **do**

    Set $N_{t+1,m,j} = N_{t,m,j}$, $Z_{t+1,m,j} = Z_{t,m,j}$, $\widehat{F}_{t+1,m,j} = \widehat{F}_{t,m,j}$, and $C^{\mathrm{th}}_{t+1,m,j} = C^{\mathrm{th}}_{t,m,j}$;

in the model of Kleinberg & Leighton (2003), the randomness is generated by an arbitrary valuation distribution, so the resulting binary sales noise is price-dependent. Together, these bounds establish that Threshold-UCB is minimax-optimal up to logarithmic factors.

4 Experiments

In this section, we compare Threshold-UCB (Algorithm 1) with existing benchmarks across different demand, inventory, and noise specifications, examining both the setting covered by C20CB (Xu et al., 2026) and settings that depart from its assumptions. These parameter specifications are not tailored to any method. They are chosen to cover both C20CB's reference setting and departures from it, and our conclusions are consistent across all 48 resulting environments (see Appendix E and Appendix F).

Demand. We consider four baseline demand curves $\mu(p)$ for $p \in [0, 1]$: (A) the linear curve $1.4 - 0.8p$ with a pronounced negative slope; (B) the nearly flat linear curve $1.10 - 0.01p$; (C) the exponential curve $\exp(0.2 - p)$; and (D) the logistic curve $[1 + \exp(-(2 - 4p))]^{-1}$. Demand A is the reference case satisfying C20CB's linearity assumption, Demand B is linear but too flat for C20CB's inventory conditions to hold, and Demands C and D are nonlinear.

Inventory. We consider three support patterns: (1) central inventories in $[0.86, 1.29]$; (2) broad inventories in $[0.02, 1.75]$; and (3) piecewise-changing inventories that cycle among low, central, and high regimes. Each pattern uses either a uniform distribution (`U`) or a clipped-Gaussian distribution (`G`). The piecewise processes simulate abrupt changes in inventory availability, as permitted by our adversarial inventory model. The precise distributions are given in Appendix D.2.

Noise. The additive model (`Add`) uses $Y_t(p) = \mu(p) + \varepsilon_t$, where ε_t is zero-mean, price-independent uniform noise with half-width 0.10 for the linear curves, 0.05 for the exponential curve, and 0.03 for the logistic curve, so that demand is always nonnegative. The multiplicative model (`Mult`) uses $Y_t(p) = \mu(p)Z_t(p)$, where the multiplier has mean 1 but a price-dependent distribution, preserving the conditional mean while changing the noise law with price. The multiplicative distributions are specified in Appendix F.

Each figure title identifies the demand curve, inventory support, inventory distribution, and noise using these codes. For example, `A-1-U-Add` denotes Demand A with central uniform inventory and additive noise. To compare performance within and beyond C20CB's assumptions, we consider six representative combinations in Figure 1. Among them, only `A-1-U-Add` satisfies both its linear-additive demand and inventory-support conditions. Starting from this, `A-2-U-Add` expands the inventory support beyond the range allowed by C20CB. In `B-1-U-Add`, the demand curve is nearly flat relative to the noise, so no inventory level can produce the censoring pattern required by C20CB

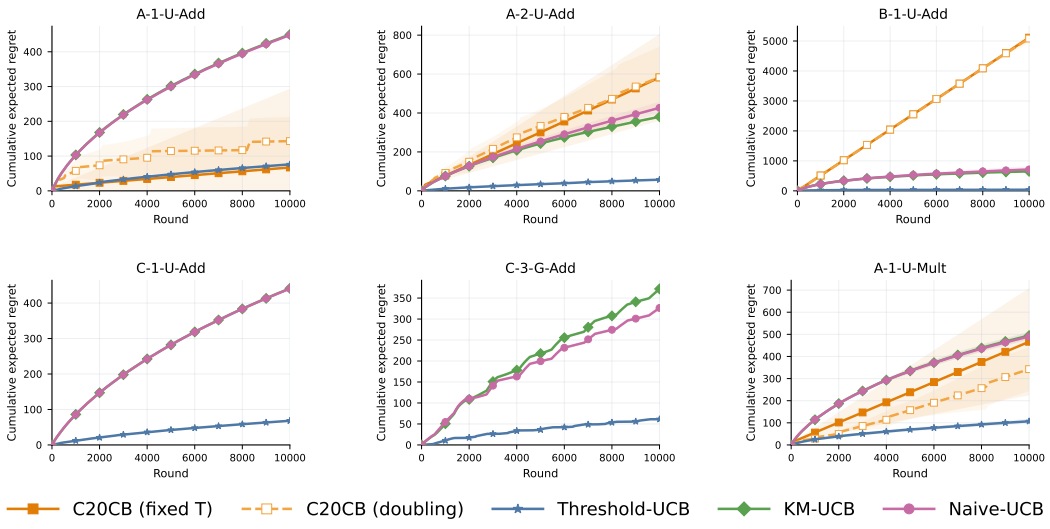


Figure 1: Mean cumulative expected regret over 30 held-out paths; shading indicates one sample standard deviation. Titles specify demand, inventory pattern, and noise type. Only `A-1-U-Add` matches C20CB's conditions. Broad inventory violates its support conditions, nearly flat Demand B makes the admissible inventory interval empty, and multiplicative noise violates its common additive-noise assumption. C20CB is omitted for nonlinear Demand C. Threshold-UCB achieves the lowest final regret in all panels except `A-1-U-Add`, where fixed-horizon C20CB is slightly lower.

(that is, demand must be fully censored at the lowest price and fully observed at the highest price). The nonlinear demand curves in `C-1-U-Add` and `C-3-G-Add` violate the linear-demand assumption, while `A-1-U-Mult` retains the reference linear demand curve but replaces the additive noise with multiplicative one. Detailed assumption checks are provided in Appendix D.2 and Appendix F.

Experimental setup. We compare tuned Threshold-UCB with the original C20CB (Xu et al., 2026) that uses a fixed horizon and a variant that uses a doubling trick on the horizon when the mean demand is linear, and with KM-UCB and Naive-UCB in all environments. KM-UCB uses Kaplan–Meier to estimate the demand distribution, whereas Naive-UCB treats censored sales as demand. We omit Double-Grid-UCB, which serves only as a theoretical warm-up in Section 3.1. See Appendix D.1 for detailed descriptions of all these benchmark algorithms. Each run lasts $T = 10000$ rounds, and regret is evaluated against an oracle on a 2001-point price grid, much finer than the learners' price grids, which have at most 45 points after tuning. Curves average 30 held-out paths. In each environment, each method receives 20 hyperparameter configurations evaluated on 5 separate tuning paths. Implementation details are given in Appendix D.

Results. Threshold-UCB achieves the lowest or nearly lowest mean final regret among the compared methods in all six environments in Figure 1. In the matched reference environment `A-1-U-Add`, fixed-horizon C20CB performs best, while Threshold-UCB attains comparable mean regret. Thus, our nonparametric method remains competitive when C20CB's stronger structural conditions hold.

Outside this reference setting, Threshold-UCB has the lowest final regret in each displayed environment. Under linear demand and additive noise, it outperforms all benchmarks in `A-2-U-Add` and `B-1-U-Add`, where C20CB's inventory conditions fail. Under nonlinear demand where C20CB does not apply, Threshold-UCB outperforms the heuristic KM-UCB and Naive-UCB baselines in both `C-1-U-Add` and `C-3-G-Add`, including the piecewise-inventory setting. Finally, in the case of `A-1-U-Mult` with price-dependent multiplicative noise, Threshold-UCB again achieves the lowest mean final regret, while both C20CB variants incur substantially higher regret. Additional results across the remaining demand, inventory, and noise combinations are reported in Appendix D, Appendix E, and Appendix F. Across all 48 environments, Threshold-UCB attains the lowest mean final regret in 46, the exceptions being `A-1-U-Add` and `A-1-G-Add`, where C20CB's conditions hold.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language editing, code development and debugging, and the presentation of mathematical arguments. We also used generative AI tools to assist with literature search and the preparation of experimental scripts and figures. All mathematical claims and proofs were independently checked by the authors, cited works were verified against the original sources, and AI-assisted code was reviewed and tested by the authors before being used in the reported experiments. We have reviewed all AI-assisted work and take responsibility for the final content of this work, including text, claims, code, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work is theoretical and uses only synthetic experiments. It does not involve human subjects, personal data, or proprietary datasets, and we are not aware of any specific ethical concerns raised by the work.

REPRODUCIBILITY STATEMENT

We provide the complete problem formulation and assumptions in Section 2, the proposed algorithms and their analyses in Section 3 and Appendix B, and the full lower-bound proof in Appendix C. Detailed experimental configurations, benchmark implementations, hyperparameter tuning, and evaluation protocols are given in Appendix D to Appendix F. The supplementary material contains the code used to run the experiments and generate the reported figures.

REFERENCES

Omar Besbes and Alp Muharremoglu. On implications of demand censoring in the newsvendor problem. *Management Science*, 59(6):1407–1424, 2013. doi: 10.1287/mnsc.1120.1654.

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. doi: 10.1287/opre.1080.0640.

Arnab Bisi, Maqbool Dada, and Surya Tokdar. A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing & Service Operations Management*, 13(4):525–533, 2011. doi: 10.1287/msom.1110.0340.

Denis Bitouzé, Béatrice Laurent, and Pascal Massart. A dvoretzky–kiefer–wolfowitz type inequality for the kaplan–meier estimator. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 35(6):735–763, 1999.

Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012. doi: 10.1287/opre.1120.1057.

Boxiao Chen, Xiuli Chao, and Yining Wang. Data-based dynamic pricing and inventory control with censored demand and limited price changes. *Operations Research*, 68(5):1445–1456, 2020. doi: 10.1287/opre.2020.1993.

Boxiao Chen, Xiuli Chao, and Cong Shi. Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Mathematics of Operations Research*, 46(2):726–756, 2021. doi: 10.1287/moor.2020.1084.

Boxiao Chen, Yining Wang, and Yuan Zhou. Optimal policies for dynamic pricing and inventory control with nonparametric censored demands. *Management Science*, 70(5):3362–3380, 2024. doi: 10.1287/mnsc.2023.4859.

