# SAFE CONTEXTUAL PRICING UNDER CENSORED FEEDBACK: JOINT INVENTORY AND ALLOCATION WITH ONLINE CONFORMAL GUARANTEES

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Many online decision problems are simultaneously profit-seeking and safety-constrained: an agent must minimise regret while keeping a downstream operational risk below a prescribed level, using only censored feedback. We study this tension in online contextual pricing with multi-supplier inventory and price-dependent allocation, where the operational risk is the stockout rate. The seller observes only whether each customer purchases, must commit inventory across heterogeneous suppliers before demand realises, and faces an unknown and possibly non-Lipschitz demand-noise distribution. Our algorithm CP-CPA prices by maximising an optimistic profit on a residual grid, provisions the estimated profit-optimal inventory, and raises it to an online conformally calibrated demand quantile whenever the coverage target requires more. Its regret against the unconstrained clairvoyant is $\widetilde{O}(T^{2/3})$, the optimal rate under agnostic noise, plus exactly the holding cost of the safety inventory added by the conformal layer; this cost is $\widetilde{O}(T^{2/3})$ when the coverage target is slack, and when it binds we prove that linear regret is unavoidable for every coverage-feasible algorithm. The realised stockout rate satisfies $\widehat{\alpha}_T \leq \alpha + O(1/(\eta T))$ deterministically, for any noise distribution and regardless of the accuracy of the demand model, and one conformal state per segment of a fixed context partition yields the same guarantee on every segment. The analysis replaces the Lipschitz and piecewise-convexity structure of prior joint pricing and allocation arguments, which fails under atomic noise, with a coupling bound on poisson-binomial tails and a monotone-demand comparator. On synthetic benchmarks with Lipschitz and atomic noise, the fitted regret exponents match the $T^{2/3}$ prediction and lie far below the distribution-free baseline; when the target binds, CP-CPA attains a stockout rate of $0.096$ with zero variance across seeds while the plug-in quantile over-provisions at higher regret. On a demand curve calibrated from the public UCI Online Retail II log, CP-CPA is the only method that meets the target, at a profit cost of $0.1\%$, and the per-segment variant never exceeds the target in any segment.

## 1 INTRODUCTION

A recurring pattern in online learning couples a *profit* objective with a *safety* objective under partial feedback: an agent repeatedly acts to minimise regret, the same action commits it to a downstream operational risk that must stay below a prescribed level, and only censored feedback is observed. This paper studies a general recipe for this pattern, wrapping a censored-feedback contextual bandit with an online conformal calibration layer whose target is the operational risk, and shows that the safety layer can be added at no asymptotic regret cost. We develop the recipe on the sharpest instance we know, in which the bandit decision *is* the safety decision: online pricing with inventory commitment. A seller posts a price, provisions inventory across several suppliers before demand realises, allocates it to heterogeneous demand, and observes only whether each customer purchased. The setting combines censored semi-bandit feedback with a joint combinatorial decision, and sits at the intersection of three lines of work studied so far in isolation. Xu & Wang (2026) establish the optimal $\widetilde{O}(T^{2/3})$ regret for linear-valuation contextual pricing under agnostic, possibly atomic, demand noise, but consider neither inventory nor allocation. Xu et al. (2025) couple a uniform



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.

price with multi-supplier inventory and a downstream allocation problem, but their analysis is non-contextual and relies on structure that fails under censored purchase feedback. Adaptive conformal inference (Gibbs & Candès, 2021) and its semi-bandit and censored-data extensions (Yang et al., 2026; Ren et al., 2026) deliver distribution-free coverage online, but do not engage with the regret problem that arises when the coverage decision is the inventory commitment itself.

**Problem and algorithm.** We formulate contextual pricing with multi-supplier inventory and price-dependent allocation under censored purchase feedback and agnostic non-Lipschitz noise (§3). The seller pursues two criteria at once: cumulative profit close to that of a clairvoyant who knows the valuation parameter and the noise law, and a realised stockout rate no larger than a user-specified target $\alpha$. We propose CP-CPA (Conformal Contextual Pricing and Allocation, §4), which prices by maximising an optimistic profit on a residual-grid survival estimate, provisions the estimated profit-optimal inventory, raises it to the demand quantile at an online conformally calibrated level whenever that quantile is larger, and allocates the result across suppliers cheapest first. The calibrated level is maintained by an adaptive conformal update driven only by the realised stockout indicator, so the layer is compatible with censored feedback and makes no assumption on the shape of the demand curve.

**Why a plug-in quantile fails.** It is tempting to run a contextual pricing algorithm and read off an inventory quantile from its demand estimate. This composition provably miscovers, and the reason is the technical crux of the paper. The pricing regret controls an *expected-profit* functional of the survival estimate, which is Lipschitz in the buy probabilities by a coupling bound on poisson-binomial tails, so estimation error averages out. The inventory decision is a *quantile* of the same estimate, and near an atom of the survival function the quantile map is discontinuous: an $O(T^{-1/3})$ estimation error moves the integer quantile by $O(1)$ units and the miscoverage does not vanish with $T$. On a demand curve calibrated from a public retail log the plug-in quantile still stocks out $15\%$ of the time after $24$ episodes against a $10\%$ target (§6). CP-CPA decouples the two functionals: the pricing layer owns the profit estimate and the conformal loop owns the quantile level, so coverage holds whatever the plug-in bias.

**Contributions.**

1. **Regret with an explicit price of safety** (Theorem 1). Under bounded agnostic noise and stochastic well-conditioned contexts, the regret of CP-CPA against the unconstrained clairvoyant is $\widetilde{O}(T^{2/3}\mathrm{poly}(m,n,d)) + \gamma_{\max}\mathbb{E}[D_T^+]$, where $D_T^+$ is the observable total safety inventory added by the conformal layer. When the coverage target is slack this term is itself $\widetilde{O}(T^{2/3})$, matching the lower bounds of Kleinberg & Leighton (2003) and Xu & Wang (2022). When it binds we prove that every coverage-feasible algorithm has linear regret against the unconstrained benchmark (Proposition A.1), so the term is the price of the constraint, not of the algorithm. The proof replaces the Lipschitz and piecewise-convexity arguments of prior joint pricing and allocation analyses, which fail under atomic noise, by a coupling bound on poisson-binomial tails and a monotone-demand comparator.
2. **Deterministic coverage** (Theorem 2). The same algorithm satisfies $\widehat{\alpha}_T \le \alpha + O(1/(\eta T))$ for every realisation, with no assumption on the noise and no accuracy requirement on the demand model.
3. **Segment-conditional coverage** (Theorem 3). One conformal state per cell of a fixed context partition gives the same deterministic guarantee on every cell at the price of its visit count, with no change to the regret bound.
4. **Empirical validation** (§6). On synthetic benchmarks with Lipschitz and atomic noise the fitted regret exponents match $T^{2/3}$ and lie far below the distribution-free baseline of Xu & Wang (2022); when the target binds, CP-CPA realises a stockout rate of $0.096$ with zero variance across seeds while the plug-in quantile over-provisions at higher regret. On a semi-real retail simulator and on an episode calibrated from the public UCI Online Retail II log, CP-CPA is the only method that meets the target, at a profit cost of $0.1\%$, and the per-segment variant never exceeds the target in any segment.

## 2 Related Work

**Contextual dynamic pricing.** Regret rates for posted-price learning with linear contextual valuations are governed by the regularity of the noise. Log-concave or parametric noise admits $O(\log T)$ rates (Javanmard & Nazerzadeh, 2019; Cohen et al., 2020); smoothness of order $k$ gives $\widetilde{O}(T^{(2k+1)/(4k-1)})$ (Fan et al., 2024); Lipschitz noise gives $\widetilde{O}(T^{2/3})$ (Luo et al., 2022; Tullii et al., 2024). The agnostic regime, with unknown bounded noise that may have atoms, is the closest to ours: Xu & Wang (2022) obtain $\widetilde{O}(T^{3/4})$ with D2-EXP4 and Xu & Wang (2026) reach the optimal $\widetilde{O}(T^{2/3})$ with a residual-grid method that we adopt as our pricing layer. We are not aware of prior work that equips such an algorithm with a coverage guarantee or with multi-supplier inventory.

**Joint pricing and inventory.** Starting from Besbes & Zeevi (2009), a large operations literature learns demand while controlling inventory, including replenishment (Chen et al., 2019), perishable goods (Chen et al., 2020; Keskin et al., 2022), and censored demand (Chen et al., 2021; 2024). These works are non-contextual and their inventory analyses rely on smoothness of the demand curve. Xu et al. (2025) add multi-supplier allocation with lower-confidence optimism over joint price and inventory decisions, again without context and with structural assumptions that atomic noise violates. We inherit their sorted-cost allocation structure and drop the smoothness.

**Online conformal prediction and decisions.** Conformal prediction (Vovk et al., 2005) gives distribution-free finite-sample coverage; adaptive conformal inference (Gibbs & Candès, 2021) extends it online under arbitrary shift through a Robbins–Monro update of the miscoverage level, and later work adds strong adaptivity (Bhatnagar et al., 2023; Gibbs & Candès, 2024), semi-bandit feedback (Yang et al., 2026; Ge et al., 2024), and right-censored outcomes (Ren et al., 2026; Candès et al., 2023). Conformal ideas have also been used offline to set newsvendor order quantities (Cao, 2024). None of these works couples an online conformal calibrator with a regret-optimal censored-feedback bandit so that an optimal regret rate and a non-asymptotic coverage rate hold for one structured operational decision, which is what CP-CPA provides.

## 3 Problem Setup

**Protocol.** At each round $t = 1, \dots, T$ the seller observes contexts $\{x_{t,j}\}_{j=1}^{n} \subset \mathbb{R}^d$, one per consumer, posts a uniform price $p_t \in [0, B]$, and provisions inventory $\vec{I}_t \in \mathbb{Z}_+^m$ across $m$ suppliers, $K_t = \sum_i I_{t,i}$. Consumer $j$ has a private valuation $v_{t,j} = \langle x_{t,j}, \theta^\star \rangle + \xi_{t,j}$ with $\theta^\star$ fixed and unknown and $\xi_{t,j}$ i.i.d. from an unknown noise law. The realised demand is $N_t = \sum_j \mathbf{1}\{p_t \le v_{t,j}\}$, the seller serves $\min(N_t, K_t)$ units cheapest supplier first, and observes only the purchase indicators $o_{t,j} = \mathbf{1}\{p_t \le v_{t,j}\}$ and the count $N_t$; valuations are never observed. Writing $c(k)$ and $\gamma(k)$ for the supply and holding cost of the $k$-th cheapest provisioned unit, the net profit is

$$\Pi_t(p, K) := p \cdot \min(N_t, K) - \textstyle\sum_{k=1}^{\min(N_t,K)} c(k) - \sum_{k=1}^{K} \gamma(k). \tag{1}$$

Let $u_{t,j} := \langle x_{t,j}, \theta^\star \rangle$ and $S(w) := \mathbb{P}(\xi \ge w)$ be the residual survival function. With a right-continuous convention, $\mathbb{E}[o_{t,j} \mid x_{t,j}, p_t] = S(p_t - u_{t,j})$ for any noise law, atoms included.

**Objective.** Let $(p_t^\star, K_t^\star)$ maximise $\mathbb{E}[\Pi_t(p, K) \mid x_t]$ over $p \in [0, B]$ and $K \in \{0, \dots, nm\}$, and let $\Pi_t^\star$ be the optimal value. The seller pursues two criteria: the profit regret $R_T := \sum_{t \le T} \big(\Pi_t^\star - \mathbb{E}[\Pi_t(p_t, K_t) \mid x_t]\big)$ against this *unconstrained* clairvoyant, and the stockout coverage $\widehat{\alpha}_T := \frac{1}{T} \sum_{t \le T} \mathbf{1}\{N_t > K_t\} \le \alpha + o(1)$ for a prescribed $\alpha \in (0, 1)$. The two can conflict: if the clairvoyant stocks out with probability above $\alpha$ (the target *binds*), any policy meeting the coverage criterion carries safety inventory every round and linear regret against the unconstrained benchmark is unavoidable (Proposition A.1). Our bound isolates this cost as an explicit observable term; everything else, and the term itself when the target is *slack*, is $\widetilde{O}(T^{2/3})$.

**Assumptions.** We inherit the assumptions of the two models we combine, and nothing more.

**Assumption 1** (Agnostic bounded noise)**.** *$\{\xi_{t,j}\}$ are i.i.d. with support in $[-c, c]$ and $\mathbb{E}[\xi] = 0$. No density, Lipschitz, smoothness, log-concavity, or parametric condition is imposed.*

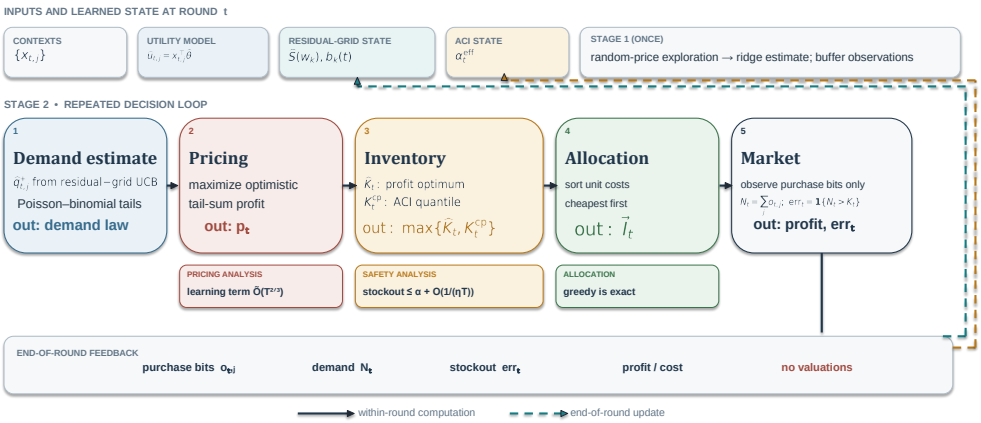



Figure 1: **CP-CPA method flow.** The learned utility model, the residual-grid survival estimate, and the conformal level enter the repeated Stage-2 loop. Each round the algorithm forms an optimistic demand law, maximises tail-sum profit to post $p_t$, combines the estimated profit optimum $\widehat{K}_t$ with the conformal quantile $K_t^{\mathrm{cp}}$, allocates $K_t$ units cheapest first, and observes only censored purchases, realised demand, profit, and the stockout indicator. Solid arrows are within-round computations; dashed arrows are the end-of-round updates of the residual-grid and conformal states.

**Assumption 2** (Well-conditioned stochastic contexts)**.** *$\{x_{t,j}\}$ are i.i.d. with $\|x\|_2 \le B_x$ and $\lambda_{\min}(\mathbb{E}[xx^\top]) \ge \lambda_0 > 0$.*

**Assumption 3** (Price buffer)**.** *There is $\kappa > 0$ with $\langle x, \theta^\star\rangle - c \ge \kappa$ and $\langle x, \theta^\star\rangle + c \le B$ almost surely, and $T$ is large enough that $8\Delta \le \min(\kappa, \gamma_{\min})$ for the Stage-1 accuracy $\Delta$ of §4, where $\gamma_{\min} = \min_i \gamma_i > 0$.*

**Assumption 4** (Sorted, perishable supply)**.** *Unit supply costs are sorted, $c_1 \le \cdots \le c_m$, and units are provisioned and served cheapest first, so the marginal value of the $k$-th unit, $g_k(p) = \mathbb{P}(N_t \ge k \mid x, p)(p - c_{(k)}) - \gamma_{(k)}$, is non-increasing in $k$ and the profit-optimal inventory $K^\star(p) = \#\{k : g_k(p) > 0\}$ is greedy. Unused units perish at the end of the round.*

Assumptions 1 to 3 are those of Xu & Wang (2026); Assumption 4 is the supply structure of Xu et al. (2025). The latter work additionally assumes that the price-to-profit map is piecewise convex on the intervals induced by the supply costs. We do not: under atomic noise that map need not even be continuous, and our analysis uses only the greedy structure above together with monotonicity of demand in the price (Appendix A, Remark A.1).

## 4 The CP-CPA Algorithm

CP-CPA runs in two stages (Figure 1; a step-by-step listing is given in Appendix D). Stage 1 lasts $T_1 = \lceil T^{2/3} \rceil$ rounds, posts uniformly random prices, provisions full capacity, and fits a ridge estimate $\widehat{\theta}$ from the censored purchase indicators. Stage 2 repeats three steps: an optimistic pricing step on a residual grid, a conformal inventory step, and a cheapest-first allocation step.

### 4.1 Residual-grid survival estimate and optimistic pricing

After Stage 1, $\max_{t,j} |\langle x_{t,j}, \widehat{\theta}\rangle - u_{t,j}| \le \Delta$ with $\Delta \asymp T^{-1/3}$ on a good event (Lemma A.4). Fix a residual grid $-c = w_0 < \cdots < w_M = c$ of mesh $2\Delta$, $M = \Theta(\Delta^{-1})$. Each censored observation $(p, \widehat{u}_{t,j}, o_{t,j})$ is assigned to the cell containing its *estimated* residual $p - \widehat{u}_{t,j}$; on the good event the true residual lies within a three-cell window, so the purchase probability lies between adjacent survival levels (Lemma A.5). This is what survives atoms: an $O(\Delta)$ utility error cannot move a purchase probability outside a window of neighbouring survival values. The seller keeps the empirical mean $\widehat{S}(w_k)$ and a confidence radius $b_k(t) = C_{\mathrm{ucb}}\sqrt{\log(nT)/\max(1, N_k(t))}$ over the $N_k(t)$ observations in cell $k$, and the optimistic survival $\widehat{S}^+(w_k) = \min(1, \widehat{S}(w_k) + b_k(t))$.

Write $\widehat{q}^{+}_{t,j}(p) = \widehat{S}^{+}(p - \langle x_{t,j}, \widehat{\theta}\rangle)$ for the optimistic buy probabilities, $\widehat{T}^{+}_{k}(p)$ for the tail $\mathbb{P}(N \geq k)$ of the poisson-binomial law with these parameters, and $(c_{(k)}, \gamma_{(k)})$ for the costs of the $k$-th cheapest unit. The price maximises optimistic *profit*, with the inventory optimised out greedily,

$$p_t \in \arg\max_{p \in \mathcal{P}} \sum_{k=1}^{mn} \Big(\widehat{T}^{+}_{k}(p)\big(p - c_{(k)}\big) - \gamma_{(k)}\Big)_{+}, \tag{2}$$

over a candidate grid $\mathcal{P} \subset [0, B]$ of mesh at most $2\Delta$. Maximising optimistic *revenue* instead would target the revenue-optimal price, which differs from the profit-optimal one by a constant and costs linear regret (Appendix A, remark after Lemma A.9).

### 4.2 Conformal inventory layer

Given $p_t$, the estimated buy probabilities are $\widehat{q}_{t,j} = \widehat{S}(p_t - \langle x_{t,j}, \widehat{\theta}\rangle)$, $\widehat{P}_t$ is the poisson-binomial law with these parameters, and the estimated profit-optimal inventory is $\widehat{K}_t = \#\{k : \widehat{T}_k(p_t)(p_t - c_{(k)}) - \gamma_{(k)} > 0\}$ with $\widehat{T}_k$ the tail of $\widehat{P}_t$. The seller maintains a level $\alpha^{\mathrm{eff}}_t \in \mathbb{R}$, initialised at $\alpha$ when Stage 2 begins, and provisions

$$K_t := \max\big(\widehat{K}_t,\, K^{\mathrm{cp}}_t\big), \qquad K^{\mathrm{cp}}_t := \min\Big\{K : \mathbb{P}_{N \sim \widehat{P}_t}(N \leq K) \geq \Pi_{[0,1]}\big(1 - \alpha^{\mathrm{eff}}_t\big)\Big\}, \tag{3}$$

with the convention $K^{\mathrm{cp}}_t = n$ when $\alpha^{\mathrm{eff}}_t \leq 0$ and $K^{\mathrm{cp}}_t = 0$ when $\alpha^{\mathrm{eff}}_t \geq 1$. The conformal quantile can only *add* inventory on top of the profit optimum; when the profit optimum already meets the target the layer is inactive and costs nothing. After the round the seller observes $\mathrm{err}_t = \mathbf{1}\{N_t > K_t\}$ and updates, without clipping,

$$\alpha^{\mathrm{eff}}_{t+1} = \alpha^{\mathrm{eff}}_t + \eta\,(\alpha - \mathrm{err}_t), \tag{4}$$

with a fixed step $\eta \in (0, 1]$. This is the adaptive conformal update of Gibbs & Candès (2021) applied to the stockout indicator; it needs only the realised count $N_t$, never a valuation.

**Why the conformal loop is necessary.** The pricing layer guarantees that the buy-probability estimates are accurate on average along the played sequence, and this is all that the profit functional in (2) needs: poisson-binomial tails are Lipschitz in the buy probabilities (Lemma A.1), so estimation error becomes regret one for one and averages out. The coverage functional is the quantile $Q_{\widehat{P}_t}(1 - \alpha)$, and the map from a pmf to its quantile is not Lipschitz when $S$ has an atom: if a mass of order $\alpha$ sits within an $\widetilde{O}(T^{-1/3})$ band of the threshold, an $\widetilde{O}(T^{-1/3})$ perturbation of $\widehat{S}$ moves the integer quantile by $O(1)$ units. The plug-in coverage error is therefore a systematic bias whose sign depends on the local shape of $S$, and no better estimator removes it. The update (4) never trusts the plug-in level; it drives $\alpha^{\mathrm{eff}}_t$ from the realised stockouts until the empirical rate is at most $\alpha$, absorbing an arbitrary and unknown plug-in bias into a deterministic $O(1/(\eta T))$ term (Theorem 2).

### 4.3 Allocation layer

With suppliers sorted by unit cost and capacities $\kappa_i$, the inventory is $I_{t,i} = \min\big(\max(0, K_t - \sum_{i' < i} \kappa_{i'}), \kappa_i\big)$, and realised demand is served cheapest first. The greedy rule is exact because the price is uniform across consumers, supply costs are consumer independent, and inventory is committed before demand realises; the allocation program of Xu et al. (2025) reduces to it under these conditions, and consumer-dependent costs change neither rate. Each Stage-2 round costs $O(|\mathcal{P}|n^2 + m)$ time; the only hyperparameter beyond the schedule of Xu & Wang (2026) is $\eta$, to which coverage is insensitive across two orders of magnitude (Appendix E.3).

## 5 Theoretical Analysis

We state the three guarantees with proof sketches; full proofs are in Appendices A to C. A single algorithm, under one set of assumptions, achieves the agnostic-pricing regret rate on its learning terms and a distribution-free non-asymptotic coverage rate, and the analysis isolates the one quantity through which the two interact: the *safety inventory*

$$D^{+}_{T} := \sum_{t > T_1} \big(K^{\mathrm{cp}}_t - \widehat{K}_t\big)_{+}, \tag{5}$$

the number of units the conformal layer adds on top of the profit optimum. It is observable and we report it in every experiment. We also use a mild condition on how often the shared price places an estimated residual next to a large increment of $S$ (the *adjacent-jump budget*, Assumption A.5 in Appendix A); it holds automatically when $S$ is Lipschitz and, for a single consumer, under the redirect rule of Xu & Wang (2026).

### 5.1 Regret

**Theorem 1** (Regret of CP-CPA)**.** *Under Assumptions 1 to 4 and the adjacent-jump budget, with* $T_1 = \lceil T^{2/3} \rceil$ *and* $\Delta \asymp T^{-1/3}$*,*

$$\mathbb{E}[R_T] \;\le\; \widetilde{O}\big(T^{2/3}\,\mathrm{poly}(m, n, d, B, B_x, 1/\lambda_0, 1/\kappa)\big) + \gamma_{\max}\,\mathbb{E}[D_T^+],$$

*against the unconstrained clairvoyant. In the* slack regime*, where every unit up to the* $(1 - \alpha + \rho)$ *demand quantile has marginal profit at least* $\rho' > 0$*,* $\gamma_{\max}\mathbb{E}[D_T^+] = O(n\gamma_{\max}/\rho^2) + \widetilde{O}(n^2 B\gamma_{\max}T^{2/3}/(\rho\rho'))$*, hence* $\mathbb{E}[R_T] = \widetilde{O}(T^{2/3})$*. In the* binding regime*, where profit maximisation alone stocks out with probability above* $\alpha$ *on a set of contexts of positive mass with constraint gap* $\rho^\alpha > 0$*, every algorithm whose stockout rate is at most* $\alpha + o(1)$ *has* $\mathbb{E}[R_T] \ge \rho^\alpha T - o(T)$ *(Proposition A.1). The exponent* $2/3$ *of the learning terms matches the lower bounds of Kleinberg & Leighton (2003) and Xu & Wang (2022).*

**Proof sketch.** On the good event of Stage-1 recovery and the uniform survival confidence bound, the per-round regret splits as

$$\begin{aligned}\Pi_t^\star - \mathbb{E}[\Pi_t(p_t, K_t)] = \underbrace{\Pi_t^\star - \mathbb{E}[\Pi_t(p_t, K^\star(p_t))]}_{\text{(A) pricing}} + \underbrace{\mathbb{E}[\Pi_t(p_t, K^\star(p_t))] - \mathbb{E}[\Pi_t(p_t, \widehat{K}_t)]}_{\text{(B) inventory estimation}} \\ + \underbrace{\mathbb{E}[\Pi_t(p_t, \widehat{K}_t)] - \mathbb{E}[\Pi_t(p_t, K_t)]}_{\text{(C) safety}},\end{aligned} \tag{6}$$

with $K^\star(p)$ the profit-optimal inventory at price $p$. For (A), the expected profit at price $p$ with optimal inventory is the tail sum $W_t(p) = \sum_k (\mathbb{P}(N_t(p) \ge k)(p - c_{(k)}) - \gamma_{(k)})_+$ and the algorithm maximises its optimistic version. Two facts replace the Lipschitz and convexity structure of the joint pricing literature, neither of which holds under atomic noise: poisson-binomial tails are 1-Lipschitz in the buy probabilities in $\ell_1$ (Lemma A.1), and demand is monotone in the price. Monotonicity yields a comparator price $\tilde{p} \in [p_t^\star - 8\Delta, p_t^\star - 6\Delta]$ at which every optimistic buy probability dominates the true one at $p_t^\star$, so the optimistic objective at $\tilde{p}$ is at least $W_t(p_t^\star) - 8\Delta mn$ without paying any adjacent jump, and the regret reduces to confidence radii and adjacent jumps at the played price, whose sums are $\widetilde{O}(\sqrt{n}\,T^{2/3})$ and $\widetilde{O}(nT^{2/3})$ (Lemma A.9). For (B), both the true and the estimated optimal inventories are thresholds of non-increasing marginal values, and misplacing a threshold costs at most $mnB\sum_j |\widehat{q}_{t,j} - q_{t,j}|$ (Lemma A.2), which inherits the budgets of (A). For (C), since $K_t = \max(\widehat{K}_t, K_t^{\mathrm{cp}})$ the conformal layer only adds units beyond the estimated optimum and each costs at most its holding cost, so (C)$\le \gamma_{\max}(K_t^{\mathrm{cp}} - \widehat{K}_t)_+$ exactly (Lemma A.11). In the slack regime the level drifts upward at rate at least $\eta\rho$ and an Azuma argument shows the conformal quantile exceeds $\widehat{K}_t$ on $O(1/\rho^2)$ rounds plus the $\widetilde{O}(T^{2/3})$ rounds where the estimate is inaccurate (Lemma A.12). In the binding regime $D_T^+$ is linear, as Proposition A.1 says it must be.

### 5.2 Coverage

**Theorem 2** (Stockout coverage of CP-CPA)**.** *For any* $\eta \in (0, 1]$ *and* $T > T_1$*, deterministically for every realisation of contexts, demand, and estimation errors,*

$$\widehat{\alpha}_T = \frac{1}{T}\sum_{t=1}^{T} \mathrm{err}_t \;\le\; \alpha + \frac{\alpha + \eta}{\eta\,(T - T_1)} = \alpha + O\Big(\frac{1}{\eta T}\Big).$$

*Moreover, on the good event of Theorem 1, the per-round conditional stockout probability satisfies* $\mathbb{P}(\mathrm{err}_t = 1 \mid \textit{past}, p_t) \le \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}}) + \varepsilon_t$ *with* $\varepsilon_t = \sum_j |\widehat{q}_{t,j} - q_{t,j}|$ *and* $\frac{1}{T}\sum_t \varepsilon_t = \widetilde{O}(nT^{-1/3})$*.*

**Proof sketch.** Summing the unclipped recursion (4) over Stage 2 gives $\sum_{t>T_1}(\mathrm{err}_t - \alpha) = (\alpha - \alpha^{\mathrm{eff}}_{T+1})/\eta$ exactly (Lemma B.2). Whenever $\alpha^{\mathrm{eff}}_t \le 0$ the quantile in (3) saturates at full capacity, no stockout occurs, and the level increases, so $\alpha^{\mathrm{eff}}_{T+1} \ge -\eta$ always (Lemma B.1); Stage 1 provisions fully and contributes no stockouts (Lemma B.5). The per-round statement follows because the quantile is taken under $\widehat{P}_t$ rather than $P_t$ and the two laws are within $\varepsilon_t$ in total variation (Lemmas A.1 and B.4); this term affects only the conditional statement, the long-run rate is exact regardless of it.

**How the two bounds meet.** The conformal layer can only increase inventory above $\widehat{K}_t$, and its entire regret cost is $\gamma_{\max} D^+_T$: $\widetilde{O}(T^{2/3})$ when the target is slack, linear and unavoidable when it binds, in which case the right comparison is the constrained clairvoyant, against which CP-CPA again loses only the learning terms plus its excess safety inventory. The coverage guarantee is one sided by design, since a coverage constraint is an inequality. Freezing the level at $1-\alpha$ (the plug-in ablation of §6) leaves the learning terms unchanged but adds a constant, uncontrolled coverage offset whenever the plug-in quantile is biased.

### 5.3 Segment-conditional coverage

Theorem 2 is marginal. Service-level agreements are usually stated per customer segment, so we strengthen it to every cell of a fixed partition. Let $g : \mathcal{X}^n \to \{1, \dots, G\}$ be a fixed map from the round context to one of $G$ anchors, $g_t = g(x_t)$, and $V_g(T) = \#\{T_1 < t \le T : g_t = g\}$. Replace (4) by

$$\alpha^{(g)}_{t+1} = \alpha^{(g)}_t + \eta\, \mathbf{1}\{g_t = g\}\,(\alpha - \mathrm{err}_t), \qquad \alpha^{(g)}_{T_1+1} = \alpha, \tag{7}$$

and use the local level $\alpha^{(g_t)}_t$ in (3). We call the result CP-CPA-WACI; it runs one conformal state per anchor, in the spirit of Mondrian conformal prediction (Vovk et al., 2005). Let $\widehat{\alpha}^{(g)}_T = V_g(T)^{-1} \sum_{T_1 < t \le T:\, g_t = g} \mathrm{err}_t$.

**Theorem 3** (Anchor-conditional stockout coverage)**.** *For any anchor map, any $\eta \in (0, 1]$, every anchor with $V_g(T) \ge 1$, and every realisation, $\widehat{\alpha}^{(g)}_T \le \alpha + (\alpha + \eta)/(\eta V_g(T))$. If round contexts are i.i.d. with $\mathbb{P}(g(x_t) = g) \ge \pi_{\min}$ for all $g$, then with probability $1 - \delta$ and simultaneously for all anchors, $\widehat{\alpha}^{(g)}_T \le \alpha + 2(\alpha + \eta)/(\eta \pi_{\min}(T - T_1))$ once $T - T_1 \ge 8 \log(G/\delta)/\pi_{\min}$. The regret bound of Theorem 1 is unchanged.*

Restricted to the rounds visiting anchor $g$, (7) is the unclipped recursion of Theorem 2, so the telescoping identity and the saturation bound apply on that subsequence with $V_g(T)$ as the horizon (Lemma C.2); a Chernoff bound on $V_g(T)$ gives the uniform statement, and the anchor-wise mechanism enters the regret only through $D^+_T$, for which Lemma A.11 holds with the local levels (Appendix C). The price of conditional validity is the effective horizon: a segment visited a fraction $\pi_g$ of the time converges at rate $1/(\eta \pi_g T)$.

## 6 Experiments

The experiments isolate the claims of §5: in the slack regime the conformal layer is inactive and the regret slope matches $T^{2/3}$ (E1); in the binding regime the stockout rate is pinned to $\alpha$ in every seed at an explicit safety-inventory cost (E2); coverage is insensitive to $\eta$ (E3); the guarantees survive a non-i.i.d. semi-real retail stream (E4), hold per segment (E5), and transfer to a demand curve calibrated from a public retail log (E6). Full environment, method, and protocol details, all tables, and the additional figures are in Appendix E; code and raw runs are in the supplementary material.

**Setup.** Each round draws $n = 10$ contexts $x_{t,j} \in \mathbb{R}^5$ (intercept, a shared per-round market shift, three per-consumer features), $\theta^\star = (2, 1, 0.3, 0.3, 0.3)$, noise support $[-1, 1]$, and one of two noise laws: *uniform* on $[-1, 1]$ (Lipschitz survival) and *cliff*, an atom of mass $0.3$ at zero mixed with the uniform (a jump of $0.3$ in $S$). Three suppliers have unit costs $\{0.50, 0.65, 0.80\}$ and holding costs $\{0.10, 0.13, 0.16\}$. With these costs the profit optimum stocks out in only 2 to 3% of rounds, so $\alpha = 0.10$ is slack; multiplying holding costs by 5 pushes the stockout rate of the profit optimum

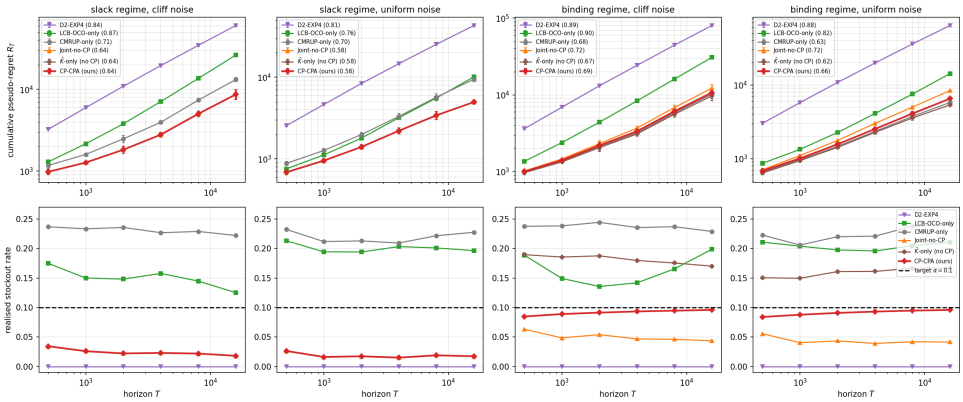



Figure 2: **Synthetic benchmark**, 5 seeds. Left two columns: default (slack) holding costs; right two columns: holding costs times 5 (binding). Top: cumulative pseudo-regret against the unconstrained oracle (log-log, fitted slopes in the legend); bottom: realised stockout rate, on cliff and uniform noise. In the slack regime CP-CPA (red) coincides with JOINT-NO-CP (orange). In the binding regime CP-CPA is the only method on the $\alpha = 0.10$ line; the plug-in JOINT-NO-CP over-provisions at a steeper regret slope, while $\widehat{K}$-only (brown) follows the unconstrained oracle by stocking out 17% of the time.