Li Chen. Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Operations Research*, 58(2):396–413, 2010. doi: 10.1287/opre.1090.0726.

Guillermo Gallego and Garrett van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994. doi: 10.1287/mnsc.40.8.999.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Woonghee Tim Huh, Ganesh Janakiraman, John A. Muckstadt, and Paat Rusmevichientong. An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand. *Mathematics of Operations Research*, 34(2):397–416, 2009. doi: 10.1287/moor.1080.0367.

Woonghee Tim Huh, Retsef Levi, Paat Rusmevichientong, and James B. Orlin. Adaptive data-driven inventory control with censored demand based on kaplan–meier estimator. *Operations Research*, 59(4):929–941, 2011. doi: 10.1287/opre.1100.0906.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

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. doi: 10.1287/opre.2014.1294.

Robert D. Kleinberg and F. Thomson Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pp. 594–605, 2003. doi: 10.1109/SFCS.2003.1238232.

Yiyun Luo, Will Wei Sun, and Yufeng Liu. Contextual dynamic pricing with unknown noise: Explore-then-ucb strategy and improved regrets. *Advances in Neural Information Processing Systems*, 35:37445–37457, 2022.

Jianyu Xu and Yu-Xiang Wang. Logarithmic regret in feature-based dynamic pricing. *Advances in Neural Information Processing Systems*, 34:13898–13910, 2021.

Jianyu Xu and Yu-Xiang Wang. Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. In *International Conference on Artificial Intelligence and Statistics*, pp. 9643–9662. PMLR, 2022.

Jianyu Xu, Yining Wang, Xi Chen, and Yu-Xiang Wang. Dynamic pricing with adversarially-censored demands. *arXiv preprint arXiv:2502.06168v2*, 2026. doi: 10.48550/arXiv.2502.06168.

A Detailed Comparison with Xu et al. (2026)

In this section, we elaborate on the discussion in Section 2.1 and show that the conditions in Eq. (3) not only are restrictive but also greatly simplify learning: the known strict gap yields a high-price interval on which demand is never censored. Specifically, define $\epsilon \triangleq \min\left\{\frac{P}{2}, \frac{\gamma_{\min}-a_{\max}+b_{\min}P-c}{2b_{\max}}\right\} > 0$. Then, for every $q \in [P-\epsilon, P]$, we have $Y_t(q) \le a_{\max} - b_{\min}P + c + b_{\max}\epsilon < \gamma_{\min} \le \gamma_t$. In particular, P and $P-\epsilon$ are admissible prices at which sales reveal demand exactly for every admissible inventory sequence. Since $\mathbb{E}[N_t] = 0$, we know that $b = \frac{\mathbb{E}[Y_t(P-\epsilon)]-\mathbb{E}[Y_t(P)]}{\epsilon}$ and $a = \mathbb{E}[Y_t(P)]+bP$. Repeated observations at these two fixed prices therefore allow a and b to be estimated by sample means, without using censored observations. Moreover, for either price q, $Y_t(q) - \mathbb{E}[Y_t(q)] = N_t$; hence, centering the uncensored observations also allows us to estimate the common noise distribution, with the centering error controlled by the sample-mean error.

In addition, the same structure supports direct revenue estimation across prices and inventory levels. An uncensored observation collected at P satisfies $X_t = a - bP + N_t$. For any candidate $p \in [0, P]$ and inventory γ, we have the identity

$$p\min\{X_t + b(P-p), \gamma\} = p\min\{a - bp + N_t, \gamma\},$$

which shows that the transformed observation has expectation $R(p, \gamma)$. Replacing b by an estimate $\widehat{b}$ changes the transformed revenue by at most $p(P-p)|\widehat{b} - b|$. Combining this error bound with a confidence band for the uncensored demand distribution gives a direct revenue-UCB construction. Further combining uniform derivative estimates with the local curvature of the revenue function yields the $\widetilde{\mathcal{O}}(\sqrt{T})$ expected regret of C20CB. Thus, these support conditions make both parameter estimation and distribution learning accessible through uncensored exploration. Our model, by contrast, assumes only monotonicity of expected sales. It allows nonlinear, heteroskedastic, and price-dependent demand laws. The $\Omega(T^{2/3})$ lower bound in Appendix C establishes the statistical cost of this broader class, even when inventory is constant and censoring disappears.

B Omitted Details in Section 3

In this section, we provide the omitted details for Section 3. First, we prove the $\widetilde{\mathcal{O}}(T^{3/4})$ expected regret guarantee for Double-Grid-UCB presented in Section 3.1. We then prove the two lemmas used in the Threshold-UCB analysis and combine them to obtain its $\widetilde{\mathcal{O}}(T^{2/3})$ expected regret guarantee in Section 3.2.

B.1 Omitted Details in Section 3.1

In this section, we provide the omitted details for Section 3.1. We first give the complete pseudocode for Double-Grid-UCB in Algorithm 2 and then prove Theorem 3.1.

B.1.1 Algorithm Description of Double-Grid-UCB

Algorithm 2: Double-Grid-UCB

Input: Price bound P, inventory bound D, integers $K, J \geq 1$, and horizon T
Set $p_m \triangleq (m-1)P/K$ for $m \in [K+1]$, $\Delta \triangleq D/J$, and $u_j \triangleq j\Delta$ for $j \in \{0, \ldots, J\}$;
For all $(m, \ell) \in [K+1] \times [J]$, initialize $N^{\mathrm{dg}}_{1,m,\ell} = Q^{\mathrm{dg}}_{1,m,\ell} = \widehat{R}^{\mathrm{dg}}_{1,m,\ell} = 0$ and $C^{\mathrm{dg}}_{1,m,\ell} = PD\sqrt{\log(2(K+1)JT^2)/2}$;
for $t = 1, \ldots, T$ **do**
  Observe γ_t and set $\ell_t \triangleq \min\{J, 1 + \lfloor \gamma_t/\Delta \rfloor\}$;
  Choose $m_t \in \arg\max_{m \in [K+1]} \{\widehat{R}^{\mathrm{dg}}_{t,m,\ell_t} + C^{\mathrm{dg}}_{t,m,\ell_t}\}$;
  Post $p_t = p_{m_t}$ and observe X_t;
  Set $N^{\mathrm{dg}}_{t+1,m_t,\ell_t} = N^{\mathrm{dg}}_{t,m_t,\ell_t} + 1$ and $Q^{\mathrm{dg}}_{t+1,m_t,\ell_t} = Q^{\mathrm{dg}}_{t,m_t,\ell_t} + p_t \min\{X_t, u_{\ell_t - 1}\}$;
  Set $\widehat{R}^{\mathrm{dg}}_{t+1,m_t,\ell_t} = Q^{\mathrm{dg}}_{t+1,m_t,\ell_t}/N^{\mathrm{dg}}_{t+1,m_t,\ell_t}$ and $C^{\mathrm{dg}}_{t+1,m_t,\ell_t} = PD\sqrt{\frac{\log(2(K+1)JT^2)}{2N^{\mathrm{dg}}_{t+1,m_t,\ell_t}}}$;
  for $(m, \ell) \in [K+1] \times [J]$ *with* $(m, \ell) \neq (m_t, \ell_t)$ **do**
    set $N^{\mathrm{dg}}_{t+1,m,\ell} = N^{\mathrm{dg}}_{t,m,\ell}$, $Q^{\mathrm{dg}}_{t+1,m,\ell} = Q^{\mathrm{dg}}_{t,m,\ell}$, $\widehat{R}^{\mathrm{dg}}_{t+1,m,\ell} = \widehat{R}^{\mathrm{dg}}_{t,m,\ell}$, and $C^{\mathrm{dg}}_{t+1,m,\ell} = C^{\mathrm{dg}}_{t,m,\ell}$;

B.1.2 Proof of Theorem 3.1

Proof. By the reduction in Eq. (2) and the inventory-rounding bound in Eq. (7), we have

$$\begin{aligned}
\mathrm{Reg}_T &\leq \frac{PDT}{K} + \mathrm{Reg}_T^{\mathcal{P}_K} && \text{(Eq. (2))} \\
&\leq \frac{PDT}{K} + \frac{PDT}{J} + \sum_{t=1}^{T} \left[\max_{m \in [K+1]} R(p_m, u_{\ell_t - 1}) - R(p_t, u_{\ell_t - 1}) \right]. && (11)
\end{aligned}$$

It therefore suffices to bound regret at the lower inventory representatives. We use the confidence event from Eq. (9): denote by $\mathcal{E}^{\mathrm{dg}}$ the event that

$$\left| \widehat{R}^{\mathrm{dg}}_{t,m,\ell} - R(p_m, u_{\ell - 1}) \right| \leq C^{\mathrm{dg}}_{t,m,\ell} \tag{12}$$

holds for every $t \in [T+1]$, $m \in [K+1]$, and $\ell \in [J]$. Lemma B.1 shows that $\mathcal{E}^{\mathrm{dg}}$ holds with probability at least $1 - \frac{1}{T}$.