Table 1: **(E2) Binding regime at** $T = 16000$, 5 seeds, $\alpha = 0.10$: regret slope $\widehat{\beta}$ with bootstrap 95% interval, realised stockout rate (mean $\pm$ std), and safety inventory per round, on cliff / uniform noise. Only CP-CPA meets the target; its zero seed-to-seed variance is the deterministic guarantee of Theorem 2.

| Method | $\widehat{\beta}$ (cliff) | $\widehat{\beta}$ (unif.) | Stockout (cliff / unif.) | $D_T^+/T$ |
|---|---|---|---|---|
| $\widehat{K}$-only | 0.669 [0.605, 0.739] | 0.623 [0.580, 0.665] | $0.170 \pm 0.019$ / $0.167 \pm 0.003$ | n/a |
| CMRUP-ONLY | 0.677 [0.624, 0.733] | 0.634 [0.589, 0.679] | $0.229 \pm 0.033$ / $0.240 \pm 0.014$ | n/a |
| LCB-OCO-ONLY | 0.904 [0.900, 0.909] | 0.816 [0.794, 0.832] | $0.199 \pm 0.040$ / $0.211 \pm 0.025$ | n/a |
| JOINT-NO-CP | 0.722 [0.668, 0.769] | 0.720 [0.693, 0.746] | $0.043 \pm 0.012$ / $0.042 \pm 0.005$ | 0.83 / 0.93 |
| D2-EXP4 | 0.893 [0.885, 0.902] | 0.882 [0.869, 0.893] | $0.000 \pm 0.000$ / $0.000 \pm 0.000$ | n/a |
| CP-CPA (ours) | 0.687 [0.631, 0.743] | 0.659 [0.620, 0.697] | $\mathbf{0.096 \pm 0.000}$ / $\mathbf{0.096 \pm 0.000}$ | 0.44 / 0.47 |

to 0.17 to 0.19 and makes the target bind. We compare $\widehat{K}$-ONLY (our stack with the conformal layer removed), CMRUP-ONLY (contextual pricing with expected-demand inventory, Xu & Wang, 2026), LCB-OCO-ONLY (non-contextual price optimism with a plug-in quantile, Xu et al., 2025), JOINT-NO-CP (our stack with the level frozen at $1-\alpha$, the ablation that isolates the conformal loop), D2-EXP4 (Xu & Wang, 2022), and CP-CPA with $\eta = 0.05$. We report pseudo-regret against the unconstrained clairvoyant, the realised stockout rate over all $T$ rounds, and $D_T^+/T$; 5 seeds per cell; slopes by log-log least squares with bootstrap 95% intervals.

**E1: slack regime.** On $T \in \{500, \ldots, 16000\}$ (Figure 2, left; Table 3), CP-CPA, JOINT-NO-CP, and $\widehat{K}$-only are bit-identical in 57 of 60 runs and differ by at most 9 safety units in the other three: $\alpha_t^{\text{eff}}$ drifts above one and the layer switches itself off, as Lemma A.12 predicts, at a stockout rate of 0.017 to 0.018. Their regret slopes, 0.59 to 0.64, are consistent with $T^{2/3}$, whereas D2-EXP4 (0.82 to 0.84) and LCB-OCO-ONLY (0.76 to 0.87) are steeper and bootstrap separated, with 7 to 9 times and 2 to 3 times the regret of CP-CPA at $T = 16000$.

**E2: binding regime.** With holding costs times 5 the target binds (Figure 2, right; Table 1). CP-CPA realises $0.096 \pm 0.000$ on both noise laws: the Stage-2 rate is at most $\alpha + (\alpha+\eta)/(\eta(T-T_1)) = 0.1002$ by Theorem 2, and the small deficit is the Stage-1 dilution $\alpha T_1/T$. The plug-in JOINT-NO-CP over-provisions to 0.043 on the same demand estimates and pays with about twice the safety inventory, 14 to 29% higher regret at $T = 16000$, and a steeper slope; $\widehat{K}$-only has the lowest regret, as it must, but stocks out in 17% of rounds. Because $\widehat{K}$-only and CP-CPA share the pricing

trajectory seed by seed, their regret gap is exactly the empirical term (C) of (6): it is $0.063$ to $0.085$ per round with no trend in $T$ from $500$ to $16000$ (Table 5), the linear price of safety that Proposition A.1 says no safe policy can avoid.

**E3: step size.** Sweeping $\eta \in \{0.01, 0.05, 0.10, 0.20\}$ at $T = 4000$ in the binding regime leaves the stockout rate at $0.091$ to $0.094$, the deterministic ceiling of Theorem 2 up to seed noise below $0.0015$, while regret grows by 9 to $15\%$ from the smallest to the largest $\eta$ through a wider oscillation of $\alpha_t^{\text{eff}}$ (Table 6). We use $\eta = 0.05$ everywhere without tuning.

**E4: semi-real retail simulator.** A retail SKU simulator with calendar seasonality, a Markov promotion regime, and a mixture of regular and deal-seeking consumers with different noise atoms ($0.10$ versus $0.55$) provides a non-i.i.d. stream with heterogeneous subgroups (Appendix E.4). In four of five seeds the target is slack and the three residual-grid variants coincide. In the fifth the Stage-1 estimate is poor, the profit optimum built from the biased pmf stocks out $33\%$ of the time, and the plug-in $0.9$-quantile of the same pmf is no help ($0.330$). CP-CPA is capped at $0.095$ in every seed, because Theorem 2 does not depend on the quality of the estimate, at lower regret ($13757$ versus $14268$) and slightly higher profit (Table 7).

**E5: per-segment coverage.** We instantiate (7) on the semi-real simulator with $G = 4$ anchors (promotion regime times deal-seeker share), holding costs times $10$ so that the target binds in most anchors, $T \in \{2, \dots, 32\} \times 10^3$, and $10$ seeds (Figure 3). The worst-anchor overshoot $\max_g(\widehat{\alpha}_T^{(g)} - \alpha)_+$ bounded by Theorem 3 is zero in $47$ of $50$ runs for CP-CPA-WACI and at most $0.002$ otherwise, below the deterministic ceiling at every horizon. Marginal CP-CPA overshoots in $30$ of $50$ runs, by up to $0.156$ in one anchor, with no downward trend in $T$: one conformal state equalises the average across anchors but cannot remove per-anchor bias. Marginal rates, profit, and regret of the two variants are indistinguishable, as Theorem 3 predicts.

**E6: real-data-calibrated episode.** We calibrate an environment from the public UCI Online Retail II log (Chen et al., 2012): the highest-volume SKU is aggregated to a daily panel of about $589$ trading days, a weighted logistic demand model in price and real calendar features defines the valuation parameter and a bounded mean-zero survival curve, and a per-consumer income covariate is drawn from the customers' own spend distribution (Appendix E.6). One episode is a chronological pass over the trading year; we run $4$ to $24$ episodes. The log provides realistic contexts and a realistic seasonal survival curve but no counterfactual valuations, so profit and regret are computed against a model-based oracle and serve as consistency checks. CP-CPA realises $0.093$, $0.094$, $0.095$, $0.096$ at $4$, $8$, $16$, $24$ episodes with zero seed variance (Table 8). On this curve the estimated profit optimum already exceeds the estimated $0.9$-quantile, so the plug-in layer of JOINT-NO-CP is inactive and its rate is whatever the biased estimate of the non-Lipschitz curve delivers: $0.223$ after $4$ episodes, $0.145$ after $24$, and $0.454$ in December. The instance is estimated slack but truly binding, the case in which only feedback on realised stockouts can restore coverage. CP-CPA pays $2.0$ safety units per round, down from $5.9$ as the estimate sharpens, at a cost of $0.1\%$ of profit and $1.7\%$ of regret relative to JOINT-NO-CP; the December versus off-season split ($0.329$ versus $0.077$) reproduces the marginal-versus-conditional gap of E5.

## 7 CONCLUSION

CP-CPA combines a residual-grid contextual pricing layer with an adaptive conformal inventory layer that never provisions below the estimated profit optimum. Its regret is $\widetilde{O}(T^{2/3})$ plus exactly the holding cost of the safety inventory, a cost that is $\widetilde{O}(T^{2/3})$ when the coverage target is slack and unavoidable when it binds; its realised stockout rate satisfies $\widehat{\alpha}_T \le \alpha + O(1/(\eta T))$ deterministically, per segment if desired. The experiments reproduce each regime the theory distinguishes.

**Limitations and outlook.** The analysis assumes a stationary $\theta^\star$ and noise law and stochastic contexts; non-stationary variants (Baby et al., 2023; Bhatnagar et al., 2023) and a kernel-smoothed continuum of anchors (Appendix C) are natural extensions. Under atomic noise with a shared price, the regret bound relies on the adjacent-jump budget (Assumption A.5); a pricing rule that provably enforces it is open.

### AI USE STATEMENT

In this work we used large language models for research ideation and execution (refining the problem formulation, suggesting the decoupling of the profit functional from the quantile functional that underlies the conformal layer, and proposing the structure of the regret decomposition), for drafting and checking the proofs in Appendices A to C, for writing the simulation, baseline, and plotting code released in the supplementary material, and for drafting and polishing sections of the paper, including this statement. We did not use generative AI to generate the datasets themselves: the synthetic and semi-real environments are sampled by deterministic programs from the distributions specified in Appendix E, and the real-data episode is calibrated from the public UCI Online Retail II log. We did not use generative AI to select or interpret experimental results beyond producing tables and figures from the raw runs. All AI-assisted work was reviewed by the authors: every lemma and theorem was re-derived line by line, the code was tested against the closed-form oracle of the synthetic benchmark and inspected for correctness, and all claims in the text were checked against the raw experimental logs. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

### ETHICS STATEMENT

Dynamic pricing is socially consequential. Two safeguards are built into our setting: the price is uniform across consumers in each round, so individual-level price discrimination is not introduced, and the conformal layer guarantees a service-level floor on the realised stockout rate, which limits the probability that any consumer is denied service through under-provisioning; the per-segment variant extends that floor to every customer segment of a fixed partition. We discourage deployments that combine the algorithm with downstream individualised pricing. The only real data used is the public and anonymised UCI Online Retail II transaction log, from which we extract aggregate daily demand and a distribution of customer spend; no individual is identifiable. No human subjects were involved.

### REPRODUCIBILITY STATEMENT

All assumptions are stated in §3 and restated with the exact constants in Appendix A; complete proofs of Theorems 1, 2, and 3 are given in Appendices A, B, and C. A step-by-step description of CP-CPA and of all baselines, with every hyperparameter, is in Appendix D. The environments, the calibration pipeline of the real-data episode, the evaluation protocol, and the full result tables are in Appendix E. The supplementary material contains the anonymised source code, the raw result files of every run reported in the paper, the calibration artefacts derived from the public UCI log, and the commands that regenerate every table and figure.

## REFERENCES

Dheeraj Baby, Jianyu Xu, and Yu-Xiang Wang. Non-stationary contextual pricing with safety constraints. *Transactions on Machine Learning Research*, 2023.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 2009.

Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved online conformal prediction via strongly adaptive online learning. In *International Conference on Machine Learning (ICML)*, 2023.

Emmanuel Candès, Lihua Lei, and Zhimei Ren. Conformalized survival analysis. *Journal of the Royal Statistical Society Series B*, 85(1):24–45, 2023.

Junyu Cao. A conformal approach to feature-based newsvendor under model misspecification. *arXiv preprint arXiv:2412.13159*, 2024.

Boxiao Chen, Xiuli Chao, and Hyun-Soo Ahn. Coordinating pricing and inventory replenishment with nonparametric demand learning. *Operations Research*, 2019.

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

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*, 2021.

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.

Daqing Chen, Sai Laing Sain, and Kun Guo. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3):197–208, 2012. UCI Machine Learning Repository: Online Retail II Data Set.

Maxime C. Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *Management Science*, 2020.

Jianqing Fan, Yongyi Guo, and Mengxin Yu. Policy optimization using semiparametric models for dynamic pricing. *Journal of the American Statistical Association*, 2024.

Haosen Ge, Hamsa Bastani, and Osbert Bastani. Stochastic online conformal prediction with semi-bandit feedback. *arXiv preprint arXiv:2405.13268*, 2024.

Isaac Gibbs and Emmanuel Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Isaac Gibbs and Emmanuel J. Candès. Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, 2024.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 2019.

N. Bora Keskin, Yuexing Li, and Jing-Sheng Song. Data-driven dynamic pricing and ordering with perishable inventory in a changing environment. *Management Science*, 68(3):1938–1958, 2022.

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

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

Sijie Ren, Meng Yan, Zhen Zhang, Xu Yinghui, and Xinwei Sun. Conformalized survival counterfactuals prediction for general right-censored data. In *International Conference on Learning Representations (ICLR)*, 2026.

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

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

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 (AISTATS)*, 2022.

Jianyu Xu and Yu-Xiang Wang. Optimal contextual pricing under agnostic non-Lipschitz demand. *arXiv preprint arXiv:2605.05609*, 2026.

Jianyu Xu, Xuan Wang, Yu-Xiang Wang, and Jiashuo Jiang. Online pricing and allocation with demand learning and fulfillment cost. *arXiv preprint arXiv:2501.18049*, 2025.

Junyoung Yang, Kyungmin Kim, and Sangdon Park. Online conformal prediction with adversarial semi-bandit feedback via regret minimization. In *International Conference on Learning Representations (ICLR)*, 2026.

# A PROOF OF THEOREM 1

This appendix gives the full proof of Theorem 1 (the regret bound for CP-CPA). Section A.1 recalls the setup, fixes the exact decision rules that are analysed, and states the assumptions; Section A.2 gives the master regret decomposition into a *pricing* term (A), an *inventory-estimation* term (B), and a *conformal-safety* term (C); Section A.3 proves the supporting lemmas; Section A.4 analyses Term (C), including the lower bound showing that its linear part is unavoidable; Section A.5 assembles the theorem.

Two conventions are used throughout. First, all statements about Stage 2 are made on the "good event" $\mathcal{G}$ of Lemma A.13 (probability $\geq 1 - 3T^{-2}$); the complement contributes $O(1)$ to the regret and is handled once, in the proof of Theorem 4. Second, $\widetilde{O}(\cdot)$ hides polylogarithmic factors in $T$ and polynomial factors in $(m, n, d, B, B_x, 1/\lambda_0, 1/\kappa)$ unless these are displayed.

## A.1 SETUP, DECISION RULES, ASSUMPTIONS

At round $t = 1, \ldots, T$ the seller observes contexts $\{x_{t,j}\}_{j=1}^{n} \subset \mathbb{R}^d$, posts a single price $p_t \in [0, B]$, and provisions total inventory $K_t \in \{0, \ldots, mn\}$ across $m$ suppliers with per-unit supply costs $c_1 \leq \cdots \leq c_m$ and holding costs $\gamma_1, \ldots, \gamma_m$; each supplier has capacity $n$. Units are provisioned and served cheapest-supplier-first, so the $k$-th provisioned unit has costs $(c_{(k)}, \gamma_{(k)})$ with $c_{(k)}$ non-decreasing in $k$. Consumer $j$ has valuation $v_{t,j} = u_{t,j} + \xi_{t,j}$, $u_{t,j} := \langle x_{t,j}, \theta^\star \rangle$, and buys one unit iff $p_t \leq v_{t,j}$. With the survival function $S(w) := \mathbb{P}(\xi \geq w)$, the purchase probability of consumer $j$ at price $p$ is $q_{t,j}(p) := S(p - u_{t,j})$, the demand $N_t(p) = \sum_j \mathbf{1}\{p \leq v_{t,j}\}$ is Poisson–binomial with parameters $q_t(p) = (q_{t,j}(p))_j$, and the per-round net profit is

$$\Pi_t(p, K) \;=\; p \cdot \min(N_t, K) \;-\; \textstyle\sum_{k=1}^{\min(N_t, K)} c_{(k)} \;-\; \sum_{k=1}^{K} \gamma_{(k)}. \tag{8}$$

The seller observes the censored indicators $o_{t,j} = \mathbf{1}\{p_t \leq v_{t,j}\}$ and $N_t$.

**Tail-sum form of the expected profit.** Writing $\bar{T}_k(p) := \mathbb{P}(N_t(p) \geq k \mid x_t)$ for the demand tail, a direct computation from (8) gives

$$F_t(p, K) := \mathbb{E}[\Pi_t(p, K) \mid x_t] = \sum_{k=1}^{K} g_{t,k}(p), \qquad g_{t,k}(p) := \bar{T}_k(p)\,(p - c_{(k)}) - \gamma_{(k)}. \tag{9}$$

The marginal value $g_{t,k}(p)$ of the $k$-th unit is non-increasing in $k$ (the tail is non-increasing and $c_{(k)}$ non-decreasing), so

$$K_t^\star(p) := \arg\max_K F_t(p, K) = \#\{k : \; g_{t,k}(p) > 0\}, \quad W_t(p) := F_t(p, K_t^\star(p)) = \sum_{k=1}^{mn} \big(g_{t,k}(p)\big)_+, \tag{10}$$

i.e. the profit-optimal inventory at price $p$ is obtained greedily. Only this monotone ("sorted-costs") structure of the supply side is used below.