Suppose that $\mathcal{E}^{\mathrm{dg}}$ holds. Applying the definition of m_t, we know that

$$\begin{aligned}
&\max_{m \in [K+1]} R(p_m, u_{\ell_t - 1}) - R(p_t, u_{\ell_t - 1}) \\
&\leq \max_{m \in [K+1]} \mathsf{U}^{\mathrm{dg}}_t(m, u_{\ell_t - 1}) - R(p_t, u_{\ell_t - 1}) && \text{(as } \mathcal{E}^{\mathrm{dg}} \text{ holds)} \\
&\leq \mathsf{U}^{\mathrm{dg}}_t(m_t, u_{\ell_t - 1}) - R(p_t, u_{\ell_t - 1}) && \text{(by definition of } m_t) \\
&\leq 2C^{\mathrm{dg}}_{t,m_t,\ell_t}. && (13)
\end{aligned}$$

The last inequality follows from Eq. (12). Since $0 \leq \mathrm{Reg}_T \leq PDT$ on every sample path, we obtain

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] &= \mathbb{E}\left[\mathrm{Reg}_T \, \mathbf{1}\{\mathcal{E}^{\mathrm{dg}}\}\right] + \mathbb{E}\left[\mathrm{Reg}_T \, \mathbf{1}\{(\mathcal{E}^{\mathrm{dg}})^c\}\right] \\
&\leq \frac{PDT}{K} + \frac{PDT}{J} + 2\mathbb{E}\left[\sum_{t=1}^{T} C^{\mathrm{dg}}_{t,m_t,\ell_t}\right] + PD. && \text{(Lemma B.1)}
\end{aligned}$$

Thus, it remains to bound the third term. For each pair (m, ℓ), the counts before its successive visits are $0, \ldots, N^{\mathrm{dg}}_{T+1,m,\ell} - 1$. Also, $\sum_{r=0}^{n-1} (\max\{1, r\})^{-1/2} \leq 2\sqrt{n}$ for every integer $n \geq 1$, with an empty sum when $n = 0$. For notational convenience, let $L \triangleq \log(2(K+1)JT^2)$. Therefore, on every sample path, we know that

$$\sum_{t=1}^{T} C^{\mathrm{dg}}_{t,m_t,\ell_t} = PD\sqrt{\frac{L}{2}} \sum_{m=1}^{K+1} \sum_{\ell=1}^{J} \sum_{r=0}^{N^{\mathrm{dg}}_{T+1,m,\ell} - 1} \frac{1}{\sqrt{\max\{1, r\}}}$$

$$\leq 3PD\sqrt{L}\sum_{m=1}^{K+1}\sum_{\ell=1}^{J}\sqrt{N^{\mathrm{dg}}_{T+1,m,\ell}}$$

$$\leq 3PD\sqrt{L(K+1)J\sum_{m=1}^{K+1}\sum_{\ell=1}^{J}N^{\mathrm{dg}}_{T+1,m,\ell}} \qquad \text{(Cauchy-Schwarz inequality)}$$

$$= 3PD\sqrt{(K+1)JTL}. \tag{14}$$

The last equality uses $\sum_{m,\ell} N^{\mathrm{dg}}_{T+1,m,\ell} = T$, since each round updates exactly one price-bin pair. Substituting this bound gives

$$\mathbb{E}[\mathrm{Reg}_T] \leq \frac{PDT}{K} + \frac{PDT}{J} + 6PD\sqrt{(K+1)JTL} + PD. \tag{15}$$

Choosing $K = J = \lceil T^{1/4}\log^{-1/4} T\rceil$ for $T \geq 2$ yields $\mathbb{E}[\mathrm{Reg}_T] = \mathcal{O}(PDT^{3/4}\log^{1/4} T)$. □

Lemma B.1. *The sequence generated by Algorithm 2 satisfies* $\mathbb{P}(\mathcal{E}^{\mathrm{dg}}) \geq 1 - 1/T$.

Proof. Fix $(m,\ell) \in [K+1]\times[J]$, and let $\tau_1 < \tau_2 < \cdots$ denote the successive rounds at which $(m_t,\ell_t) = (m,\ell)$. On such a round, $u_{\ell-1} \leq \gamma_t$, and hence $p_t \min\{X_t, u_{\ell-1}\} = p_m \min\{Y_t(p_m), u_{\ell-1}\}$. Moreover, the event $\{(m_t,\ell_t) = (m,\ell)\}$ is determined before the current demand is realized. Since $Y_t(\cdot)$ is independent of $\mathcal{G}_t$ and has the same law in every period, the variables $W_r \triangleq p_m \min\{Y_{\tau_r}(p_m), u_{\ell-1}\}, r \geq 1$, are i.i.d., take values in $[0, PD]$, and have mean $R(p_m, u_{\ell-1})$.

For each fixed $n \in [T]$, Hoeffding's inequality (Lemma G.1) gives

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{r=1}^{n} W_r - R(p_m,u_{\ell-1})\right| > PD\sqrt{\frac{\log(2(K+1)JT^2)}{2n}}\right) \leq \frac{1}{(K+1)JT^2}.$$

At zero count, the confidence bound holds by initialization. Taking a union bound over all pairs (m,ℓ) and positive sample counts $n \in [T]$ yields

$$\mathbb{P}\big(\mathcal{E}^{\mathrm{dg}} \text{ does not hold}\big) \leq \sum_{m=1}^{K+1}\sum_{\ell=1}^{J}\sum_{n=1}^{T}\frac{1}{(K+1)JT^2} = \frac{1}{T}.$$

□

B.2 Omitted Details in Section 3.2

In this section, we provide the omitted details for Section 3.2. The pseudocode of Threshold-UCB is given in Algorithm 1. We first prove Theorem 3.2 using three supporting lemmas, and then prove these lemmas. Throughout, recall that $\Delta = D/J$, $u_j = j\Delta$ for $j \in \{0,\ldots,J\}$, and

$$\mathsf{U}^{\mathrm{th}}_t(m,\gamma) = p_m\Delta \sum_{j\in[J]:u_{j-1}<\gamma}\left(1 - \widehat{F}_{t,m,j}\right) + B_{t,m}(\gamma),$$

$$B_{t,m}(\gamma) = p_m\Delta \sum_{j\in[J]:u_{j-1}<\gamma} C^{\mathrm{th}}_{t,m,j},$$

with width $w_t(m,\gamma) = 2B_{t,m}(\gamma)$. For notational convenience, let $L \triangleq \log(2(K+1)JT^2)$.

B.2.1 Proof of Theorem 3.2

Proof. Unlike Double-Grid-UCB, Threshold-UCB constructs its indices for the discretized revenues $R^{(J)}_m$ in Eq. (10) instead of rounding the inventory. By the reduction in Eq. (2) and the integration-error bound in Eq. (21) of Lemma B.3, we have

$$\mathrm{Reg}_T \leq \frac{PDT}{K} + \mathrm{Reg}^{\mathcal{P}_K}_T \qquad \text{(Eq. (2))}$$

$$\leq \frac{PDT}{K} + \frac{PDT}{J} + \sum_{t=1}^{T} \left[\max_{m \in [K+1]} R_m^{(J)}(\gamma_t) - R_{m_t}^{(J)}(\gamma_t) \right]. \tag{16}$$

Here, only one integration error is paid in each round: every grid revenue satisfies $R(p_m, \gamma_t) \leq R_m^{(J)}(\gamma_t)$, while the selected revenue satisfies $R(p_t, \gamma_t) \geq R_{m_t}^{(J)}(\gamma_t) - P\Delta$, and $TP\Delta = PDT/J$. It therefore suffices to bound regret with respect to the discretized revenues. Denote by $\mathcal{E}^{\text{th}}$ the event that

$$\left| \widehat{F}_{t,m,j} - F_{p_m}(u_{j-1}) \right| \leq C_{t,m,j}^{\text{th}} \tag{17}$$

holds for every $t \in [T+1]$, $m \in [K+1]$, and $j \in [J]$. Lemma B.2 shows that $\mathcal{E}^{\text{th}}$ holds with probability at least $1 - \frac{1}{T}$.

Suppose that $\mathcal{E}^{\text{th}}$ holds. Applying the definition of m_t, we know that

$$\begin{aligned}
&\max_{m \in [K+1]} R_m^{(J)}(\gamma_t) - R_{m_t}^{(J)}(\gamma_t) \\
&\leq \max_{m \in [K+1]} \mathsf{U}_t^{\text{th}}(m, \gamma_t) - R_{m_t}^{(J)}(\gamma_t) && \text{(as } \mathcal{E}^{\text{th}} \text{ holds, by Lemma B.3)} \\
&\leq \mathsf{U}_t^{\text{th}}(m_t, \gamma_t) - R_{m_t}^{(J)}(\gamma_t) && \text{(by definition of } m_t) \\
&\leq w_t(m_t, \gamma_t), && (18)
\end{aligned}$$

where the last inequality follows from Eq. (20) in Lemma B.3. Since $0 \leq \text{Reg}_T \leq PDT$ on every sample path, we obtain

$$\begin{aligned}
\mathbb{E}[\text{Reg}_T] &= \mathbb{E}\left[\text{Reg}_T \, \mathbf{1}\{\mathcal{E}^{\text{th}}\}\right] + \mathbb{E}\left[\text{Reg}_T \, \mathbf{1}\{(\mathcal{E}^{\text{th}})^c\}\right] \\
&\leq \frac{PDT}{K} + \frac{PDT}{J} + \mathbb{E}\left[\sum_{t=1}^{T} w_t(m_t, \gamma_t)\right] + PD. && \text{(Lemma B.2)}
\end{aligned}$$

Thus, it remains to bound the third term. Since $w_t(m_t, \gamma_t) = 2B_{t,m_t}(\gamma_t)$, Lemma B.4 shows that, on every sample path,

$$\sum_{t=1}^{T} w_t(m_t, \gamma_t) = 2 \sum_{t=1}^{T} B_{t,m_t}(\gamma_t) \leq 6PD\sqrt{(K+1)TL}.$$

Substituting this bound gives

$$\mathbb{E}[\text{Reg}_T] \leq \frac{PDT}{K} + \frac{PDT}{J} + 6PD\sqrt{(K+1)TL} + PD. \tag{19}$$

Choosing $K = J = \lceil T^{1/3} \log^{-1/3} T \rceil$ for $T \geq 2$ yields $\mathbb{E}[\text{Reg}_T] = \mathcal{O}(PDT^{2/3} \log^{1/3} T)$. □

We now prove the three supporting lemmas. The first shows that the confidence event $\mathcal{E}^{\text{th}}$ holds with high probability.

Lemma B.2. *The sequence generated by Algorithm 1 satisfies* $\mathbb{P}(\mathcal{E}^{\text{th}}) \geq 1 - 1/T$.

Proof. Fix $(m, j) \in [K+1] \times [J]$. Whenever $m_t = m$ and $\gamma_t > u_{j-1}$, the observable indicator satisfies $\mathbf{1}\{X_t \leq u_{j-1}\} = \mathbf{1}\{Y_t(p_m) \leq u_{j-1}\}$. Moreover, the event $\{m_t = m, \gamma_t > u_{j-1}\}$ is determined before the current demand is realized. Since $Y_t(\cdot)$ is independent of $\mathcal{G}_t$ and has the same law in every period, the same argument as in the proof of Lemma B.1 shows that the retained indicators are consecutive entries $W_1, W_2, \ldots$ of an i.i.d. Bernoulli sequence with mean $F_{p_m}(u_{j-1})$.