Let $\Pi_t^\star = \max_{p \in [0,B]} W_t(p)$ be the clairvoyant per-round optimum under known $(\theta^\star, S)$, attained at $p_t^\star$, and let $K_t^\star := K_t^\star(p_t^\star)$. The (pseudo-)regret against this *unconstrained* clairvoyant is

$$R_T \;=\; \sum_{t=1}^{T} \Big(\Pi_t^\star - \mathbb{E}[\Pi_t(p_t, K_t) \mid x_t]\Big). \tag{11}$$

**Decision rules analysed.** Stage 1 ($t \leq T_1$, exploration) posts $p_t \sim \mathrm{Unif}[0, B]$, provisions fully, and buffers the censored observations. At the end of Stage 1 the ridge estimate $\widehat{\theta}$ is frozen and *all* observations so far are assigned to residual cells using this single estimate. In Stage 2 ($t > T_1$, optimistic pricing) the algorithm holds predicted utilities $\widehat{u}_{t,j} = \langle x_{t,j}, \widehat{\theta} \rangle$, a residual grid $-c = w_0 < w_1 < \cdots < w_M = c$ of mesh $2\Delta$, survival estimates $\widehat{S}(w_k)$ with confidence radii $b_k(t)$ (Section A.3.3), and a candidate price set $\mathcal{P} \subset [0, B]$ of mesh $\leq 2\Delta$. For a candidate price $p$ and consumer $j$ let $k_j(p)$ be the grid index with $p - \widehat{u}_{t,j} \in (w_{k_j(p)-1}, w_{k_j(p)}]$, and define the point and optimistic buy-probability estimates

$$\widehat{q}_{t,j}(p) := \widehat{S}\big(w_{k_j(p)}\big), \qquad \widehat{q}^{+}_{t,j}(p) := \min\big\{1,\; \widehat{S}\big(w_{k_j(p)}\big) + b_{k_j(p)}(t)\big\},$$

and the corresponding Poisson–binomial tails $\widehat{T}_k(p)$ and $\widehat{T}_k^+(p)$. Then CP-CPA plays

$$p_t \in \arg\max_{p\in\mathcal{P}} \widetilde{W}_t(p), \quad \widetilde{W}_t(p) := \sum_{k=1}^{mn} \Big(\widehat{T}_k^+(p)\,(p - c_{(k)}) - \gamma_{(k)}\Big)_+ \qquad \text{(optimistic profit)} \tag{12}$$

$$\widehat{K}_t := \#\big\{k :\ \widehat{T}_k(p_t)\,(p_t - c_{(k)}) - \gamma_{(k)} > 0\big\} \qquad \text{(estimated profit optimum)} \tag{13}$$

$$K_t^{\mathrm{cp}} := \min\big\{K :\ \widehat{\mathbb{P}}_t(N_t \le K) \ge 1 - \alpha_t^{\mathrm{eff}}\big\} \qquad \text{(conformal quantile)} \tag{14}$$

$$K_t := \max\big(\widehat{K}_t,\ K_t^{\mathrm{cp}}\big), \qquad \text{(provisioned inventory)} \tag{15}$$

where $\widehat{\mathbb{P}}_t$ is the Poisson–binomial law with parameters $\widehat{q}_t(p_t)$, the level $1-\alpha_t^{\mathrm{eff}}$ is clipped to $[0,1]$ inside the quantile (so $K_t^{\mathrm{cp}} = 0$ if $\alpha_t^{\mathrm{eff}} \ge 1$ and $K_t^{\mathrm{cp}} = n$ if $\alpha_t^{\mathrm{eff}} \le 0$), and $\alpha_{t+1}^{\mathrm{eff}} = \alpha_t^{\mathrm{eff}} + \eta(\alpha - \mathrm{err}_t)$, $\mathrm{err}_t = \mathbf{1}\{N_t > K_t\}$, is the (unclipped) ACI recursion of Appendix B. Rule (15) says that the conformal layer can only *add* safety inventory on top of the estimated profit optimum. These are exactly the rules of Algorithm 1 in the main text and of the released implementation.

**Assumption A.1** (Agnostic bounded noise)**.** *$\{\xi_{t,j}\}$ are i.i.d. with support in $[-c, c]$, $\mathbb{E}[\xi] = 0$, and an arbitrary (possibly non-Lipschitz, atomic) distribution. No density, smoothness, log-concavity, or parametric structure is assumed.*

**Assumption A.2** (Stochastic contexts)**.** *$\{x_{t,j}\}$ are i.i.d. from $\mathcal{D}_x$ supported on $\{\|x\|_2 \le B_x\}$ with $\lambda_{\min}(\mathbb{E}[xx^\top]) \ge \lambda_0 > 0$.*

**Assumption A.3** (Price buffer)**.** *There is $\kappa > 0$ with $\langle x, \theta^\star\rangle - c \ge \kappa$ and $\langle x, \theta^\star\rangle + c \le B$ a.s. for $x \sim \mathcal{D}_x$. The horizon is large enough that $8\Delta \le \min(\kappa, \gamma_{\min})$, where $\Delta$ is the Stage-1 target accuracy in* (16) *and $\gamma_{\min} = \min_i \gamma_i > 0$.*

**Assumption A.4** (Sorted supply side; perishable inventory)**.** *Supply costs are sorted, $c_1 \le \cdots \le c_m$, units are provisioned and served cheapest-first, and inventory is perishable (not carried across rounds).*

**Remark A.1** (What is, and is not, used from the supply side)**.** *The joint pricing–allocation literature additionally assumes that the population profit $W(p)$ is piecewise convex on the $O(mn)$ intervals induced by the supply costs. Our analysis does* not *use this: the pricing lemma (Lemma A.9) only needs (i) the greedy characterization* (10)*, which follows from sorted costs, and (ii) monotonicity of demand in the price. In fact, under Assumption A.1 $W_t$ need not be Lipschitz or piecewise convex at all: an atom of $\xi$ at $w$ produces a downward jump of $W_t$ at every price $u_{t,j} + w$, and these jump locations depend on the (unknown) $u_{t,j}$. This is precisely why the argument below works with grid slack and adjacent-jump budgets instead of Lipschitz constants.*

Assumptions A.1–A.3 are those of the contextual pricing model (CMRUP); Assumption A.4 is the sorted-costs half of the joint pricing–allocation model (cf. Remark A.1). The grid scale and stage lengths are

$$\Delta \asymp T^{-1/3}, \qquad M = \lceil c/\Delta \rceil = \Theta(T^{1/3}), \qquad T_1 = \lceil T^{2/3} \rceil, \qquad |\mathcal{P}| \le \lceil B/\Delta \rceil. \tag{16}$$

### A.2 Master regret decomposition

Fix a Stage-2 round $t$ and write the per-round regret as a telescoping sum through two intermediate inventories, $K_t^\star(p_t)$ (profit-optimal at the played price) and $\widehat{K}_t$ (its estimate, (13)):

$$\begin{aligned}\Pi_t^\star - F_t(p_t, K_t) = &\underbrace{\big(W_t(p_t^\star) - W_t(p_t)\big)}_{\text{(A) pricing regret}} + \underbrace{\big(F_t(p_t, K_t^\star(p_t)) - F_t(p_t, \widehat{K}_t)\big)}_{\text{(B) inventory-estimation gap}} \\ &+ \underbrace{\big(F_t(p_t, \widehat{K}_t) - F_t(p_t, K_t)\big)}_{\text{(C) conformal-safety cost}}.\end{aligned} \tag{17}$$

Terms (A) and (B) are non-negative by optimality of $p_t^\star$ and $K_t^\star(p_t)$; Term (C) is non-negative because $K_t \ge \widehat{K}_t$ and, by (13), every unit beyond $\widehat{K}_t$ has non-positive *estimated* marginal value (its true marginal value may be slightly positive, which only helps and is absorbed in Term (B)). The proof bounds $\sum_t$(A) and $\sum_t$(B) by $\widetilde{O}(T^{2/3})$ and identifies $\sum_t$(C) with the holding cost of the safety inventory.

### A.3 SUPPORTING LEMMAS

#### A.3.1 TWO FACTS ABOUT POISSON–BINOMIAL TAILS

**Lemma A.1** (Coupling bound for Poisson–binomial laws). *Let $q, q' \in [0,1]^n$ and let $N, N'$ be Poisson–binomial with parameters $q, q'$. Then*

$$\sup_k \big|\mathbb{P}(N \geq k) - \mathbb{P}(N' \geq k)\big| \;\leq\; \mathrm{TV}\big(\mathcal{L}(N), \mathcal{L}(N')\big) \;\leq\; \sum_{j=1}^{n} |q_j - q'_j|.$$

*If moreover $q \leq q'$ coordinatewise, then $\mathbb{P}(N \geq k) \leq \mathbb{P}(N' \geq k)$ for all $k$.*

*Proof.* Let $U_1, \ldots, U_n$ be i.i.d. $\mathrm{Unif}[0,1]$ and set $B_j = \mathbf{1}\{U_j \leq q_j\}$, $B'_j = \mathbf{1}\{U_j \leq q'_j\}$, so that $N := \sum_j B_j$ and $N' := \sum_j B'_j$ have the two laws. Then $\mathbb{P}(N \neq N') \leq \mathbb{P}(\exists j : B_j \neq B'_j) \leq \sum_j \mathbb{P}(B_j \neq B'_j) = \sum_j |q_j - q'_j|$, and $\mathrm{TV} \leq \mathbb{P}(N \neq N')$ for any coupling; the sup over tail events is at most the total variation. If $q \leq q'$ then $B_j \leq B'_j$ for every $j$, hence $N \leq N'$ pointwise and the tails are ordered. □

**Lemma A.2** (Threshold loss). *Let $g_1 \geq g_2 \geq \cdots \geq g_L$ and $\widehat{g}_1 \geq \cdots \geq \widehat{g}_L$ be two non-increasing sequences with $\max_k |g_k - \widehat{g}_k| \leq \eta$, and let $K^\star = \#\{k : g_k > 0\}$, $\widehat{K} = \#\{k : \widehat{g}_k > 0\}$. Then $0 \leq \sum_{k \leq K^\star} g_k - \sum_{k \leq \widehat{K}} g_k \leq L\eta$.*

*Proof.* If $\widehat{K} < K^\star$, the loss is $\sum_{k=\widehat{K}+1}^{K^\star} g_k$; for these $k$, $g_k > 0 \geq \widehat{g}_k$ so $0 < g_k \leq \eta$. If $\widehat{K} > K^\star$, the loss is $\sum_{k=K^\star+1}^{\widehat{K}} (-g_k)$ and for these $k$, $\widehat{g}_k > 0 \geq g_k$ so $0 \leq -g_k \leq \eta$. Either way at most $L$ terms, each in $[0, \eta]$. □

#### A.3.2 STAGE 1: PARAMETER RECOVERY

**Lemma A.3** (Censored linear signal). *Under Assumptions A.1–A.2, if in Stage 1 the price is $p \sim \mathrm{Unif}[0, B]$ drawn independently of $(x, \xi)$ and $v \in [0, B]$ (Assumption A.3), then $z_{t,j} := B\, o_{t,j}$ satisfies $\mathbb{E}[z_{t,j} \mid x_{t,j}] = \langle x_{t,j}, \theta^\star \rangle$.*

*Proof.* $\mathbb{E}[o_{t,j} \mid x, v] = \mathbb{P}(p \leq v \mid v) = v/B$ for $v \in [0, B]$, hence $\mathbb{E}[B\, o_{t,j} \mid x] = \mathbb{E}[v \mid x] = \langle x, \theta^\star \rangle$ using $\mathbb{E}[\xi] = 0$ and the independence of $p$. □

**Lemma A.4** (Parameter estimation). *Let $\widehat{\theta}$ be the ridge least-squares estimator on the $nT_1$ Stage-1 samples $\{(x_{t,j}, z_{t,j})\}$. Under Assumptions A.1–A.2 and (16), with probability at least $1 - T^{-2}$,*

$$\max_{t,j} \big|\langle x_{t,j}, \widehat{\theta} \rangle - u_{t,j}\big| \;\leq\; \Delta \;=\; O\Big(\tfrac{BB_x}{\lambda_0} \sqrt{\tfrac{d \log T}{nT_1}}\Big) \;=\; O\big(T^{-1/3}\big).$$

*Proof.* The design has bounded second moment with $\lambda_{\min} \geq \lambda_0$ (Assumption A.2); the noise $z - \mathbb{E}[z \mid x]$ is bounded in $[-B, B]$ and, by Lemma A.3, centred. Standard bounded-design ridge concentration gives $\|\widehat{\theta} - \theta^\star\|_2 \lesssim (BB_x/\lambda_0)\sqrt{d \log T/(nT_1)}$ w.p. $1 - T^{-2}$ (union bound over the $nT$ contexts is absorbed in the $\log T$); multiplying by $\|x\|_2 \leq B_x$ and substituting $T_1 = \lceil T^{2/3} \rceil$ yields the claim. □

Denote by $\mathcal{E}_\theta$ the event of Lemma A.4, on which $|\widehat{u}_{t,j} - u_{t,j}| \leq \Delta$ for all $t, j$.

#### A.3.3 CELL CONTAINMENT, ADJACENT JUMPS, AND UNIFORM OPTIMISM

Every censored observation $(p, \widehat{u}_{t,j}, o_{t,j})$ is assigned to the grid cell $k = k_j(p)$ of its *estimated* residual $p - \widehat{u}_{t,j} \in (w_{k-1}, w_k]$. Let $N_k(t)$ be the number of observations assigned to cell $k$ before round $t$, $\widehat{S}(w_k)$ their empirical purchase frequency, and

$$b_k(t) := \sqrt{\frac{5 \log(nT)}{\max(1, N_k(t))}}. \tag{18}$$

**Lemma A.5** (Cell containment). *On $\mathcal{E}_\theta$, an observation assigned to cell $k$ has* true *residual $r = p - u_{t,j} \in [w_{k-2}, w_{k+1}]$ (with $w_{-1} := w_0$, $w_{M+1} := w_M$), hence its conditional purchase probability $S(r)$ lies in $[S(w_{k+1}), S(w_{k-2})]$. In particular, all observations assigned to cell $k$, and the point estimate $\widehat{q}_{t,j}(p)$ of any consumer assigned to cell $k$, refer to survival values within the window $J_k := S(w_{k-2}) - S(w_{k+1})$ of each other.*

*Proof.* $|\widehat{u} - u| \le \Delta$ and $p - \widehat{u} \in (w_{k-1}, w_k]$ give $r \in (w_{k-1} - \Delta, w_k + \Delta]$; since the mesh is $2\Delta$, $w_{k-1} - \Delta = w_k - 3\Delta \ge w_{k-2}$ and $w_k + \Delta \le w_{k+1}$. Monotonicity of $S$ gives the survival window. (This is the mechanism that survives atoms: a residual error of size $\Delta$ cannot move the conditional probability outside a window of three adjacent grid cells.) □

We call $J_k$ the *adjacent jump* at cell $k$. Since every increment of $S$ across the grid is counted in at most three windows, $\sum_k J_k \le 3$.

**Lemma A.6** (Uniform optimism). *With probability at least $1 - T^{-2}$, simultaneously for all rounds $t$ and all cells $k$ with $N_k(t) \ge 1$,*

$$\left|\widehat{S}(w_k) - \bar{S}_k(t)\right| \le b_k(t), \qquad \bar{S}_k(t) := \frac{1}{N_k(t)} \sum_{(s,j)\ \text{assigned to } k,\ s<t} S(p_s - u_{s,j}).$$

*Consequently, on $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$ (the events of Lemmas A.4 and A.6), for every candidate price $p$ and consumer $j$, with $k = k_j(p)$,*

$$q_{t,j}(p) - J_k \;\le\; \widehat{q}^{+}_{t,j}(p) \;\le\; q_{t,j}(p) + J_k + 2b_k(t), \qquad \left|\widehat{q}_{t,j}(p) - q_{t,j}(p)\right| \le J_k + b_k(t). \tag{19}$$

*Proof. Filtration.* Let $\mathcal{F}_t$ be the $\sigma$-field generated by everything revealed up to the end of round $t$, and $\mathcal{F}_t^- := \mathcal{F}_{t-1} \vee \sigma(x_t)$. The price $p_t$ (hence every cell assignment $k_j(p_t)$) is $\mathcal{F}_t^-$-measurable, while, conditionally on $\mathcal{F}_t^-$, the indicators $o_{t,1}, \dots, o_{t,n}$ are independent with means $S(p_t - u_{t,j})$ (Assumption A.1: the $\xi_{t,j}$ are i.i.d. and independent of the past). Refine the filtration within each round by revealing $o_{t,1}, o_{t,2}, \dots$ one at a time; by conditional independence the mean of $o_{t,j}$ given the refined past is still $S(p_t - u_{t,j})$. Hence, for each fixed cell $k$, if we enumerate the observations assigned to $k$ in this order as $s = 1, 2, \dots$, the centred variables $\xi_s^{(k)} := o - S(p - u)$ form a martingale-difference sequence with $|\xi_s^{(k)}| \le 1$ with respect to the refined filtration (the event "the $s$-th observation is assigned to $k$" is a stopping time, and optional sampling preserves the martingale-difference property). Note that this holds for *whatever* adaptive rule selects $p_t$, in particular it does not require the price to be independent of the past.