For each fixed $n \in [T]$, Hoeffding's inequality (Lemma G.1) gives

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{r=1}^{n} W_r - F_{p_m}(u_{j-1}) \right| > \sqrt{\frac{L}{2n}} \right) \leq 2e^{-L} = \frac{1}{(K+1)JT^2}.$$

Whenever $N_{t,m,j} = n \geq 1$, the estimate $\widehat{F}_{t,m,j}$ is the average of the first n entries. At zero count, the error is at most one and the confidence bound holds by initialization, since $C_{1,m,j}^{\text{th}} = \sqrt{L/2} \geq 1$

when $K \geq 1$ and $J \geq 2$. Taking a union bound over all pairs (m, j) and positive sample counts $n \in [T]$ yields

$$\mathbb{P}\big(\mathcal{E}^{\mathrm{th}} \text{ does not hold}\big) \leq \sum_{m=1}^{K+1} \sum_{j=1}^{J} \sum_{n=1}^{T} \frac{1}{(K+1)JT^2} = \frac{1}{T}.$$

□

The next lemma shows that, on $\mathcal{E}^{\mathrm{th}}$, $\mathsf{U}_t^{\mathrm{th}}(m, \gamma)$ is a valid upper confidence bound on the discretized revenue $R_m^{(J)}(\gamma)$ with width at most $w_t(m, \gamma)$. It also bounds the error of approximating $R(p_m, \gamma)$ by $R_m^{(J)}(\gamma)$.

Lemma B.3. *On $\mathcal{E}^{\mathrm{th}}$, for every $t \in [T]$, $m \in [K+1]$, and $\gamma \in [0, D]$, we have*

$$0 \leq \mathsf{U}_t^{\mathrm{th}}(m, \gamma) - R_m^{(J)}(\gamma) \leq 2B_{t,m}(\gamma) = w_t(m, \gamma). \tag{20}$$

Moreover, for every $m \in [K+1]$ and $\gamma \in [0, D]$, we have

$$0 \leq R_m^{(J)}(\gamma) - R(p_m, \gamma) \leq p_m \Delta \leq P\Delta. \tag{21}$$

Proof. We first establish Eq. (20). Since $S_{p_m}(u_{j-1}) = 1 - F_{p_m}(u_{j-1})$, the definitions of $\mathsf{U}_t^{\mathrm{th}}$ and $R_m^{(J)}$ give

$$\mathsf{U}_t^{\mathrm{th}}(m, \gamma) - R_m^{(J)}(\gamma) = p_m \Delta \sum_{j \in [J]: u_{j-1} < \gamma} \left(F_{p_m}(u_{j-1}) - \widehat{F}_{t,m,j} + C_{t,m,j}^{\mathrm{th}} \right).$$

On $\mathcal{E}^{\mathrm{th}}$, each summand in parentheses lies in $[0, 2C_{t,m,j}^{\mathrm{th}}]$. Summing gives Eq. (20) simultaneously for every $\gamma \in [0, D]$, since changing inventory only selects a subset of the same threshold estimates.

For the integration error, if $\gamma = 0$, then both $R_m^{(J)}(\gamma)$ and $R(p_m, \gamma)$ equal zero. Otherwise, let $n \triangleq |\{j \in [J] : u_{j-1} < \gamma\}|$, so that $n \geq 1$ and $u_{n-1} < \gamma \leq u_n \leq D$. Since $S_{p_m}(\cdot)$ is nonincreasing, for every $j \in [n]$ we have $\int_{u_{j-1}}^{u_j} S_{p_m}(u)\,\mathrm{d}u \leq \Delta S_{p_m}(u_{j-1})$ and if $j \geq 2$, $\Delta S_{p_m}(u_{j-1}) \leq \int_{u_{j-2}}^{u_{j-1}} S_{p_m}(u)\,\mathrm{d}u$. Summing the first inequality over $j \in [n]$ and the second inequality over $2 \leq j \leq n$, and using $[0, u_n] \supseteq [0, \gamma]$ and $[0, u_{n-1}] \subseteq [0, \gamma]$, we have

$$\begin{aligned} R(p_m, \gamma) = p_m \int_0^{\gamma} S_{p_m}(u)\,\mathrm{d}u \leq R_m^{(J)}(\gamma) &\leq p_m \Delta S_{p_m}(u_0) + p_m \int_0^{u_{n-1}} S_{p_m}(u)\,\mathrm{d}u \\ &\leq p_m \Delta + R(p_m, \gamma), \end{aligned} \tag{22}$$

where the last inequality uses $S_{p_m}(u) \leq 1$ for any u and $u_{n-1} < \gamma$. This proves Eq. (21). □

The final lemma controls the cumulative bonus, and hence the cumulative selected width $\sum_{t=1}^{T} w_t(m_t, \gamma_t) = 2\sum_{t=1}^{T} B_{t,m_t}(\gamma_t)$. Each threshold estimate pools all observations at its price whose inventories exceed that threshold. Consequently, every radius appearing in the selected bonus is charged to an update of its shared count.

Lemma B.4. *On every sample path, the sequence generated by Algorithm 1 satisfies*

$$\sum_{t=1}^{T} B_{t,m_t}(\gamma_t) \leq 3PD\sqrt{(K+1)TL}. \tag{23}$$

Proof. For each pair (m, j), every selected bonus containing $C_{t,m,j}^{\mathrm{th}}$ also increments $N_{t,m,j}$ (Line 8 of Algorithm 1). The counts before these visits are therefore $0, \ldots, N_{T+1,m,j} - 1$. As in the proof of Theorem 3.1, we use $\sum_{r=0}^{n-1} (\max\{1, r\})^{-1/2} \leq 2\sqrt{n}$ for every integer $n \geq 1$, with an empty sum when $n = 0$. Therefore, on every sample path, we know that

$$\sum_{t=1}^{T} B_{t,m_t}(\gamma_t)$$

$$
\begin{aligned}
&= \Delta \sum_{j=1}^{J} \sum_{t:u_{j-1}<\gamma_t} p_{m_t} C^{\mathrm{th}}_{t,m_t,j} && \text{(definition of } B_{t,m_t}\text{)} \\
&\le P\Delta \sqrt{\frac{L}{2}} \sum_{j=1}^{J} \sum_{m=1}^{K+1} \sum_{r=0}^{N_{T+1,m,j}-1} \frac{1}{\sqrt{\max\{1,r\}}} && \text{(grouping visits by } (m,j)\text{)} \\
&\le 3P\Delta\sqrt{L} \sum_{j=1}^{J} \sum_{m=1}^{K+1} \sqrt{N_{T+1,m,j}} && \text{(summing the radii over visits)} \\
&\le 3P\Delta\sqrt{(K+1)L} \sum_{j=1}^{J} \sqrt{\sum_{m=1}^{K+1} N_{T+1,m,j}} && \text{(Cauchy–Schwarz over prices)} \\
&= 3P\Delta\sqrt{(K+1)L} \sum_{j=1}^{J} \sqrt{\sum_{t=1}^{T} \mathbf{1}\{u_{j-1} < \gamma_t\}} && \text{(one price update per eligible threshold)} \\
&= 3PD\sqrt{(K+1)TL},
\end{aligned}
$$

where the last equality uses $J\Delta = D$. Note that unlike the per-bin counts in Eq. (14), $N_{T+1,m,j}$ pools observations from all inventories exceeding u_{j-1}, so no separate count remains for an inventory bin. This proves Eq. (23). □

C Lower Bound

Section 3 gives $\widetilde{\mathcal{O}}(T^{3/4})$ expected regret for Double-Grid-UCB (Algorithm 2) and $\widetilde{\mathcal{O}}(T^{2/3})$ expected regret for Threshold-UCB (Algorithm 1). We now show that the latter exponent is optimal for the nonparametric model. The hard instances use constant inventory and binary demand, so censoring disappears. This restriction embeds classical stochastic posted pricing (Kleinberg & Leighton, 2003) in our model.

Theorem C.1. *There is a universal constant $c > 0$ such that, for all sufficiently large T, every dynamic-pricing policy admits an instance satisfying Section 2 and Assumption 2.1 such that $\mathbb{E}[\mathrm{Reg}_T] \ge cT^{2/3}$.*

Proof. It suffices to show that the classical unit-demand pricing problem studied by Kleinberg & Leighton (2003) is a special case of our model. Their lower-bound construction shows that, for every pricing policy and all sufficiently large T, there exists a fixed valuation distribution on $[0,1]$ under which expected regret relative to a price maximizing expected revenue is at least $cT^{2/3}$, for a universal constant $c > 0$.

In this classical problem, buyer valuations $V_1, \dots, V_T$ are drawn i.i.d. from an unknown distribution ν on $[0,1]$, independently of the learner's random seed. Denote by $\mathbb{P}_\nu$ and $\mathbb{E}_\nu$ probability and expectation, respectively, over the valuation sequence and the learner's randomization under valuation law ν. At each round, the learner posts a price $p_t \in [0,1]$ based on past observations, observes only the purchase indicator $\mathbf{1}\{V_t \ge p_t\}$, and earns revenue $p_t\mathbf{1}\{V_t \ge p_t\}$. For any fixed price p, the purchase probability is $\mathbb{P}_\nu(V_t \ge p) = \nu([p,1])$, which does not depend on t. Hence, the expected regret is

$$
T \max_{p\in[0,1]} p\mathbb{P}_\nu(V_t \ge p) - \mathbb{E}_\nu\left[\sum_{t=1}^{T} p_t \mathbf{1}\{V_t \ge p_t\}\right].
$$

To embed this problem in our model, set $P = D = 1$, choose $\gamma_t = 1$ in every round, and define $Y_t(p) \triangleq \mathbf{1}\{V_t \ge p\}$. The potential-demand functions are stationary and independent of the pre-demand information $\mathcal{G}_t$, as required by Section 2. Moreover, for every $\gamma \in [0,1]$, we have $s(p,\gamma) = \gamma\mathbb{P}_\nu(V_t \ge p)$, which is nonincreasing in p and therefore satisfies Assumption 2.1. Since demand is binary, $X_t = Y_t(p_t)$, so the observations and realized revenues coincide with those in the classical problem. Finally, constant inventory gives $R(p,1) = p\mathbb{P}_\nu(V_t \ge p)$, and conditioning on $\mathcal{G}_t$ shows that $\mathbb{E}_\nu[\mathrm{Reg}_T]$ equals the expected regret displayed above. The classical lower bound therefore applies to this instance, proving the theorem. □

D Experimental Details and Additional Additive Results

This section provides the experimental details referred to in Section 4. We first describe the benchmark algorithms, then specify the inventory processes and their relation to C20CB's assumptions, give the shared tuning and evaluation protocol, and finally present results for all 24 additive-noise environments. Appendix F gives the multiplicative-noise specifications and the corresponding additional results.

D.1 Benchmark Algorithms

Here we provide the omitted details of the benchmark algorithms used in our experiments. We first describe KM-UCB and Naive-UCB, which use different estimators on the same price grid, and then present the doubling version of C20CB.

KM-UCB and Naive-UCB use the uniform price grid $\mathcal{P}_K$ and a tunable exploration parameter $\alpha > 0$. Both select the price with the largest estimated revenue plus an exploration bonus, breaking ties in favor of the higher price. Their hyperparameters are selected using the tuning protocol described in Appendix D.3.

KM-UCB. KM-UCB uses the Kaplan–Meier estimator described in Section 3.2 to estimate demand at each price. As discussed there, adaptive price selection means that the retained observations need not satisfy the i.i.d. sampling assumptions required by the classical confidence guarantee. We therefore implement KM-UCB as a heuristic benchmark: it integrates the estimated survival function up to the current inventory and adds an exploration bonus, as summarized in Algorithm 3. We do not establish a regret guarantee for this benchmark under our adaptive sampling protocol.

Algorithm 3: KM-UCB

Input: Grid size K and exploration parameter $\alpha > 0$
Set $p_m = (m-1)P/K$ and initialize an empty observation list $\mathcal{H}_{1,m}$ for each $m \in [K+1]$;
for $t = 1, \dots, T$ **do**
  Observe γ_t;
  foreach $m \in [K+1]$ **do**
    Set $N_{t,m} = |\mathcal{H}_{t,m}|$ and let $\mathcal{Z}_{t,m}$ contain the distinct sales in $\mathcal{H}_{t,m}$;
    foreach $z \in \mathcal{Z}_{t,m}$ **do**
      Set $n_{t,m}(z) = \sum_{(x,e)\in\mathcal{H}_{t,m}} \mathbf{1}\{x \geq z\}$ and $d_{t,m}(z) = \sum_{(x,e)\in\mathcal{H}_{t,m}} e\,\mathbf{1}\{x = z\}$;
    Set $\widehat{S}^{\mathrm{km}}_{t,m}(u) = \prod_{z\in\mathcal{Z}_{t,m}:\, z\leq u}(1 - d_{t,m}(z)/n_{t,m}(z))$, with an empty product equal to one;
    Let $b_0, \dots, b_q$ be the sorted distinct points in $\{0, \gamma_t\} \cup \{z \in \mathcal{Z}_{t,m} : 0 < z < \gamma_t\}$ and set $\widehat{R}^{\mathrm{km}}_{t,m}(\gamma_t) = p_m \sum_{r=0}^{q-1}(b_{r+1} - b_r)\widehat{S}^{\mathrm{km}}_{t,m}(b_r)$, with an empty sum equal to zero;
    Set $\mathsf{U}^{\mathrm{km}}_t(m, \gamma_t) = \widehat{R}^{\mathrm{km}}_{t,m}(\gamma_t) + PD\sqrt{\alpha \log(\max\{2,T\})/(1+N_{t,m})}$;
  Choose the largest $m_t \in \arg\max_{m\in[K+1]} \mathsf{U}^{\mathrm{km}}_t(m, \gamma_t)$;
  Post $p_t = p_{m_t}$ and observe X_t;
  Set $\mathcal{H}_{t+1,m} = \mathcal{H}_{t,m}$ for all $m \in [K+1]$ and append $(X_t, \mathbf{1}\{X_t < \gamma_t\})$ to $\mathcal{H}_{t+1,m_t}$;

Naive-UCB. Naive-UCB uses the same price-selection rule but treats every sale as a fully observed demand realization. Specifically, it estimates revenue at the current inventory by averaging $p_m \min\{X_s, \gamma_t\}$ over previous sales at price p_m, as summarized in Algorithm 4. Thus, it ignores the information lost to earlier stockouts and can underestimate revenue at larger inventories.

Algorithm 4: Naive-UCB

Input: Grid size K and exploration parameter $\alpha > 0$
Set $p_m = (m-1)P/K$ and initialize an empty sales list $\mathcal{H}_{1,m}$ for each $m \in [K+1]$;
for $t = 1, \ldots, T$ **do**
  Observe γ_t;
  foreach $m \in [K+1]$ **do**
    Set $N_{t,m} = |\mathcal{H}_{t,m}|$;
    if $N_{t,m} = 0$ **then** Set $\widehat{R}^{\text{naive}}_{t,m}(\gamma_t) = p_m \gamma_t$;
    else Set $\widehat{R}^{\text{naive}}_{t,m}(\gamma_t) = \frac{p_m}{N_{t,m}} \sum_{x \in \mathcal{H}_{t,m}} \min\{x, \gamma_t\}$;
    Set $\mathsf{U}^{\text{naive}}_t(m, \gamma_t) = \widehat{R}^{\text{naive}}_{t,m}(\gamma_t) + PD\sqrt{\alpha \log(\max\{2,T\})/(1+N_{t,m})}$;
  Choose the largest $m_t \in \arg\max_{m \in [K+1]} \mathsf{U}^{\text{naive}}_t(m, \gamma_t)$;
  Post $p_t = p_{m_t}$ and observe X_t;
  Set $\mathcal{H}_{t+1,m} = \mathcal{H}_{t,m}$ for all $m \in [K+1]$ and append X_t to $\mathcal{H}_{t+1,m_t}$;

C20CB-doubling. The fixed-horizon benchmark runs C20CB (Xu et al., 2026) with horizon T. Its doubling version instead starts a fresh C20CB instance on each epoch of length $1, 2, 4, \ldots$, so its online rule does not require the terminal horizon. As shown in Algorithm 5, each instance uses its epoch length to set the exploration length and noise-grid resolution. Only the learner is restarted and the inventory and demand process continues unchanged. If the experiment ends during an epoch, that epoch is truncated.

Algorithm 5: C20CB-doubling

Input: Exploration multiplier $s_\tau > 0$, grid multiplier $s_M > 0$, and confidence multiplier $c_{\text{conf}} > 0$
Initialize epoch length $H = 1$ and completed rounds within the epoch $h = 0$;
for $t = 1, 2, \ldots$ **do**
  if $h = H$ **then** Set $H \leftarrow 2H$ and $h = 0$;
  if $h = 0$ **then**
    Set $\tau = \min\{H, \max\{1, \lceil s_\tau \sqrt{H} \rceil\}\}$ and $M = \max\{1, \lceil s_M H^{1/4} \rceil\}$;
    Initialize a fresh instance $\mathcal{A} = \text{C20CB}(H, \tau, M, c_{\text{conf}})$;
  Observe γ_t and obtain $p_t = \mathcal{A}.\textsc{SelectPrice}(\gamma_t)$, post p_t, and observe X_t;
  Call $\mathcal{A}.\textsc{Update}(\gamma_t, p_t, X_t)$;
  Set $h \leftarrow h + 1$;

D.2 Inventory configurations and assumption checks

In this section, we provide the complete inventory configurations used in Section 4 and verify which demand-inventory combinations satisfy the conditions of C20CB. The demand curves and additive-noise distributions are specified in Section 4. Table 1 completes the inventory specifications, using the same support and distribution codes as the figures. For the piecewise processes, we use $\lceil T^{1/3} \rceil$ blocks of length $\lceil T/\lceil T^{1/3} \rceil \rceil$, truncating the last block at T. The regimes cycle from low to central to high, and inventories are sampled independently within each block. All inventory paths are generated before the interaction starts.

Table 1: Inventory processes used in both noise suites. The codes combine support pattern with distribution: `U` denotes uniform and `G` denotes clipped Gaussian. Piecewise processes cycle through three regimes.

Code	Type	Exact configuration
`1-U`	Uniform	$\gamma_t \sim \mathrm{Unif}[0.86, 1.29]$.
`2-U`	Uniform	$\gamma_t \sim \mathrm{Unif}[0.02, 1.75]$.
`1-G`	Clipped Gaussian	$\gamma_t = \mathrm{clip}\{N(1.075, 0.14^2), 0.86, 1.29\}$.
`2-G`	Clipped Gaussian	$\gamma_t = \mathrm{clip}\{N(0.85, 0.55^2), 0.02, 1.75\}$.
`3-U`	Piecewise uniform	Cycle $[0.02, 0.30]$, $[0.86, 1.29]$, and $[1.45, 1.75]$.
`3-G`	Piecewise Gaussian	Cycle clipped $N(0.12, 0.05^2)$, $N(1.075, 0.14^2)$, and $N(1.60, 0.06^2)$ on the same three supports.

We next examine more closely which combinations satisfy or violate the conditions of C20CB in Xu et al. (2026). Both additive linear models use noise $\varepsilon_t \sim \mathrm{Unif}[-0.10, 0.10]$, so $c = 0.10$ in the context of Xu et al. (2026) is the tight noise bound. For Demand A, the true parameters are $(a, b) = (1.4, 0.8)$, and the experiments supply $(a_{\max}, b_{\min}, b_{\max}, c) = (1.5, 0.75, 0.85, 0.10)$. These bounds require $\gamma_t \in (0.85, 1.30)$ and $\gamma_t > 0.20$, which both the central uniform and central clipped-Gaussian inventory processes satisfy. Even the tightest valid bounds only widen the interval to $(0.70, 1.30)$ and the broad and piecewise inventory supports extend outside this interval, so their violations are not caused by loose parameter bounds.