*Concentration.* By Azuma–Hoeffding, for fixed $k$ and fixed count $N$, $\mathbb{P}\big(|\sum_{s \le N} \xi_s^{(k)}| \ge \sqrt{2N \log(2/\delta)}\big) \le \delta$. Taking $\delta = (nT)^{-4}$ and a union bound over $k \le M \le T$ and $N \le nT$ gives, with probability $\ge 1 - T^{-2}$, $|\sum_{s \le N} \xi_s^{(k)}| \le \sqrt{2N \cdot 5 \log(nT)}$ for all $(k, N)$ simultaneously, i.e. $|\widehat{S}(w_k) - \bar{S}_k(t)| \le b_k(t)$ with $b_k$ as in (18) (we absorbed $\sqrt{2}$ into the constant 5).

*Optimism.* On $\mathcal{E}_\theta$, Lemma A.5 places both $\bar{S}_k(t)$ and $q_{t,j}(p) = S(p - u_{t,j})$ in $[S(w_{k+1}), S(w_{k-2})]$, so $|\bar{S}_k(t) - q_{t,j}(p)| \le J_k$. Combining with the concentration bound gives $|\widehat{q}_{t,j}(p) - q_{t,j}(p)| \le J_k + b_k(t)$ and, adding $b_k(t)$, the two-sided bound (19) for $\widehat{q}^{+}$ (the truncation at 1 can only move $\widehat{q}^{+}$ towards $q \le 1$). □

**Lemma A.7** (Confidence budget). *Let $\mathcal{B}_T := \sum_{t \in \text{Stage 2}} \sum_{j=1}^{n} b_{k_j(p_t)}(t)$ be the confidence budget accumulated at the played prices. Then, deterministically, $\mathcal{B}_T \le 2\sqrt{5\, M\, nT \log(nT)} = \widetilde{O}(\sqrt{n}\, T^{2/3})$.*

*Proof.* Each played consumer contributes one observation to its cell, so the sum over all observations of cell $k$ is $\sum_{s=1}^{N_k(T)} \sqrt{5 \log(nT)/s} \le 2\sqrt{5 N_k(T) \log(nT)}$; summing over $k$ and using Cauchy–Schwarz with $\sum_k N_k(T) \le nT$ gives $\mathcal{B}_T \le 2\sqrt{5 \log(nT)}\sqrt{M \cdot nT}$, and $M = \Theta(T^{1/3})$ yields $\widetilde{O}(\sqrt{n}\, T^{2/3})$. □

The second budget concerns the adjacent jumps at the played prices,

$$\mathcal{J}_T := \sum_{t \in \text{Stage 2}} \sum_{j=1}^{n} J_{k_j(p_t)}. \tag{20}$$

Its size is a property of how often the shared price places a consumer's estimated residual within three cells of a region where $S$ has a large increment.

**Assumption A.5** (Adjacent-jump budget)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\text{conf}}$, $\mathcal{J}_T \le C_J\, n\, T^{2/3} \log T$ for a constant $C_J$ not depending on $T$.*

**Lemma A.8** (When the jump budget holds)**.** *Assumption A.5 holds in each of the following cases.*

*(i) $S$ is $L_S$-Lipschitz on $[-c, c]$ (equivalently, $\xi$ has a density bounded by $L_S$). Then $J_k \le 6\Delta L_S$ for every $k$ and $\mathcal{J}_T \le 6L_S\Delta\, nT = O(L_S n T^{2/3})$ deterministically.*

*(ii) $n = 1$ and prices are chosen by the CMRUP conservative-markdown and one-step redirect rules (Xu & Wang, 2026). Then a cell with $J_k > 2\Delta$ is played only while $b_k(t) > \Delta$, i.e. at most $5 \log T/\Delta^2$ times, after which the redirect moves the played residual one cell away from the window; hence $\mathcal{J}_T \le \sum_k J_k \cdot 5 \log T/\Delta^2 + 2\Delta T \le 15 \log T/\Delta^2 + 2\Delta T = O(T^{2/3} \log T)$.*

*(iii) $\xi$ has finitely many atoms $a_1, \dots, a_A$ of total mass $\bar{a}$ plus an $L_S$-Lipschitz part, and the shared price places some consumer's estimated residual within $3\Delta$ of an atom in at most $\nu_T$ rounds. Then $\mathcal{J}_T \le 6L_S\Delta nT + \bar{a}\, n\, \nu_T$, so Assumption A.5 holds whenever $\nu_T = O(T^{2/3} \log T)$.*

*Proof.* (i) is immediate from $J_k = S(w_{k-2}) - S(w_{k+1})$ and $w_{k+1} - w_{k-2} = 6\Delta$. (ii) is the CMRUP counting argument, restated with our constants; we refer to Xu & Wang (2026) for the redirect mechanics. (iii) splits $J_k$ into its Lipschitz part, bounded as in (i), and its atomic part, which is non-zero only in the $\nu_T$ rounds where an atom lies inside some played window. □

**Remark A.2** (Scope of Assumption A.5)**.** *For a single consumer the redirect makes the jump budget a theorem (case (ii)); with $n$ consumers sharing one price the redirect of one consumer moves the residuals of all others, and we do not know a rule that provably controls $\nu_T$ in case (iii) for all atomic noise distributions. The released implementation uses the plain optimistic argmax* (12) *without a redirect, and we therefore state the jump budget as an explicit assumption rather than claim a proof. It is fully rigorous under Lipschitz noise (case (i)), and the "cliff" experiments of Section 6 (atom mass $0.3$, $n = 10$) are the empirical check of case (iii): the fitted regret exponents there are $0.73$–$0.76$, consistent with a $T^{2/3}$ budget. The coverage guarantees of Theorems 2 and 3 do not depend on this assumption.*

### A.3.4 Pricing regret

**Lemma A.9** (Term (A): cumulative pricing regret)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\text{conf}}$, the optimistic-profit rule* (12) *satisfies*

$$\sum_{t \in \text{Stage 2}} \big(W_t(p_t^\star) - W_t(p_t)\big) \;\le\; 8\Delta\, mn\, T \;+\; mnB\,\big(2\mathcal{B}_T + \mathcal{J}_T\big),$$

*which is $\widetilde{O}(mn^2 B T^{2/3})$ under Assumption A.5.*

*Proof.* Fix a Stage-2 round $t$ and drop the index $t$. Write $L := mn$ for the number of provisionable units and recall $W(p) = \sum_{k \le L} (g_k(p))_+$, $\widetilde{W}(p) = \sum_{k \le L} (\widehat{g}_k^+(p))_+$ with $\widehat{g}_k^+(p) := \widehat{T}_k^+(p)(p - c_{(k)}) - \gamma_{(k)}$.

*Step 1 (the optimistic objective is accurate at the played price).* For $x \mapsto (x(p-c) - \gamma)_+$ with $p \in [0, B]$, the map is non-decreasing and $B$-Lipschitz in $x \in [0, 1]$. The upper half of (19) and Lemma A.1 give $\widehat{T}_k^+(p) \le \bar{T}_k(p) + \sum_j \big(J_{k_j(p)} + 2b_{k_j(p)}\big)$ for every $k$, hence for every $p \in \mathcal{P}$

$$\widetilde{W}(p) \;\le\; W(p) + L\, B \sum_j \big(J_{k_j(p)} + 2b_{k_j(p)}\big). \tag{21}$$

*Step 2 (a comparator price at which the optimistic objective is provably optimistic, without paying any jump).* Let $\tilde{p} \in \mathcal{P}$ with $p^\star - 8\Delta \le \tilde{p} \le p^\star - 6\Delta$ (such a grid price exists since the mesh of $\mathcal{P}$ is $\le 2\Delta$). Fix a consumer $j$ and let $k = k_j(\tilde{p})$, so $\tilde{p} - \widehat{u}_j \in (w_{k-1}, w_k]$. On $\mathcal{E}_\theta$,

$$w_{k+1} = w_{k-1} + 4\Delta < \tilde{p} - \widehat{u}_j + 4\Delta \le (p^\star - u_j) + \Delta - 6\Delta + 4\Delta < p^\star - u_j,$$

so by monotonicity of $S$ and Lemma A.5 (the window of cell $k$ has lower end $S(w_{k+1})$, and $\bar{S}_k \ge S(w_{k+1})$),

$$\widehat{q}_j^{+}(\tilde{p}) \;\ge\; \min\{1, \bar{S}_k(t)\} \;\ge\; S(w_{k+1}) \;\ge\; S(p^\star - u_j) = q_j(p^\star).$$

In words: at the slightly lower comparator price, every consumer's optimistic buy-probability dominates her *true* buy-probability at the optimal price, because the whole containment window of her cell lies below the optimal residual. No adjacent jump enters. By the monotone part of Lemma A.1, $\widehat{T}_k^{+}(\tilde{p}) \ge \bar{T}_k(p^\star)$ for all $k$. Every unit $k \le K^\star := K^\star(p^\star)$ has $g_k(p^\star) > 0$, hence $p^\star - c_{(k)} > \gamma_{(k)} \ge \gamma_{\min} \ge 8\Delta$ (Assumption A.3) and $\tilde{p} - c_{(k)} \ge 0$. Therefore

$$\begin{aligned}\widetilde{W}(\tilde{p}) \;&\ge\; \sum_{k \le K^\star} \Big(\widehat{T}_k^{+}(\tilde{p})(\tilde{p} - c_{(k)}) - \gamma_{(k)}\Big) \;\ge\; \sum_{k \le K^\star} \Big(\bar{T}_k(p^\star)(\tilde{p} - c_{(k)}) - \gamma_{(k)}\Big) \\ &\ge\; W(p^\star) - 8\Delta K^\star \;\ge\; W(p^\star) - 8\Delta L.\end{aligned}$$

No continuity of $W$ is used: only that lowering the price never loses demand.

*Step 3 (assembly).* Using optimality of $p_t$ for $\widetilde{W}$ over $\mathcal{P}$,

$$\begin{aligned}W(p^\star) - W(p_t) &\le \big[W(p^\star) - \widetilde{W}(\tilde{p})\big] + \big[\widetilde{W}(\tilde{p}) - \widetilde{W}(p_t)\big] + \big[\widetilde{W}(p_t) - W(p_t)\big] \\ &\le 8\Delta L + 0 + LB \sum_j \big(J_{k_j(p_t)} + 2b_{k_j(p_t)}\big),\end{aligned}$$

by Step 2 and (21) at $p_t$. Only the *played* price's confidence radii and adjacent jumps appear. Summing over the at most $T$ Stage-2 rounds gives the displayed bound; Lemma A.7, Assumption A.5 and $\Delta \asymp T^{-1/3}$ make all three contributions $\widetilde{O}(mn^2BT^{2/3})$. □

**Remark A.3** (Why optimistic *profit*, not optimistic revenue)**.** *If the pricing rule maximized the optimistic* revenue $\sum_j p\,\widehat{q}_j^{+}(p)$ *instead of* (12)*, Step 3 would compare $p_t$ to the revenue-optimal price, and the difference between the revenue- and profit-optimal prices is a constant that does not vanish with $T$ (under* $\mathrm{Unif}[-c, c]$ *noise and a single supplier with cost $c_1$, e.g., the revenue-optimal residual is $(c-u)/2$ while the profit-optimal one is $(c-u+c_1)/2$). Term (A) would then be $\Theta(T)$. The objective in* (12) *is what the released implementation optimizes.*

### A.3.5 Inventory-estimation gap

**Lemma A.10** (Term (B): inventory-estimation gap)**.** *On $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$,*

$$\sum_{t \in \textit{Stage 2}} \big(F_t(p_t, K_t^\star(p_t)) - F_t(p_t, \widehat{K}_t)\big) \;\le\; mnB\,(\mathcal{B}_T + \mathcal{J}_T),$$

*which is $\widetilde{O}(mn^2BT^{2/3})$ under Assumption A.5.*

*Proof.* Fix $t$ and write $g_k := g_{t,k}(p_t)$ and $\widehat{g}_k := \widehat{T}_k(p_t)(p_t - c_{(k)}) - \gamma_{(k)}$, both non-increasing in $k$ (Poisson–binomial tails are non-increasing in $k$, and $c_{(k)}$ is non-decreasing). By Lemma A.1 and the second half of (19), $\max_k |\widehat{g}_k - g_k| \le B \sum_j |\widehat{q}_{t,j}(p_t) - q_{t,j}(p_t)| \le B \sum_j (J_{k_j(p_t)} + b_{k_j(p_t)}(t)) =: \eta_t$. Since $K_t^\star(p_t) = \#\{k : g_k > 0\}$ and $\widehat{K}_t = \#\{k : \widehat{g}_k > 0\}$, Lemma A.2 with $L = mn$ bounds the per-round loss by $mn\,\eta_t$. Summing over Stage 2 gives $mnB(\mathcal{B}_T + \mathcal{J}_T)$. □

## A.4 The conformal-safety term

Term (C) is where CP-CPA differs from an unconstrained profit maximizer, and it is *not* in general $\widetilde{O}(T^{2/3})$: if the coverage target $\alpha$ forces more inventory than profit maximization would provision, every coverage-feasible policy pays a per-round holding cost, and so does the clairvoyant constrained policy. We therefore (i) bound Term (C) exactly by the holding cost of the safety inventory, (ii) show that this cost is unavoidable against the unconstrained benchmark whenever the constraint binds, and (iii) show that it is $\widetilde{O}(T^{2/3})$ whenever the constraint is slack by a margin.

**Lemma A.11** (Term (C) equals the holding cost of the safety inventory)**.** *Deterministically, for every Stage-2 round,*

$$0 \;\le\; F_t(p_t, \widehat{K}_t) - F_t(p_t, K_t) \;\le\; \bar{\gamma}_t \,(K_t^{\mathrm{cp}} - \widehat{K}_t)_+, \qquad \bar{\gamma}_t := \gamma_{\max} + \big(c_m - p_t\big)_+ \;\le\; \gamma_{\max} + c_m,$$

*where $\gamma_{\max} = \max_i \gamma_i$; in particular $\bar{\gamma}_t = \gamma_{\max}$ whenever the posted price covers the most expensive supplier. Consequently*

$$\sum_{t \in \textit{Stage 2}} \big(F_t(p_t, \widehat{K}_t) - F_t(p_t, K_t)\big) \;\le\; \gamma_{\max}\, D_T^+ \;+\; c_m \sum_t \mathbf{1}\{p_t < c_m\}\,(K_t^{\mathrm{cp}} - \widehat{K}_t)_+,$$

*where $D_T^+ := \sum_{t \in \textit{Stage 2}} \big(K_t^{\mathrm{cp}} - \widehat{K}_t\big)_+$.*

*Proof.* By (9) and (15), $F_t(p_t, \widehat{K}_t) - F_t(p_t, K_t) = -\sum_{k=\widehat{K}_t+1}^{K_t} g_{t,k}(p_t)$ and $K_t - \widehat{K}_t = (K_t^{\mathrm{cp}} - \widehat{K}_t)_+$. Each summand is $-g_{t,k}(p_t) = \gamma_{(k)} + \bar{T}_k(p_t)\,(c_{(k)} - p_t) \le \gamma_{(k)} + (c_{(k)} - p_t)_+ \le \bar{\gamma}_t$. Non-negativity was argued after (17). □

In the experiments the posted price always exceeds $c_m$ in Stage 2 (the optimistic profit of any price below $c_m$ is dominated), so we write the bound as $\gamma_{\max} D_T^+$ below and keep in mind the degenerate correction.

The quantity $D_T^+$ is the total *safety inventory*: the number of units that the conformal layer adds on top of the estimated profit optimum. It is an observable quantity, and it is reported in the experiments.