For Demand B, changing the price over $[0, 1]$ changes mean demand by only 0.01, less than the noise support width of 0.20. Thus, no inventory level can lie below all demand realizations at the lowest price and above all demand realizations at the highest price, as required by C20CB. Specifically, any valid bounds satisfy $a_{\max} \geq a$, $b_{\min} \leq b$, and $c \geq 0.10$. With $(a, b) = (1.10, 0.01)$ and $P = 1$, C20CB requires both $\gamma_t > a_{\max} - b_{\min} P + c \geq 1.19$ and $\gamma_t < a - c \leq 1$. These requirements are incompatible for every valid choice of bounds. C20CB is not tested on Demands C and D because their conditional means are nonlinear.

D.3 Tuning and evaluation protocol

All experiments use prices in $[0, P]$ with $P = 1$ and horizon $T = 10{,}000$. We use the same protocol for every environment, that is, for every combination of demand family, inventory process, and noise model in Section 4.

Hyperparameter tuning. In each environment, every method is tuned separately with the same budget of 20 hyperparameter configurations. The first eight configurations include the reference configuration in Table 2; the remaining twelve are proposed sequentially by Gaussian-process Bayesian optimization based on the results so far. Each configuration is run on the same five tuning paths, and we select the configuration with the smallest mean cumulative regret at round T over these paths. The selected configuration is then evaluated on 30 held-out paths that are not used for tuning. Thus, every method receives the same number of tuning runs and evaluation runs.

Table 2 lists the tuned hyperparameters for the additive-noise experiments. For Threshold-UCB, K and J are the numbers of price-grid intervals and thresholds in Algorithm 1, and c_{conf} scales the confidence radii $C^{\mathrm{th}}_{t,m,j}$. For KM-UCB and Naive-UCB (Algorithm 3 and Algorithm 4), K is the price-grid size and α scales the exploration bonus. For C20CB and C20CB-doubling (Algorithm 5), s_τ scales the length τ of the initial exploration phase, s_M scales the size M of the noise grid, and c_{conf} scales the confidence widths. The reference value of K and J is $\lceil T^{1/3} \rceil = 22$, and the reference values of the other hyperparameters are given in parentheses in Table 2. Each search range spans from half to twice the reference value and is searched on a logarithmic scale; K and J are rounded to the nearest integer, with ties rounded up.

Table 2: Hyperparameters tuned in the additive-noise experiments with $T = 10{,}000$. Reference values, which are included among the initial configurations, are given in parentheses.

Method	Hyperparameters	Search ranges (reference values)
Threshold-UCB	K, J, c_{conf}	$K, J \in \{11, \dots, 44\}$ (22 each); $c_{\text{conf}} \in [0.5, 2]$ (1).
C20CB	$s_\tau, s_M, c_{\text{conf}}$	$s_\tau, s_M, c_{\text{conf}} \in [0.5, 2]$ (1 each), with $\tau = \lceil s_\tau \sqrt{T} \rceil$ and $M = \lceil s_M T^{1/4} \rceil$.
C20CB-doubling	$s_\tau, s_M, c_{\text{conf}}$	Same as C20CB, with T replaced by the epoch length $H \in \{1, 2, 4, \dots\}$ and τ capped at H.
KM-UCB	K, α	$K \in \{11, \dots, 44\}$ (22); $\alpha \in [1, 4]$ (2).
Naive-UCB	K, α	Same as KM-UCB.

Fixed implementation choices. The remaining settings are fixed across environments. Some of them differ from the choices used in our analysis; we list them explicitly.

- *Threshold-UCB.* The threshold grid uses $D = 1.8$, which upper-bounds every inventory level in Table 1, so $\Delta = 1.8/J$. The index omits the term $j = 1$, that is, the threshold $u_0 = 0$, and sums only over the positive thresholds $u_1, \dots, u_{J-1}$ below γ_t. With $\delta = 0.05$, the confidence radius is

$$C^{\text{th}}_{t,m,j} = c_{\text{conf}} \sqrt{\frac{\log(2(K+1)JT/\delta)}{2N_{t,m,j}}},$$

  if $N_{t,m,j} > 0$, and $C^{\text{th}}_{t,m,j} = c_{\text{conf}}$ if $N_{t,m,j} = 0$ in place of the radius in Algorithm 1. Ties in the index are broken in favor of the higher price, as in KM-UCB and Naive-UCB.
- *C20CB and C20CB-doubling.* The original C20CB analysis involves several problem-dependent constants controlling the uncertainty from the initial parameter-estimation phase. To keep the benchmark implementation tractable and avoid tuning a large number of C20CB-specific constants, we tie these first-stage uncertainty corrections to a single fixed coefficient, set to 0.25, for the corresponding $1/\sqrt{\tau}$ terms. The resulting confidence widths are further scaled by the tuned multiplier c_{conf}. Both variants also clip the parameter estimates to the supplied bounds $(a_{\max}, b_{\min}, b_{\max}, c)$ in Appendix D.2, and use the fallback price 0.5 if the checks after the initial exploration phase fail.

The modifications for multiplicative noise are described in Appendix F.

Evaluation. Within each tuning or held-out path, all methods face the same inventory sequence $\gamma_1, \dots, \gamma_T$ and the same underlying noise draws, so differences in performance come only from the posted prices. The resulting demands can still differ because they depend on the price. Tuning and held-out paths use separate random seeds. We report the cumulative expected regret

$$\sum_{t=1}^{T} \left[\max_{p \in \mathcal{P}_{2000}} R(p, \gamma_t) - R(p_t, \gamma_t) \right],$$

where $\mathcal{P}_{2000}$ is the uniform grid of 2001 prices in $[0, 1]$. This fine grid approximates the continuous oracle in Eq. (1) and is independent of the learner's price grid. The expected revenue $R(p, \gamma)$ is computed in closed form whenever available. For additive Demand D, the revenue is instead approximated numerically using a fixed 41-point quadrature rule. Each curve shows the mean over the 30 held-out paths, and the shaded band shows the mean plus or minus one standard deviation, with its lower edge clipped at zero.

E Additional additive results

In this section, we provide the complete results for the additive-noise experiments. Figure 2–Figure 5 show one demand family each, with the six inventory processes arranged in the same 2×3 layout: columns vary the inventory support and rows vary the inventory distribution.

These figures complete the additive comparisons in Figure 1. Threshold-UCB has the smallest mean final regret in 22 of the 24 environments. Fixed-horizon C20CB has the smallest mean in the

two central-inventory environments for Demand A, where its linear-additive and inventory-support conditions hold. For all other environments, the figures show the effect of relaxing those conditions.

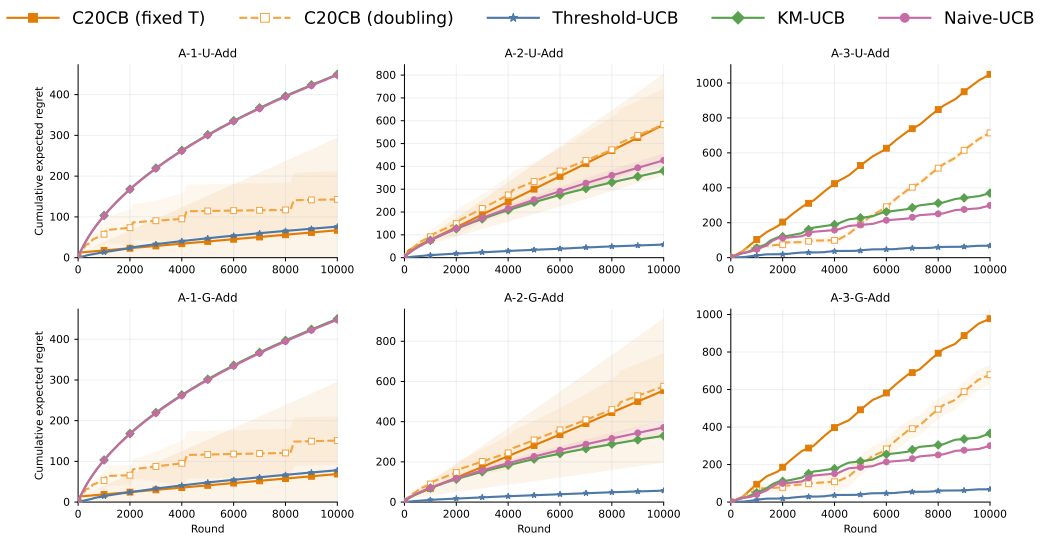


Figure 2: Additive-noise Demand A. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively. Only `A-1-U-Add` and `A-1-G-Add` satisfy C20CB's inventory conditions; the broad and piecewise processes test performance when those conditions fail.

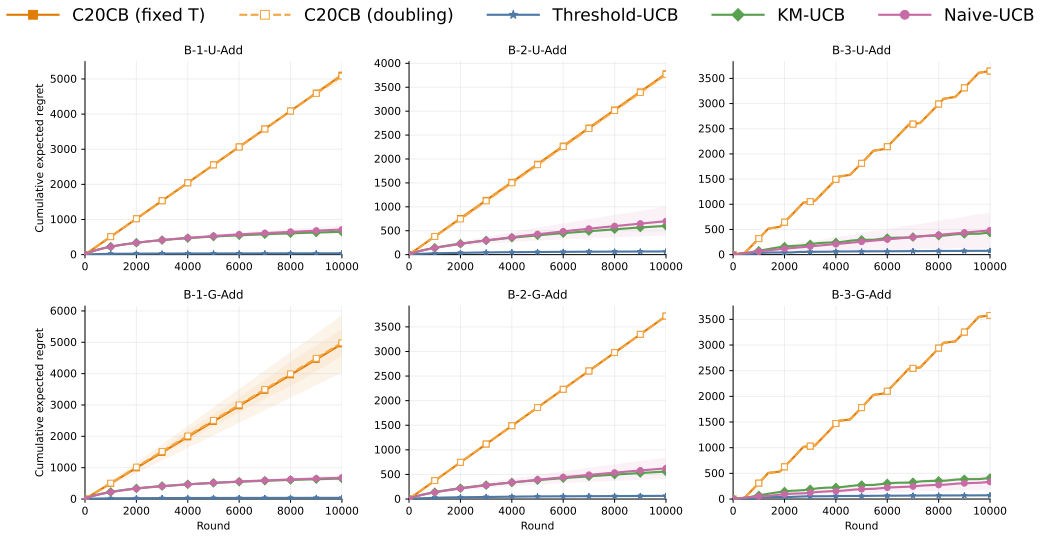


Figure 3: Additive-noise Demand B. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively.