**Proposition A.1** (Linear regret is unavoidable when the constraint binds)**.** *Fix a distribution of $(x_t, \xi_t)$ satisfying Assumptions A.1–A.4. For a price $p$ and context $x$, let $\Pi^\alpha(x) := \max_p \max\{F(p,K) : \mathbb{P}(N(p) > K \mid x) \le \alpha\}$ be the best expected profit compatible with a per-round stockout probability of at most $\alpha$, and let $\rho^\alpha := \mathbb{E}_x[\Pi^\star(x) - \Pi^\alpha(x)] \ge 0$ be the expected* constraint gap*. Then any policy whose average stockout rate satisfies $\mathbb{E}\big[\frac{1}{T}\sum_t \mathrm{err}_t\big] \le \alpha + \epsilon_T$ incurs*

$$\mathbb{E}[R_T] \;\ge\; \rho^\alpha\, T \;-\; \Pi_{\max}\, \epsilon_T\, T\, \frac{1}{\alpha' - \alpha} \quad \textit{for every } \alpha' \in (\alpha, 1],$$

*where $\Pi_{\max} = nB$ bounds the per-round profit; in particular $\mathbb{E}[R_T] = \Omega(T)$ whenever $\rho^\alpha > 0$ and $\epsilon_T \to 0$. Conversely, the constrained clairvoyant that plays the maximizer in $\Pi^\alpha(x_t)$ each round is $\alpha$-feasible and has regret exactly $\rho^\alpha T$ in expectation.*

*Proof.* Let $\beta_t := \mathbb{P}(\mathrm{err}_t = 1 \mid x_t, p_t, K_t)$ be the conditional stockout probability of the policy at round $t$. On rounds with $\beta_t \le \alpha'$ the policy's expected profit is at most $\Pi^{\alpha'}(x_t)$ by definition of the constrained optimum; on the remaining rounds it is at most $\Pi^\star(x_t)$. Hence $\mathbb{E}[R_T] \ge \sum_t \mathbb{E}\big[(\Pi^\star(x_t) - \Pi^{\alpha'}(x_t))\mathbf{1}\{\beta_t \le \alpha'\}\big] \ge \rho^{\alpha'} T - \Pi_{\max}\mathbb{E}\big[\#\{t : \beta_t > \alpha'\}\big]$. By Markov's inequality applied to $\frac{1}{T}\sum_t \beta_t$, whose expectation equals $\mathbb{E}[\frac{1}{T}\sum_t \mathrm{err}_t] \le \alpha + \epsilon_T$, the expected number of rounds with $\beta_t > \alpha'$ is at most $T(\alpha + \epsilon_T)/\alpha'$; a sharper count follows from $\sum_t \beta_t \ge \alpha' \#\{t : \beta_t > \alpha'\}$, which gives $\mathbb{E}\#\{t : \beta_t > \alpha'\} \le T(\alpha + \epsilon_T)/\alpha'$. Letting $\alpha' \downarrow \alpha$ and using that $\rho^{\alpha'} \to \rho^\alpha$ from below (the constrained value is continuous from the right in $\alpha$ for the finitely supported $N$) yields, after rearranging, the displayed inequality with the stated constant; the second claim is immediate from the definitions. □

**Remark A.4** (Reading Proposition A.1)**.** *The proposition separates two regimes. If the profit-optimal decision already stocks out with probability $\le \alpha$ ("slack" regime, $\rho^\alpha = 0$), then the coverage constraint costs nothing in principle and Lemma A.12 below shows that* CP-CPA *pays only $\widetilde{O}(T^{2/3})$ for it. If profit maximization would stock out more often than $\alpha$ ("binding" regime, $\rho^\alpha > 0$), then every coverage-feasible algorithm, including the constrained clairvoyant, has linear regret against the* unconstrained *clairvoyant; the meaningful comparison is then to the constrained clairvoyant, and Term (C) is the price of safety, reported explicitly as $\gamma_{\max} D_T^+$. In the real-data experiment (Section 6.6) the target binds, and this is exactly the regret excess that is observed.*

**Assumption A.6** (Slack coverage target with margin)**.** *There are $\rho, \rho' > 0$ such that, for every context $x$ in the support of $\mathcal{D}_x$ and every $p \in \mathcal{P}$, every unit up to the $(1 - \alpha + \rho)$-quantile of demand has marginal profit at least $\rho'$: $g_k(p) \ge \rho'$ for all $k \le Q_{x,p}(1 - \alpha + \rho) := \min\{K : \mathbb{P}(N(p) \le K \mid x) \ge 1 - \alpha + \rho\}$.*

In words: the units needed to reach a stockout probability of $\alpha - \rho$ are all strictly profitable, so a profit maximizer over-covers the target by a margin. This is the regime of the synthetic benchmark in Section 6 (where the observed stockout rate of the profit-optimal inventory is 2–4% against $\alpha = 10\%$).

**Lemma A.12** (Term (C) in the slack regime)**.** *Under Assumption A.6, on $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$,*

$$\mathbb{E}\big[D_T^+\big] \;\le\; n\Big(1+\frac{8}{\rho^2}\Big) \;+\; \frac{10\,nB}{\rho\,\rho'}\,\big(\mathcal{B}_T+\mathcal{J}_T\big) \;=\; O\Big(\frac{n}{\rho^2}\Big) + \widetilde{O}\Big(\frac{n^2B}{\rho\rho'}\,T^{2/3}\Big),$$

*uniformly in the ACI step $\eta \in (0,1]$. Hence $\gamma_{\max}\mathbb{E}[D_T^+] = \widetilde{O}(T^{2/3})$ for fixed $(\rho,\rho')$.*

*Proof.* Call a Stage-2 round $t$ *good* if $\eta_t := B\sum_j (J_{k_j(p_t)} + b_{k_j(p_t)}(t)) < \rho'/2$ and *bad* otherwise. The number of bad rounds is at most $\sum_t \eta_t/(\rho'/2) = 2B(\mathcal{B}_T+\mathcal{J}_T)/\rho'$, and on every round $(K_t^{\mathrm{cp}} - \widehat{K}_t)_+ \le n$.

*Good rounds over-cover.* On a good round, for every $k \le Q_t := Q_{x_t,p_t}(1-\alpha+\rho)$ we have $\widehat{g}_k \ge g_k - \eta_t > \rho'/2 > 0$ (proof of Lemma A.10), hence $\widehat{K}_t \ge Q_t$ and $\mathbb{P}(N_t > \widehat{K}_t \mid x_t, p_t) \le \alpha-\rho$. Since $K_t \ge \widehat{K}_t$, the stockout probability $\beta_t := \mathbb{P}(\mathrm{err}_t = 1 \mid \mathcal{F}_t^-, p_t, K_t) \le \alpha - \rho$ on good rounds, and $\beta_t \le 1$ always.

*When is safety inventory added?* By Lemma A.1, $\widehat{\mathbb{P}}_t(N \le \widehat{K}_t) \ge \mathbb{P}(N_t \le \widehat{K}_t \mid x_t) - \eta_t/B \ge 1 - \alpha+\rho-\rho'/(2B)$ on a good round; we may assume $\rho' \le B\rho$ (otherwise shrink $\rho'$), so $\widehat{\mathbb{P}}_t(N \le \widehat{K}_t) \ge 1-\alpha+\rho/2$. If $\alpha_t^{\mathrm{eff}} \ge \alpha - \rho/2$, then $1-\alpha_t^{\mathrm{eff}} \le 1-\alpha+\rho/2 \le \widehat{\mathbb{P}}_t(N \le \widehat{K}_t)$, so by (14) $K_t^{\mathrm{cp}} \le \widehat{K}_t$ and no safety inventory is added. Hence on good rounds $(K_t^{\mathrm{cp}} - \widehat{K}_t)_+ \le n\,\mathbf{1}\{\alpha_t^{\mathrm{eff}} < \alpha - \rho/2\}$.

*Occupation time below the threshold.* Write $\alpha_{t+1}^{\mathrm{eff}} - \alpha_t^{\mathrm{eff}} = \eta(\alpha - \beta_t) + \eta(\beta_t - \mathrm{err}_t)$. The second term is a martingale difference bounded by $\eta$ (given $\mathcal{F}_t^-$, $p_t$, $K_t$, the indicator $\mathrm{err}_t$ has mean $\beta_t$); the first (drift) is $\ge \eta\rho$ on good rounds and $\ge -\eta(1-\alpha) \ge -\eta$ on bad rounds. Let $t_0$ be the first Stage-2 round. Stage 1 provisions fully, so $\mathrm{err}_t = 0$ there and $\alpha_{t_0}^{\mathrm{eff}} \ge \alpha$. For $t \ge t_0$ let $s := t - t_0$ and let $N_t^{\mathrm{bad}}$ be the number of bad rounds in $[t_0, t)$. Then

$$\alpha_t^{\mathrm{eff}} \;\ge\; \alpha + \eta\rho\,(s - N_t^{\mathrm{bad}}) - \eta N_t^{\mathrm{bad}} - M_t \;=\; \alpha + \eta\rho s - \eta(1+\rho)N_t^{\mathrm{bad}} - M_t,$$

with $M_t := \sum_{t_0 \le s' < t} \eta(\mathrm{err}_{s'} - \beta_{s'})$ a martingale with increments in $[-\eta, \eta]$. Hence

$$\big\{\alpha_t^{\mathrm{eff}} < \alpha - \rho/2\big\} \;\subseteq\; \big\{s < 2(1+\rho)N_T^{\mathrm{bad}}/\rho\big\} \;\cup\; \big\{M_t > \eta\rho s/2\big\},$$

because on the complement of the first event $\eta(1+\rho)N_t^{\mathrm{bad}} \le \eta\rho s/2$, and then $\alpha_t^{\mathrm{eff}} < \alpha - \rho/2$ forces $M_t > \rho/2 + \eta\rho s/2 > \eta\rho s/2$. By Azuma–Hoeffding, $\mathbb{P}(M_t > \eta\rho s/2) \le \exp\big(-(\eta\rho s/2)^2/(2s\eta^2)\big) = e^{-\rho^2 s/8}$. Summing over $t$,

$$\mathbb{E}\Big[\sum_{t\ \mathrm{good}} \mathbf{1}\{\alpha_t^{\mathrm{eff}} < \alpha-\rho/2\}\Big] \;\le\; \frac{2(1+\rho)}{\rho}\,\mathbb{E}[N_T^{\mathrm{bad}}] + \sum_{s\ge 0} e^{-\rho^2 s/8} \;\le\; \frac{2(1+\rho)}{\rho}\,\mathbb{E}[N_T^{\mathrm{bad}}] + 1 + \frac{8}{\rho^2}.$$

*Assembly.* $\mathbb{E}[D_T^+] \le n\,\mathbb{E}[N_T^{\mathrm{bad}}] + n\,\mathbb{E}\big[\sum_{t\ \mathrm{good}} \mathbf{1}\{\alpha_t^{\mathrm{eff}} < \alpha - \rho/2\}\big]$. Substituting $\mathbb{E}[N_T^{\mathrm{bad}}] \le 2B(\mathcal{B}_T+\mathcal{J}_T)/\rho'$ and using $1 + 2(1+\rho)/\rho \le 5/\rho$ for $\rho \le 1$ gives the claim; Lemma A.7 and Assumption A.5 give the rate. The ACI step $\eta$ cancels in the Azuma exponent, which is why the bound is uniform in $\eta$. □

**Remark A.5** (What the slack lemma does and does not say)**.** *Lemma A.12 does not claim that $\alpha_t^{\mathrm{eff}}$ stays in a band around $\alpha$; in the slack regime it drifts* upward *(the algorithm stocks out less often than $\alpha$, so the ACI recursion keeps raising $\alpha_t^{\mathrm{eff}}$, and the unclipped recursion allows $\alpha_t^{\mathrm{eff}} > 1$, at which point $K_t^{\mathrm{cp}} = 0$). The lemma only needs that, apart from an $O(1/(\eta\rho) + 1/\rho^2)$ initial period and the $\widetilde{O}(T^{2/3})$ rounds on which the demand estimate is inaccurate, $\alpha_t^{\mathrm{eff}}$ stays above the level $\alpha - \rho/2$ at which the conformal quantile could exceed the estimated profit optimum. In the binding regime the same drift argument shows that $\alpha_t^{\mathrm{eff}}$ settles where the stockout rate equals $\alpha$ (Theorem 2), and $D_T^+$ is then linear in $T$, as it must be by Proposition A.1.*

### A.5 Proof of Theorem 1

**Lemma A.13** (Good event). *Let $\mathcal{G} := \mathcal{E}_\theta \cap \mathcal{E}_{\text{conf}}$. Then $\mathbb{P}(\mathcal{G}^c) \le 2T^{-2}$ and, on $\mathcal{G}^c$, the total regret is at most $T\Pi_{\max} = nBT$, so $\mathbb{E}[R_T \mathbf{1}\{\mathcal{G}^c\}] \le 2nB/T \le 2nB$.*

*Proof.* Union bound over Lemmas A.4 and A.6; the per-round regret is at most $\Pi_{\max} = nB$ since profits lie in $[-\sum_k (c_{(k)} + \gamma_{(k)}), nB]$ and the oracle's is non-negative. □

**Theorem 4** (Regret bound). *Under Assumptions A.1–A.4 and A.5 and the schedule* (16)*,* CP-CPA *(rules* (12)–(15)*) satisfies*

$$\mathbb{E}[R_T] \;\le\; \widetilde{O}\big(T^{2/3} \cdot \mathrm{poly}(m, n, d, B, B_x, 1/\lambda_0, 1/\kappa)\big) \;+\; \gamma_{\max}\, \mathbb{E}\big[D_T^+\big],$$

*where $D_T^+ = \sum_t (K_t^{\text{cp}} - \widehat{K}_t)_+$ is the total safety inventory added by the conformal layer. If in addition the coverage target is slack with margin (Assumption A.6), then $\gamma_{\max}\mathbb{E}[D_T^+] = O\big(\frac{n\gamma_{\max}}{\rho^2}\big) + \widetilde{O}\big(\frac{n^2 B \gamma_{\max}}{\rho\rho'} T^{2/3}\big)$ and hence $\mathbb{E}[R_T] = \widetilde{O}(T^{2/3})$. If the target binds ($\rho^\alpha > 0$ in Proposition A.1), then $\mathbb{E}[R_T] = \Omega(\rho^\alpha T)$ for* every *algorithm satisfying the coverage guarantee of Theorem 2, and the excess of* CP-CPA *over the constrained clairvoyant is again bounded by the $\widetilde{O}(T^{2/3})$ terms plus the safety inventory in excess of the constrained clairvoyant's.*

*Proof.* Split $\mathbb{E}[R_T] = \mathbb{E}[R_T \mathbf{1}\{\mathcal{G}\}] + \mathbb{E}[R_T \mathbf{1}\{\mathcal{G}^c\}]$; the second term is $O(nB)$ by Lemma A.13. On $\mathcal{G}$, Stage 1 lasts $T_1 = \lceil T^{2/3} \rceil$ rounds with per-round regret at most $\Pi_{\max} = nB$, contributing $O(nBT^{2/3})$. For Stage 2 apply (17) and Lemmas A.9, A.10, A.11:

$$\sum_{t \in \text{Stage 2}} \big(\Pi_t^\star - F_t(p_t, K_t)\big) \le \underbrace{8\Delta mnT + mnB(2\mathcal{B}_T + \mathcal{J}_T)}_{(A)} + \underbrace{mnB(\mathcal{B}_T + \mathcal{J}_T)}_{(B)} + \underbrace{\gamma_{\max} D_T^+}_{(C)}.$$

Lemma A.7 bounds $\mathcal{B}_T = \widetilde{O}(\sqrt{n}T^{2/3})$, Assumption A.5 bounds $\mathcal{J}_T = \widetilde{O}(nT^{2/3})$, and $\Delta \asymp T^{-1/3}$ gives $8\Delta mnT = O(mnT^{2/3})$; the $d, B, B_x, 1/\lambda_0, 1/\kappa$ dependence enters through the constant in $\Delta$ (Lemma A.4). This proves the first display. The slack-regime statement is Lemma A.12; the binding-regime statements are Proposition A.1 together with the observation that, relative to the constrained clairvoyant, Terms (A) and (B) are unchanged and Term (C) is reduced by the constrained clairvoyant's own safety inventory. □

**Corollary A.1** (Optimal $T$-dependence of the learning terms). *The exponent $2/3$ in the $\widetilde{O}(T^{2/3})$ part is unimprovable: the non-contextual lower bound $\Omega(T^{2/3})$ of Kleinberg–Leighton for agnostic posted-price learning embeds into our model with constant valuation, dummy covariates (Assumption A.2 is satisfied by independent dummies), a single supplier, and a coverage target that is slack (e.g. $\alpha = 1$), so no algorithm achieves $o(T^{2/3})$ regret in the agnostic non-Lipschitz regime even without a coverage constraint.*

## B Proof of Theorem 2

This appendix proves Theorem 2: under CP-CPA, the long-run *realized stockout rate* never exceeds the user-specified target $\alpha$ by more than an $O(1/(\eta T))$ term, and this holds *deterministically*, for every realization of demand, contexts, and estimation error. A second, conditional statement quantifies the per-round stockout probability in terms of the demand-estimation error and shows that the guarantee does not degenerate into over-provisioning when the target binds. We write the per-round stockout indicator as

$$\mathrm{err}_t \;:=\; \mathbf{1}\{N_t > K_t\}, \tag{22}$$

where $N_t$ is realized demand and $K_t$ the provisioned capacity, and the empirical stockout rate as $\widehat{\alpha}_T := \frac{1}{T}\sum_{t=1}^T \mathrm{err}_t$.

### B.1 INVENTORY LAYER AND CONFORMAL CALIBRATION

Fix a Stage-2 round $t$ with posted price $p_t$ and contexts $x_t$. Let $q_{t,j} := S(p_t - u_{t,j})$ be the true buy-probabilities, so that $N_t = \sum_j B_{t,j}$ with independent $B_{t,j} \sim \text{Bernoulli}(q_{t,j})$ is Poisson–binomial with parameters $q_t$; let $\widehat{q}_{t,j} = \widehat{S}(p_t - \widehat{u}_{t,j})$ be the point estimates of Appendix A and $\widehat{\mathbb{P}}_t$ the Poisson–binomial law with parameters $\widehat{q}_t$.

CP-CPA maintains an *Adaptive Conformal Inference* (ACI) state $\alpha_t^{\text{eff}} \in \mathbb{R}$ and provisions

$$\begin{aligned} K_t &:= \max\big(\widehat{K}_t,\ K_t^{\text{cp}}\big), \qquad K_t^{\text{cp}} := \widehat{Q}_t\big(\Pi_{[0,1]}(1-\alpha_t^{\text{eff}})\big), \\ \widehat{Q}_t(\beta) &:= \min\Big\{K \in \{0,\dots,n\} : \widehat{\mathbb{P}}_t(N_t \le K) \ge \beta\Big\}, \end{aligned} \tag{23}$$

with the convention $\widehat{Q}_t(1) := n$, where $\widehat{K}_t$ is the estimated profit-optimal inventory (eq. (13) in Appendix A) and $\Pi_{[0,1]}$ clips the level to $[0,1]$; thus $K_t^{\text{cp}} = n$ (full provisioning) when $\alpha_t^{\text{eff}} \le 0$ and $K_t^{\text{cp}} = 0$ when $\alpha_t^{\text{eff}} \ge 1$. The convention matters: the estimated Poisson–binomial law may have truncated support (some $\widehat{q}_{t,j} = 0$), in which case the literal $\min\{K : \widehat{\mathbb{P}}_t(N_t \le K) \ge 1\}$ would be smaller than $n$ and a stockout could occur while $\alpha_t^{\text{eff}} \le 0$, breaking Lemma B.1 below. The released implementation enforces the convention explicitly. After observing $\text{err}_t$, ACI updates, *without clipping*,

$$\alpha_{t+1}^{\text{eff}} \ = \ \alpha_t^{\text{eff}} + \eta\,(\alpha - \text{err}_t), \qquad \alpha_{t_0}^{\text{eff}} = \alpha, \tag{24}$$

where $t_0 = T_1 + 1$ is the first Stage-2 round; the recursion is not run in Stage 1, where the algorithm provisions fully and $\text{err}_t = 0$. This is the original recursion of Gibbs & Candès (2021); the implementation default is $\eta = 0.05$.

**Assumption B.1** (ACI step size)**.** *The step $\eta \in (0,1]$ is fixed.*

No relation between $\eta$ and $\alpha$ is needed.

### B.2 THE SATURATION PROPERTY AND THE TELESCOPING IDENTITY

**Lemma B.1** (Saturation keeps $\alpha^{\text{eff}}$ bounded below)**.** *For every $t \ge t_0$, $\alpha_t^{\text{eff}} \ \ge\ -\eta(1-\alpha) \ >\ -\eta$, deterministically.*

*Proof.* Induction on $t$. The claim holds at $t_0$. If $\alpha_t^{\text{eff}} \le 0$ then $K_t^{\text{cp}} = \widehat{Q}_t(1) = n \ge N_t$ by (23), so $K_t = n$, $\text{err}_t = 0$, and $\alpha_{t+1}^{\text{eff}} = \alpha_t^{\text{eff}} + \eta\alpha > \alpha_t^{\text{eff}}$. If $\alpha_t^{\text{eff}} > 0$ then $\alpha_{t+1}^{\text{eff}} \ge \alpha_t^{\text{eff}} - \eta(1-\alpha) > -\eta(1-\alpha)$. □

**Lemma B.2** (Telescoping identity and long-run stockout rate)**.** *For every $T \ge t_0$, deterministically,*

$$\sum_{t=t_0}^{T} \big(\text{err}_t - \alpha\big) \ = \ \frac{\alpha_{t_0}^{\text{eff}} - \alpha_{T+1}^{\text{eff}}}{\eta}, \tag{25}$$

*and consequently*

$$\frac{1}{T - t_0 + 1} \sum_{t=t_0}^{T} \text{err}_t \ \le\ \alpha + \frac{\alpha + \eta}{\eta\,(T - t_0 + 1)}. \tag{26}$$

*Proof.* Summing (24) over $t_0 \le t \le T$ gives $\alpha_{T+1}^{\text{eff}} - \alpha_{t_0}^{\text{eff}} = \eta \sum_t (\alpha - \text{err}_t)$, which is (25). Lemma B.1 gives $\alpha_{T+1}^{\text{eff}} \ge -\eta$, and $\alpha_{t_0}^{\text{eff}} = \alpha$, so the right-hand side of (25) is at most $(\alpha+\eta)/\eta$. □

**Remark B.1** (Why the update is not clipped)**.** *The clipped variant $\alpha_{t+1}^{\text{eff}} = \Pi_{[0,1]}(\alpha_t^{\text{eff}} + \eta(\alpha - \text{err}_t))$ breaks the exact identity* (25)*: every active clip at* 1 *discards a positive increment, and the discarded mass is not controlled without further argument. The unclipped recursion, with the level clipped* inside *the quantile as in* (23)*, is the form for which Gibbs & Candès (2021) prove their guarantee, and it is what the released implementation runs. Note that $\alpha_t^{\text{eff}}$ may exceed* 1 *(then $K_t^{\text{cp}} = 0$ and $K_t = \widehat{K}_t$); this is harmless for the upper bound* (26) *and is exactly what happens in the slack regime of Appendix A, Lemma A.12.*

### B.3 PER-ROUND STOCKOUT PROBABILITY

The identity gives the long-run rate without any reference to the quality of the demand estimate. The following two lemmas describe what happens *per round.*

**Lemma B.3** (Quantile miscoverage under the true law)**.** *If $K_t^{\mathrm{cp}}$ in* (23) *is computed from the* true *law of $N_t$ given $(x_t, p_t)$, then* $\mathbb{P}\big(N_t > K_t \mid x_t, p_t, \alpha_t^{\mathrm{eff}}\big) \le \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}})$.

*Proof.* $K_t \ge K_t^{\mathrm{cp}}$, so $\mathbb{P}(N_t > K_t) \le \mathbb{P}(N_t > K_t^{\mathrm{cp}}) = 1 - \mathbb{P}(N_t \le K_t^{\mathrm{cp}}) \le 1 - \Pi_{[0,1]}(1-\alpha_t^{\mathrm{eff}}) = \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}})$ by the definition of the quantile. □

**Lemma B.4** (Estimation-induced miscoverage)**.** *On the good event $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$ of Appendix A (Lemmas A.4 and A.6), for every Stage-2 round,*

$$\mathbb{P}\big(\mathrm{err}_t = 1 \mid \mathcal{F}_t^-, p_t, \alpha_t^{\mathrm{eff}}\big) \;\le\; \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}}) + \varepsilon_t, \quad \varepsilon_t := \sum_{j=1}^{n} \big|\widehat{q}_{t,j} - q_{t,j}\big| \le \sum_{j=1}^{n} \big(J_{k_j(p_t)} + b_{k_j(p_t)}(t)\big),$$

*with $J_k$ the adjacent jumps and $b_k(t)$ the confidence radii of Appendix A. Consequently* $\frac{1}{T}\sum_{t\in\text{Stage 2}} \varepsilon_t \le (\mathcal{B}_T + \mathcal{J}_T)/T = \widetilde{O}(nT^{-1/3})$ *under Assumption A.5.*

*Proof.* Let $\widehat{Q}_t$ and $Q_t$ be the estimated and true quantile functions. By Lemma A.1 (coupling bound for Poisson–binomial laws), $\big|\mathbb{P}(N_t > K) - \widehat{\mathbb{P}}_t(N_t > K)\big| \le \sum_j |\widehat{q}_{t,j} - q_{t,j}| = \varepsilon_t$ for every $K$. Apply this at $K = K_t^{\mathrm{cp}}$: by construction $\widehat{\mathbb{P}}_t(N_t > K_t^{\mathrm{cp}}) \le \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}})$, hence $\mathbb{P}(N_t > K_t^{\mathrm{cp}} \mid \cdot) \le \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}}) + \varepsilon_t$, and $K_t \ge K_t^{\mathrm{cp}}$ only lowers the left side. The bound on $\varepsilon_t$ is the second half of (19); the rate follows from Lemma A.7 and Assumption A.5. □

### B.4 STAGE 1

**Lemma B.5** (Stage-1 contribution)**.** *Stage 1 provisions $K_t = n \ge N_t$, so* $\mathrm{err}_t = 0$ *for $t \le T_1$ and*

$$\widehat{\alpha}_T = \frac{T - T_1}{T} \cdot \frac{1}{T - T_1} \sum_{t=t_0}^{T} \mathrm{err}_t \;\le\; \frac{1}{T - T_1} \sum_{t=t_0}^{T} \mathrm{err}_t.$$

### B.5 PROOF OF THEOREM 2

**Theorem 5** (Marginal stockout coverage)**.** *Under Assumption B.1, for every $T > T_1$ and every realization of contexts, demand, and estimation errors,* CP-CPA *satisfies*

$$\widehat{\alpha}_T \;\le\; \frac{1}{T - T_1} \sum_{t=t_0}^{T} \mathrm{err}_t \;\le\; \alpha + \frac{\alpha + \eta}{\eta\,(T - T_1)} \;=\; \alpha + O\Big(\frac{1}{\eta T}\Big). \tag{27}$$

*Moreover, on the good event of Appendix A and under Assumption A.5, the per-round conditional stockout probabilities satisfy* $\mathbb{P}(\mathrm{err}_t = 1 \mid \mathcal{F}_t^-, p_t, \alpha_t^{\mathrm{eff}}) \le \Pi_{[0,1]}(\alpha_t^{\mathrm{eff}}) + \varepsilon_t$ *with* $\frac{1}{T}\sum_t \varepsilon_t = \widetilde{O}(nT^{-1/3})$.

*Proof.* The first inequality in (27) is Lemma B.5, the second is (26). The conditional statement is Lemma B.4. □

**Corollary B.1** (Service-level guarantee)**.** *The realized fill rate satisfies* $1 - \widehat{\alpha}_T \ge 1 - \alpha - O(1/(\eta T))$ *deterministically:* CP-CPA *meets the target service level in the long run regardless of the demand model's bias, the noise distribution (atoms included), and the censoring of the pricing feedback.*

**Remark B.2** (One-sided by design; two-sided when the target binds)**.** *The coverage requirement is an inequality, and* (27) *is correspondingly one-sided: in the slack regime (Appendix A, Assumption A.6) the estimated profit optimum $\widehat{K}_t$ already stocks out less often than $\alpha$, $\alpha_t^{\mathrm{eff}}$ drifts upward, the conformal quantile becomes inactive, and $\widehat{\alpha}_T$ settles* below $\alpha$ *at the stockout rate of the profit-optimal inventory. This is the desired behaviour (there is nothing to pay for safety), and it is what the synthetic benchmark shows. When the target binds, the identity* (25) *also gives the matching lower*

*bound* $\frac{1}{T-T_1}\sum_t \mathrm{err}_t \geq \alpha - (\alpha_{T+1}^{\mathrm{eff}} - \alpha)/(\eta(T-T_1))$*, so the rate converges to* $\alpha$ *from both sides as long as* $\alpha_{T+1}^{\mathrm{eff}} = o(T)$*, which is the case whenever the conformal quantile is active a positive fraction of the time (each activation with a stockout lowers* $\alpha^{\mathrm{eff}}$*). The real-data experiment (Section 6.6) and the binding-regime synthetic runs are in this case.*

**Remark B.3** (Comparison to the plug-in baseline)**.** *The plug-in quantile baseline (*JOINT-NO-CP*) uses* $K_t = \max(\widehat{K}_t, \widehat{Q}_t(1-\alpha))$ *with the level frozen at* $1-\alpha$*. Lemma B.4 then gives* $\mathbb{P}(\mathrm{err}_t = 1 \mid \cdot) \leq \alpha + \varepsilon_t$ *with no mechanism to remove the bias* $\varepsilon_t$*: when the survival estimate is systematically optimistic (a cliff that the grid places on the wrong side), the baseline miscovers by a constant. In the real-data benchmark it stocks out at* $14$–$25\%$ *against* $\alpha = 10\%$*, whereas* (27) *caps* CP-CPA *at* $\alpha + O(1/(\eta T))$ *and the observed rate is* $10.2\%$ *at the longest horizon.*

**Remark B.4** (From marginal to conditional coverage)**.** *Theorem 5 is a statement about the marginal rate over rounds. A conditional guarantee* $\mathbb{P}(N_t > K_t \mid x_t \in \text{anchor } g) \leq \alpha + o(1)$ *for a finite family of anchors is obtained by running one ACI state per anchor; this is Theorem 3, proved in Appendix C with the same two ingredients (the telescoping identity per anchor and Lemma B.4).*

# C Proof of Theorem 3

This appendix proves the conditional version of the coverage guarantee (Theorem 3 in the main text). The statement is for a *finite anchor partition* of the context space, one ACI state per anchor, which is exactly what CP-CPA-WACI implements and what the experiments of Section 6.5 evaluate ($G = 2$ and $G = 4$ anchors). The proof has the same two ingredients as Theorem 2 (Appendix B): the telescoping identity, now applied to the subsequence of rounds that visit a given anchor, and the estimation-induced miscoverage bound (Lemma B.4). A kernel-smoothed version with a continuum of anchors is discussed, but not claimed, in Remark C.2.

Throughout we keep the notation of Appendices A and B: $\widehat{\mathbb{P}}_t$ is the plug-in Poisson–binomial law of the demand at $(x_t, p_t)$, $\widehat{K}_t$ the estimated profit-optimal inventory, $\mathrm{err}_t = \mathbf{1}\{N_t > K_t\}$, and $t_0 = T_1 + 1$ the first Stage-2 round.

## C.1 Anchor-wise conformal inventory layer

Let $g : \mathcal{X}^n \to \{1, \ldots, G\}$ be a fixed, measurable *anchor map* from the round context $x_t = (x_{t,1}, \ldots, x_{t,n})$ to one of $G$ anchors (segments). In the experiments $g$ is the deal-heavy/regular split ($G = 2$) or the promo $\times$ deal-share grid ($G = 4$); any partition may be used, including one obtained by binning a kernel density. Write $g_t := g(x_t)$ and let

$$\mathcal{T}_g := \{t \geq t_0 : \ g_t = g\}, \qquad V_g(T) := |\mathcal{T}_g \cap [t_0, T]|$$

be the rounds visiting anchor $g$ and their number. The seller maintains one ACI state per anchor, $\alpha_t^{(g)} \in \mathbb{R}$, initialised at $\alpha$ and updated only when the anchor is visited:

$$\alpha_{t+1}^{(g)} \ = \ \alpha_t^{(g)} + \eta\, \mathbf{1}\{g_t = g\}\,(\alpha - \mathrm{err}_t), \qquad \alpha_{t_0}^{(g)} = \alpha. \tag{28}$$

The Stage-2 inventory at round $t$ is that of Algorithm 1 with the marginal level replaced by the local one:

$$K_t = \max\big(\widehat{K}_t,\ K_t^{\mathrm{cp}}\big), \qquad K_t^{\mathrm{cp}} = \widehat{Q}_t\big(\Pi_{[0,1]}(1 - \alpha_t^{(g_t)})\big), \tag{29}$$

again with the convention $\widehat{Q}_t(1) := n$ of Appendix B. The anchor-conditional stockout rate after $T$ rounds is

$$\widehat{\alpha}_T^{(g)} := \frac{1}{V_g(T)} \sum_{t \in \mathcal{T}_g \cap [t_0, T]} \mathrm{err}_t \qquad (\text{defined when } V_g(T) \geq 1). \tag{30}$$

## C.2 Anchor-wise oracle quantile

**Lemma C.1** (Anchor-wise oracle coverage)**.** *If* $K_t^{\mathrm{cp}}$ *in* (29) *were computed from the* true *Poisson–binomial law of* $N_t$ *given* $(x_t, p_t)$*, then* $\mathbb{P}(N_t > K_t \mid x_t, p_t, \alpha_t^{(g_t)}) \leq \Pi_{[0,1]}(\alpha_t^{(g_t)})$.

*Proof.* Identical to Lemma B.3: $K_t \geq K_t^{\mathrm{cp}}$ and, by the definition of the integer quantile, $\mathbb{P}(N_t > K_t^{\mathrm{cp}} \mid \cdot) \leq 1 - \Pi_{[0,1]}(1 - \alpha_t^{(g_t)})$. The bound is pointwise in $x_t$ because the level $\alpha_t^{(g_t)}$ is a fixed number once $x_t$ (hence $g_t$) and the past are fixed. □

### C.3 Anchor-wise telescoping

**Lemma C.2** (Anchor-wise saturation and identity)**.** *For every anchor $g$ and every $T \geq t_0$, deterministically, $\alpha_{T+1}^{(g)} \geq -\eta(1-\alpha)$ and*

$$\sum_{t \in \mathcal{T}_g \cap [t_0, T]} \big(\mathrm{err}_t - \alpha\big) \;=\; \frac{\alpha_{t_0}^{(g)} - \alpha_{T+1}^{(g)}}{\eta} \;\leq\; \frac{\alpha + \eta}{\eta}. \tag{31}$$

*Consequently, whenever $V_g(T) \geq 1$,*

$$\widehat{\alpha}_T^{(g)} \;\leq\; \alpha + \frac{\alpha + \eta}{\eta\, V_g(T)}. \tag{32}$$

*Proof.* Restricted to $\mathcal{T}_g$, the recursion (28) is the unclipped ACI recursion (24) of Appendix B, run on the subsequence of rounds visiting $g$; on other rounds $\alpha^{(g)}$ does not move. If $\alpha_t^{(g)} \leq 0$ on a visiting round then $K_t^{\mathrm{cp}} = n$, $\mathrm{err}_t = 0$ and $\alpha^{(g)}$ increases; otherwise it decreases by at most $\eta(1-\alpha)$. This gives the lower bound as in Lemma B.1. Summing (28) over the visiting rounds gives the identity, and the lower bound on $\alpha_{T+1}^{(g)}$ gives the inequality; dividing by $V_g(T)$ gives (32). □

### C.4 Estimation slack at an anchor

**Lemma C.3** (Anchor-wise estimation slack)**.** *On the good event $\mathcal{E}_\theta \cap \mathcal{E}_{\mathrm{conf}}$ of Appendix A, for every Stage-2 round, $\mathbb{P}(\mathrm{err}_t = 1 \mid \mathcal{F}_t^-, p_t, \alpha_t^{(g_t)}) \leq \Pi_{[0,1]}(\alpha_t^{(g_t)}) + \varepsilon_t$ with $\varepsilon_t = \sum_j |\widehat{q}_{t,j} - q_{t,j}|$ as in Lemma B.4, and for every anchor $g$,*

$$\frac{1}{V_g(T)} \sum_{t \in \mathcal{T}_g \cap [t_0, T]} \varepsilon_t \;\leq\; \frac{\mathcal{B}_T + \mathcal{J}_T}{V_g(T)}.$$