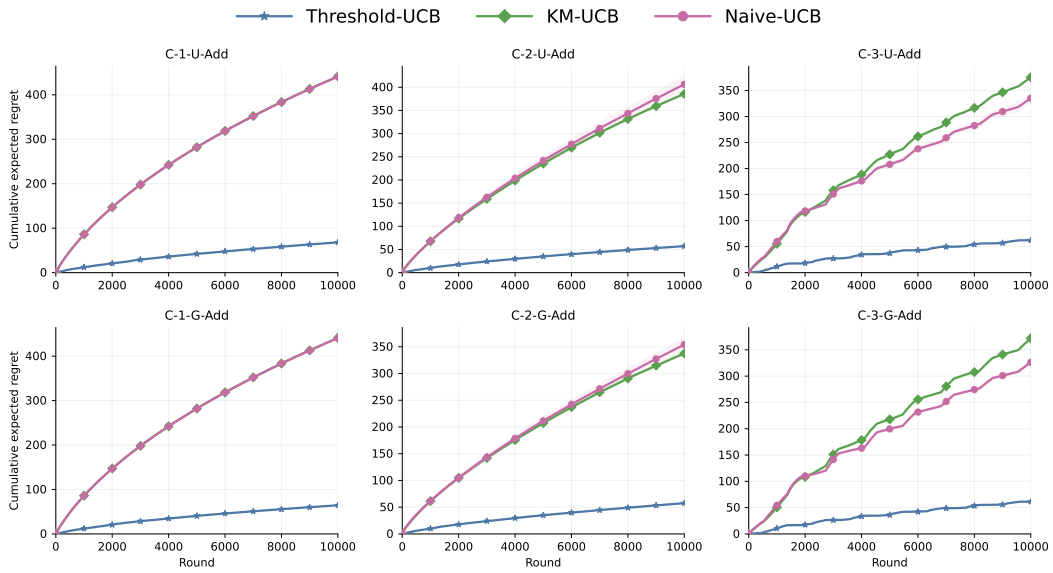


Figure 4: Additive-noise Demand C. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively. C20CB is omitted because the mean demand is nonlinear.

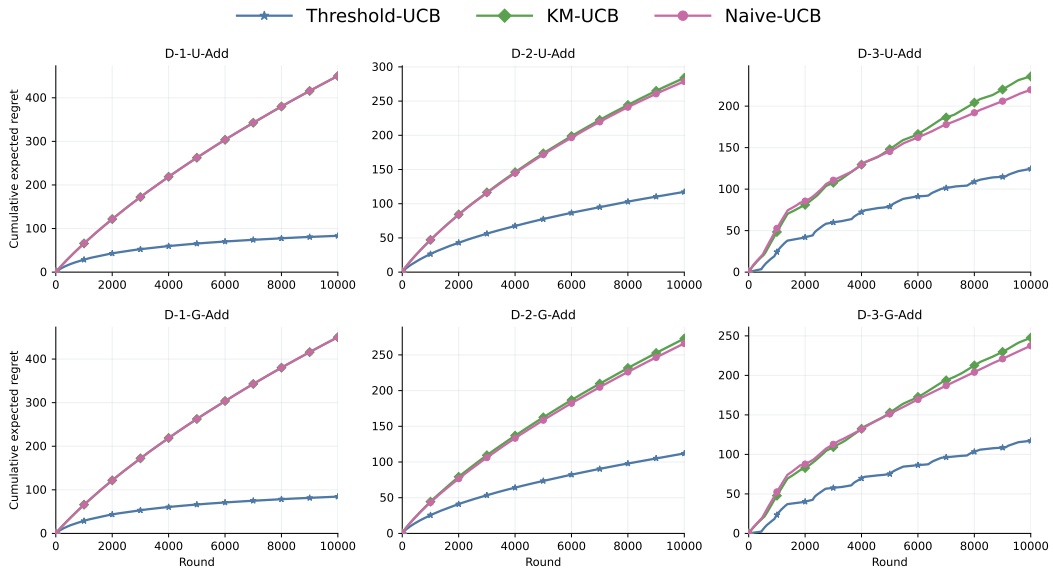


Figure 5: Additive-noise Demand D. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively. C20CB is omitted because the mean demand is nonlinear.

F Additional Results with Multiplicative Noise

In this section, we provide the details and complete results for the multiplicative-noise experiments (`Mult`) in Section 4. We use the same four demand curves $\mu(p)$ and the same six inventory processes in Table 1, so there are again 24 environments. The only change is the noise model: unlike the additive case, the shape of the demand distribution now changes with price, even after centering at its

mean. We first specify the noise model and check which assumptions it satisfies, then describe how the protocol in Appendix D.3 is adapted, and finally report the results.

F.1 Noise specifications and assumption checks

For each demand family, potential demand is $Y_t(p) = \mu(p)Z_t(p)$, where the multiplier $Z_t(p)$ takes two values:

$$Z_t(p) = \begin{cases} \ell, & \text{with probability } q(p), \\ h(p) \triangleq \dfrac{1-\ell q(p)}{1-q(p)}, & \text{with probability } 1-q(p). \end{cases}$$

Here, $\ell < 1$ is a fixed low multiplier and $q(p) < 1$ is the price-dependent probability of the low state. The high multiplier $h(p) \geq 1$ is chosen so that $\mathbb{E}[Z_t(p)] = 1$. Hence, expected potential demand equals $\mu(p)$ at every price, exactly as in the additive case, and only the shape of the distribution changes. Table 3 specifies ℓ and $q(p)$ for each family, where, for Demand B, $x(p) \triangleq \mathrm{clip}((p-0.66)/0.18, 0, 1)$, so that $q(p)$ increases smoothly from 0 to 0.95 as p increases from 0.66 to 0.84. For Demands A and C, $q(p) = 0$ for $p \leq 0.65$, so demand is deterministic at low prices and becomes random only at high prices.

Table 3: Multiplier distributions for multiplicative noise. The last column gives the demand bound $D_{\mathrm{exp}} \geq \max_{p\in[0,1]} \mu(p)h(p)$, which upper-bounds every realization of potential demand.

Demand	Low multiplier ℓ	Low-state probability $q(p)$	Demand bound D_{exp}
A	0.20	$2(p-0.65)_+$	1.8
B	0.63	$0.95\big(3x(p)^2 - 2x(p)^3\big)$	9
C	0.20	$2(p-0.65)_+$	1.35
D	0.20	$\min\{0.85, 1.4p\}$	2.25

Monotone expected sales. All four environments have expected sales that are nonincreasing in price, satisfying Assumption 2.1. Fix $p \leq p'$ and $\gamma \geq 0$. Since $\mu(p)$ is nonincreasing and $q(p)$ is nondecreasing, we have $\mu(p) \geq \mu(p')$ and $q(p) \leq q(p')$. Recall that $Z_t(p)$ equals ℓ with probability $q(p)$ and $h(p)$ otherwise, where $h(p)$ is chosen so that $\mathbb{E}[Z_t(p)] = 1$. Hence, we have

$$(1-q(p))h(p) = 1 - q(p)\ell = (q(p') - q(p))\ell + (1-q(p'))h(p').$$

Since $z \mapsto \min\{\mu(p')z, \gamma\}$ is concave, this identity implies

$$(1-q(p))\min\{\mu(p')h(p), \gamma\} \geq (q(p')-q(p))\min\{\mu(p')\ell, \gamma\} + (1-q(p'))\min\{\mu(p')h(p'), \gamma\}.$$

Adding $q(p)\min\{\mu(p')\ell, \gamma\}$ to both sides gives $\mathbb{E}[\min\{\mu(p')Z_t(p), \gamma\}] \geq \mathbb{E}[\min\{\mu(p')Z_t(p'), \gamma\}]$. Finally, since $\mu(p) \geq \mu(p')$ and $Z_t(p) \geq 0$, we have

$$\mathbb{E}[\min\{\mu(p)Z_t(p), \gamma\}] \geq \mathbb{E}[\min\{\mu(p')Z_t(p), \gamma\}].$$

Therefore, $s(p, \gamma) \geq s(p', \gamma)$.

F.2 Changes to the tuning and evaluation protocol

We use the tuning and evaluation protocol in Appendix D.3, including the budget of 20 configurations on five tuning paths and the 30 held-out paths, with the following changes.

- *Demand bound.* Potential demand can now exceed the largest inventory level, and its range differs across families. We therefore replace $D = 1.8$ by the family-specific bound D_{exp} in Table 3, both in Threshold-UCB's threshold grid and in the exploration bonuses $PD\sqrt{\alpha \log(\max\{2, T\})/(1+N_{t,m})}$ of KM-UCB and Naive-UCB. Since $Y_t(p) \leq D_{\mathrm{exp}}$, we have $S_p(u) = 0$ for every $u \geq D_{\mathrm{exp}}$, so thresholds in $[0, D_{\mathrm{exp}}]$ suffice even when inventory exceeds D_{exp}.
- *Threshold count of Threshold-UCB.* To keep the threshold spacing $\Delta = D_{\mathrm{exp}}/J$ comparable across families, the reference value of J is $J_0 = \lceil D_{\mathrm{exp}} T^{1/3} \rceil$, and J is tuned over the integers in $[\max\{2, \lfloor 0.5J_0 \rfloor\}, \lceil 2J_0 \rceil]$. The ranges of K and c_{conf} are unchanged.

- *C20CB and C20CB-doubling.* Since no true additive-noise bound exists, both variants additionally tune the noise bound $c \in [0.025, 0.4]$ that they use internally, on a logarithmic scale with reference value 0.1, and the fallback price in $[0.1, 0.9]$, on a linear scale with reference value 0.5. This tuning gives C20CB more flexibility but does not restore its common additive-noise assumption.

All other hyperparameter ranges and fixed settings are as in Table 2 and Appendix D.3. For evaluation, the expected revenue has the closed form

$$R(p, \gamma) = p\Big[q(p)\min\{\ell\mu(p), \gamma\} + \big(1 - q(p)\big)\min\{h(p)\mu(p), \gamma\}\Big],$$

and regret is computed against the same 2001-point oracle grid $\mathcal{P}_{2000}$ as in the additive case.

F.3 Additional multiplicative results

In this section, we report the complete results for the multiplicative-noise experiments. Figure 6–Figure 9 each show one demand family, with the six inventory processes arranged in the same 2×3 layout as in Appendix E. Threshold-UCB still has the smallest mean final regret in all 24 environments. For Demand A, the mean is the same linear curve as in the additive reference and only the noise distribution changes, so the gap between Threshold-UCB and C20CB cannot be attributed to a nonlinear mean. For the nonlinear Demands C and D, we compare only Threshold-UCB, KM-UCB, and Naive-UCB.

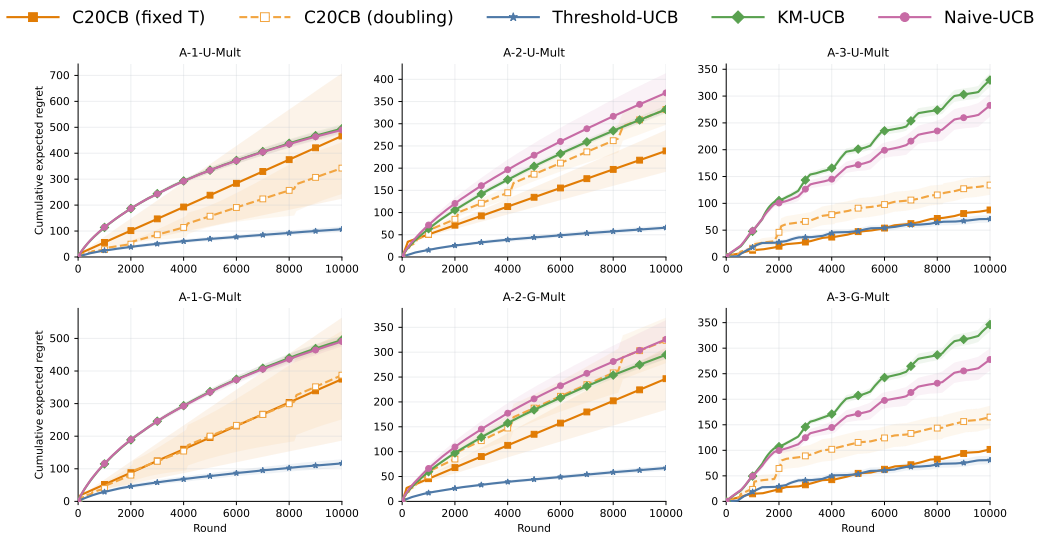


Figure 6: Multiplicative-noise Demand A. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively. The mean matches additive Demand A, but the price-dependent noise violates C20CB's common additive-noise assumption.

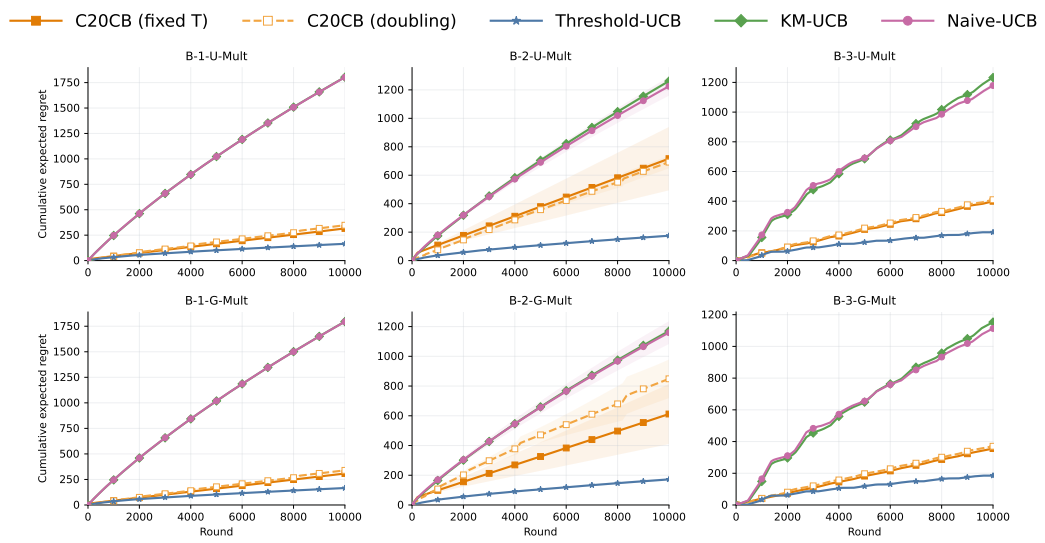


Figure 7: Multiplicative-noise Demand B. Columns correspond to the central (1), broad (2), and piecewise-changing (3) inventory supports; the top and bottom rows use uniform (U) and clipped-Gaussian (G) inventory distributions, respectively. The nearly flat linear mean is combined with a price-dependent probability of low demand, which violates C20CB's common additive-noise assumption.

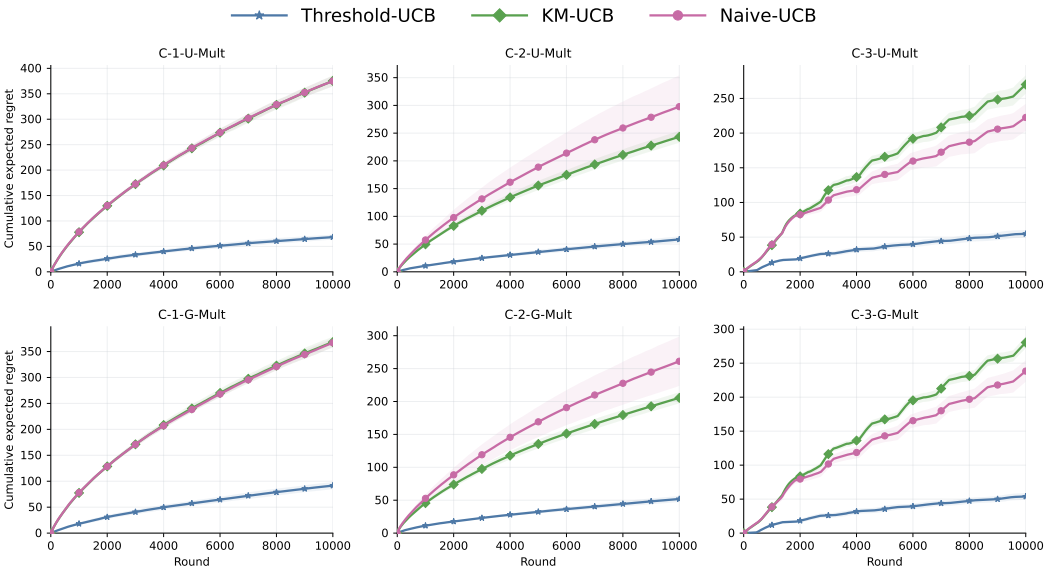


Figure 8: Multiplicative-noise Demand C. Columns correspond to the central (1), broad (2), and piecewise-changing (3) inventory supports; the top and bottom rows use uniform (U) and clipped-Gaussian (G) inventory distributions, respectively. C20CB is omitted because the mean demand is nonlinear.

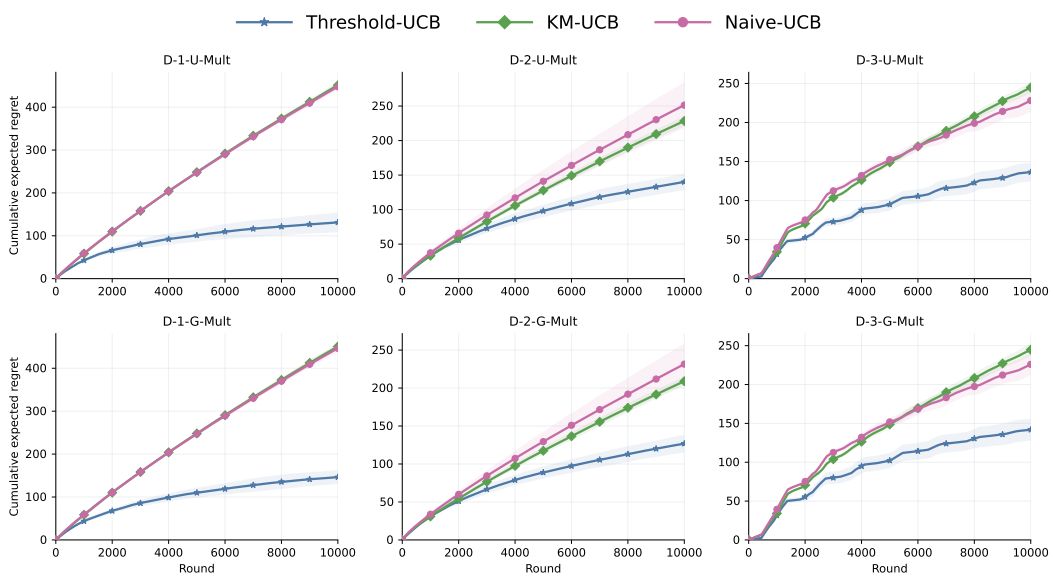


Figure 9: Multiplicative-noise Demand D. Columns correspond to the central (`1`), broad (`2`), and piecewise-changing (`3`) inventory supports; the top and bottom rows use uniform (`U`) and clipped-Gaussian (`G`) inventory distributions, respectively. C20CB is omitted because the mean demand is nonlinear.

G Auxiliary Lemmas

Lemma G.1 (Hoeffding's inequality; Hoeffding, 1963)**.** *Let $W_1, \ldots, W_n$ be independent random variables taking values in $[a, b]$, where $a < b$, with common mean μ. Then, for every $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{r=1}^{n} W_r - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$