*Proof.* The first statement is Lemma B.4 with the marginal level replaced by the local one; its proof uses only Lemma A.1 and the definition of the quantile, and is insensitive to which level is plugged in. The second statement bounds the sum over $\mathcal{T}_g$ by the sum over all Stage-2 rounds, which is $\mathcal{B}_T + \mathcal{J}_T$ by (19). □

### C.5 Proof of Theorem 3

**Theorem 6** (Anchor-conditional stockout coverage)**.** *Fix any anchor map $g$ with $G$ anchors and run* CP-CPA-WACI (28)–(29) *with step $\eta \in (0, 1]$. Then, for every anchor $g$ with $V_g(T) \geq 1$ and every realization,*

$$\widehat{\alpha}_T^{(g)} \;\leq\; \alpha + \frac{\alpha + \eta}{\eta\, V_g(T)}.$$

*If, in addition, the round contexts are i.i.d. and $\pi_g := \mathbb{P}(g(x_t) = g) > 0$, then with probability at least $1 - \delta$ simultaneously for all $g$ with $\pi_g \geq \pi_{\min}$, $V_g(T) \geq \pi_{\min}(T - T_1)/2$ once $T - T_1 \geq 8\log(G/\delta)/\pi_{\min}$, and hence*

$$\max_{g:\ \pi_g \geq \pi_{\min}} \widehat{\alpha}_T^{(g)} \;\leq\; \alpha + \frac{2(\alpha + \eta)}{\eta\, \pi_{\min}\, (T - T_1)} \;=\; \alpha + O\Big(\frac{1}{\eta\, \pi_{\min}\, T}\Big).$$

*On the good event of Appendix A the per-round conditional stockout probabilities at anchor $g$ satisfy Lemma C.3.*

*Proof.* The first display is (32). For the second, $V_g(T)$ is a $\mathrm{Bin}(T - T_1, \pi_g)$ variable; the multiplicative Chernoff bound $\mathbb{P}(V_g \leq \pi_g(T - T_1)/2) \leq \exp(-\pi_g(T - T_1)/8)$ and a union bound over the at most $G$ anchors give the visit-count statement, and substituting into (32) gives the rate. The last statement is Lemma C.3. □

**Remark C.1** (What is gained and what is paid)**.** *The anchor-wise guarantee is as strong as the marginal one (Theorem 5) on every anchor, at the price of a slower constant: the effective horizon at anchor $g$ is its visit count $V_g(T) \approx \pi_g T$, so rarely visited segments converge more slowly, and a partition that is too fine (small $\pi_{\min}$) trades conditional validity for a larger $O(1/(\eta\pi_{\min}T))$ term. This is exactly the trade-off visible in the $G = 4$ experiment (Figure 3), where the worst-anchor gap of* CP-CPA-WACI *decays with $T$ while that of marginal* CP-CPA *plateaus.*

**Remark C.2** (Kernel-smoothed anchors)**.** *A continuum version, in which the update at round $t$ is weighted by a kernel $\kappa_h(x^{(g)}, x_t)$ and the level at $x_t$ is interpolated between anchors, is the weighted ACI of Gibbs & Candès (2024). Under a density lower bound on the contexts and Lipschitz dependence of the conditional stockout probability on $x$, one expects a rate of the form $O(h) + O(1/(\eta T h^d))$, i.e. $O(T^{-1/(d+3)})$ after balancing. We do not claim this rate: the interpolated level is no longer piecewise constant along the visiting subsequence, so the exact identity* (31) *does not hold verbatim, and the released implementation uses the hard partition* (28)*. Extending Theorem 6 to kernel weights is left to future work.*

### C.6 Compatibility with the regret bound

The anchor-wise conformal layer enters the regret decomposition (6) only through Term (C), and Lemma A.11 applies verbatim: Term (C) is at most $\gamma_{\max} D_T^+$ with $D_T^+ = \sum_t (K_t^{\mathrm{cp}} - \widehat{K}_t)_+$ now computed with the local levels. In the slack regime of Assumption A.6, the proof of Lemma A.12 goes through anchor by anchor (the drift argument is run on each visiting subsequence, with $\alpha_{t_0}^{(g)} = \alpha$ for every $g$), giving $\mathbb{E}[D_T^+] \le G\, n(1 + 8/\rho^2) + 10nB(\mathcal{B}_T + \mathcal{J}_T)/(\rho\rho')$; the only change is the factor $G$ in the horizon-free term. In the binding regime, $D_T^+$ is linear in $T$ for the same reason as in the marginal case (Proposition A.1): every anchor-conditionally coverage-feasible policy must carry the safety inventory on every anchor where the target binds.

## D Algorithm Listing, Complexity, and Baselines

This appendix gives the step-by-step listing of CP-CPA, its complexity and hyperparameters, and a summary of how the baselines of §6 differ from it along the two axes that matter for the problem.

**Algorithm 1** CP-CPA: Conformal Contextual Pricing and Allocation

**Input:** horizon $T$; target $\alpha$; step $\eta$; bounds $B, B_x, c, \kappa$; residual grid $\{w_k\}$; candidate price grid $\mathcal{P}$.
Initialise $\widehat{\theta} = 0$ and grid counters $\{(N_k, \Sigma_k)\} = 0$.
**for** $t = 1, \dots, T_1$ **do** ▷ Stage 1: parameter recovery
  Post $p_t \sim \mathrm{Unif}[0, B]$ and provision $K_t = n$.
  Observe purchase indicators $\{o_{t,j}\}$ and the count $N_t$; buffer $(x_t, p_t, o_t)$.
  Update $\widehat{\theta}$ by ridge least squares on $(x_{t,j}, B\, o_{t,j})$.
**end for**
Freeze $\widehat{\theta}$; assign all buffered observations to residual cells; set $\alpha_{T_1+1}^{\mathrm{eff}} = \alpha$.
**for** $t = T_1 + 1, \dots, T$ **do** ▷ Stage 2: optimistic pricing and conformal safety
  Compute $p_t$ by (2), the optimistic profit over $\mathcal{P}$; break ties by optimistic revenue.
  Compute $\widehat{q}_{t,j} = \widehat{S}(p_t - \langle x_{t,j}, \widehat{\theta}\rangle)$ for $j = 1, \dots, n$; form $\widehat{P}_t$ and $\widehat{K}_t$.
  Set $K_t = \max(\widehat{K}_t, K_t^{\mathrm{cp}})$ by (3).
  Provision $\vec{I}_t$ across suppliers cheapest first.
  Realise $N_t$; observe $\mathrm{err}_t = \mathbf{1}\{N_t > K_t\}$ and $\{o_{t,j}\}$.
  Update grid statistics $(N_k, \Sigma_k)$; update $\alpha_{t+1}^{\mathrm{eff}}$ by (4).
**end for**

**Tie-breaking in** (2)**.** Ties, in particular the degenerate case in which every candidate price has non-positive optimistic profit, are broken by optimistic revenue $\sum_j p\, \widehat{q}_{t,j}^{+}(p)$. The analysis is indifferent to the tie-break.

**Complexity.** Each Stage-2 round costs $O(|\mathcal{P}|\, n^2)$ for the optimistic-profit scan (one poisson-binomial pmf per candidate price), $O(n^2)$ for $\widehat{P}_t$, $\widehat{K}_t$, and the quantile, and $O(m)$ for the allocation. With $|\mathcal{P}| = \Theta(T^{1/3})$ the per-round cost is $O(n^2 T^{1/3} + m)$ and the total is $O(n^2 T^{4/3} + mT)$. No matrix inversion is needed beyond the $d \times d$ ridge updates of Stage 1.

**Hyperparameters.** The schedule $T_1 = \lceil T^{2/3} \rceil$ and the grid mesh $\Delta \asymp T^{-1/3}$ are those of Xu & Wang (2026). The implementation uses a fixed residual grid of $M = 30$ cells and a fixed candidate price grid of $|\mathcal{P}| = 60$ points for all horizons, the confidence radius $b_k(t) = \sqrt{\log T/(2\max(1, N_k(t)))}$, and ridge parameter 1 in Stage 1. The conformal layer adds the single step size $\eta$, set to $0.05$ in every experiment; Appendix E.3 shows that coverage is insensitive to $\eta$ across two orders of magnitude while a larger $\eta$ widens the oscillation of $\alpha_t^{\text{eff}}$ and hence the safety-inventory cost.

Table 2: Method differences along the two key axes. Only CP-CPA both estimates a contextual demand model from censored feedback and provides distribution-free coverage of the realised stockout rate. "Plug-in" denotes the demand quantile of the estimated pmf at level $1-\alpha$ without online calibration.

| Method | Contextual demand model | Online coverage calibration |
|---|---|---|
| CMRUP-ONLY (Xu & Wang, 2026) | yes (UCB on residual grid) | none (provisions $\lceil \mathbb{E}_{\widehat{P}_t} N \rceil$) |
| LCB-OCO-ONLY (Xu et al., 2025) | no (price-only optimism) | plug-in only |
| $\widehat{K}$-only (ablation) | yes | none (provisions $\widehat{K}_t$) |
| JOINT-NO-CP (ablation) | yes | plug-in only, $\max(\widehat{K}_t, Q_{\widehat{P}_t}(1-\alpha))$ |
| D2-EXP4 (Xu & Wang, 2022) | yes (sampled linear pricers) | none (full capacity) |
| CP-CPA (ours) | yes | yes (adaptive update on $\mathrm{err}_t$) |
| CP-CPA-WACI (ours) | yes | yes, one state per anchor |

# E Experimental Details and Complete Results

This appendix collects the environment and protocol details, the complete result tables, and the additional figures behind §6. All numbers are produced by the code in the supplementary material.

## E.1 Synthetic benchmark: environment, methods, and protocol

**Environment.** At each round we sample $n = 10$ consumer contexts $x_{t,j} \in \mathbb{R}^5$: coordinate 0 is an intercept, coordinate 1 is a per-round shared market shift $\mathrm{Unif}(-0.7, 0.7)$ (so the optimal price moves from round to round and contextual learning is genuinely useful), and coordinates 2 to 4 are per-consumer features $\mathrm{Unif}(0,1)$. The valuation parameter is $\theta^\star = (2.0, 1.0, 0.3, 0.3, 0.3)^\top$ and the noise support is $[-1, 1]$. Two noise laws are used. *Uniform*: $\xi \sim \mathrm{Unif}[-1,1]$, whose survival function is Lipschitz. *Cliff*: $\xi = 0$ with probability $0.3$ and otherwise $\xi \sim \mathrm{Unif}[-1,1]$, whose survival function has a jump of $0.3$ at $w = 0$, a canonical agnostic non-Lipschitz instance. Three suppliers have sorted unit costs $c \in \{0.50, 0.65, 0.80\}$ and holding costs $\gamma \in \{0.10, 0.13, 0.16\}$; inventory is perishable. With these holding costs the estimated profit-optimal inventory $\widehat{K}_t$ stocks out in only 2 to 3% of rounds, so the target $\alpha = 0.10$ is slack in the sense of Assumption A.6. To create a binding instance we multiply all holding costs by 5 ($\gamma \in \{0.50, 0.65, 0.80\}$), which pushes the stockout rate of $\widehat{K}_t$ to 0.17 to 0.19; every other parameter is unchanged.

**Methods.** $\widehat{K}$-ONLY is our pricing and allocation stack with the conformal layer removed, $K_t = \widehat{K}_t$; it is the learned analogue of the unconstrained oracle and is used to measure term (C) of (6). CMRUP-ONLY uses the same contextual pricing with the naive expected-demand inventory $\lceil \mathbb{E}_{\widehat{P}_t} N \rceil$. LCB-OCO-ONLY is a non-contextual price optimism rule with a plug-in quantile inventory. JOINT-NO-CP uses our stack with inventory $\max(\widehat{K}_t, Q_{\widehat{P}_t}(1-\alpha))$ at the fixed plug-in level; it is the ablation that isolates the conformal loop. D2-EXP4 is the distribution-free sampled-policy contextual bandit of Xu & Wang (2022) with full-capacity provisioning, a pool of 40 random linear pricers plus one constant-price expert per grid slot, step size $\sqrt{\log N/(KT)}$, and exploration mix

0.05. All residual-grid methods share $T_1 = \lceil T^{2/3} \rceil$, $M = 30$ residual cells, a 60-point candidate price grid, and $\alpha = 0.10$; CP-CPA uses $\eta = 0.05$.

**Metrics.** We report cumulative pseudo-regret $R_T = \sum_t (\Pi_t^\star - \mathbb{E}[\Pi_t(p_t, K_t) \mid x_t])$ against the unconstrained clairvoyant that knows $(\theta^\star, S)$ and the supply schedule; in the binding regime this oracle violates the coverage target, and Proposition A.1 shows that a linear gap to it is unavoidable for any safe policy. Stockouts are measured on realised demand over all $T$ rounds, Stage 1 included, where full provisioning never stocks out. For the two methods that use the $\max(\widehat{K}_t, K_t^{\mathrm{cp}})$ rule we record the safety inventory $D_T^+$ of (5). We use 5 seeds per cell and fit slopes by least squares in log-log coordinates with bootstrap 95% intervals over seeds.

### E.2 E1 AND E2: COMPLETE TABLES

Table 3: **Regret slope** $\widehat{\beta}$ by log-log least squares, bootstrap 95% interval over 5 seeds, in the slack (left) and binding (right) regimes. Theoretical reference: $2/3 \approx 0.667$ for residual-grid methods (Theorem 1) and $3/4$ for D2-EXP4. In the slack regime CP-CPA and JOINT-NO-CP coincide; in the binding regime the frozen plug-in level of JOINT-NO-CP costs a steeper slope while CP-CPA stays within the interval of the unconstrained $\widehat{K}$-only.

| | | slack ($\gamma \times 1$) | | binding ($\gamma \times 5$) | |
|---|---|---|---|---|---|
| Noise | Method | $\widehat{\beta}$ | 95% interval | $\widehat{\beta}$ | 95% interval |
| cliff | $\widehat{K}$-only | 0.637 | [0.553, 0.706] | 0.669 | [0.605, 0.739] |
| cliff | CMRUP-ONLY | 0.707 | [0.647, 0.761] | 0.677 | [0.624, 0.733] |
| cliff | LCB-OCO-ONLY | 0.874 | [0.857, 0.892] | 0.904 | [0.900, 0.909] |
| cliff | JOINT-NO-CP | 0.637 | [0.560, 0.706] | 0.722 | [0.668, 0.769] |
| cliff | D2-EXP4 | 0.844 | [0.838, 0.850] | 0.893 | [0.885, 0.902] |
| cliff | CP-CPA (ours) | 0.637 | [0.560, 0.708] | 0.687 | [0.631, 0.743] |
| uniform | $\widehat{K}$-only | 0.585 | [0.519, 0.644] | 0.623 | [0.580, 0.665] |
| uniform | CMRUP-ONLY | 0.696 | [0.635, 0.754] | 0.634 | [0.589, 0.679] |
| uniform | LCB-OCO-ONLY | 0.757 | [0.748, 0.771] | 0.816 | [0.794, 0.832] |
| uniform | JOINT-NO-CP | 0.585 | [0.524, 0.644] | 0.720 | [0.693, 0.746] |
| uniform | D2-EXP4 | 0.815 | [0.807, 0.822] | 0.882 | [0.869, 0.893] |
| uniform | CP-CPA (ours) | 0.585 | [0.519, 0.642] | 0.659 | [0.620, 0.697] |

Table 4: **Realised stockout rate at** $T = 16000$ (mean $\pm$ std over 5 seeds), target $\alpha = 0.10$, and average safety inventory $D_T^+/T$ per round for the two methods that use the $\max(\widehat{K}_t, K_t^{\mathrm{cp}})$ rule. Only CP-CPA meets the target in the binding regime; in the slack regime the conformal layer is inactive and both methods inherit the over-coverage of the profit optimum.

| | | slack ($\gamma \times 1$) | | binding ($\gamma \times 5$) | |
|---|---|---|---|---|---|
| Noise | Method | Stockout | $D_T^+/T$ | Stockout | $D_T^+/T$ |
| cliff | $\widehat{K}$-only | $0.018 \pm 0.008$ | n/a | $0.170 \pm 0.019$ | n/a |
| cliff | CMRUP-ONLY | $0.222 \pm 0.031$ | n/a | $0.229 \pm 0.033$ | n/a |
| cliff | LCB-OCO-ONLY | $0.125 \pm 0.116$ | n/a | $0.199 \pm 0.040$ | n/a |
| cliff | JOINT-NO-CP | $0.018 \pm 0.008$ | 0.00 | $0.043 \pm 0.012$ | 0.83 |
| cliff | D2-EXP4 | $0.000 \pm 0.000$ | n/a | $0.000 \pm 0.000$ | n/a |
| cliff | CP-CPA (ours) | $0.018 \pm 0.008$ | 0.00 | $\mathbf{0.096 \pm 0.000}$ | 0.44 |
| uniform | $\widehat{K}$-only | $0.017 \pm 0.003$ | n/a | $0.167 \pm 0.003$ | n/a |
| uniform | CMRUP-ONLY | $0.227 \pm 0.013$ | n/a | $0.240 \pm 0.014$ | n/a |
| uniform | LCB-OCO-ONLY | $0.196 \pm 0.025$ | n/a | $0.211 \pm 0.025$ | n/a |
| uniform | JOINT-NO-CP | $0.017 \pm 0.003$ | 0.00 | $0.042 \pm 0.005$ | 0.93 |
| uniform | D2-EXP4 | $0.000 \pm 0.000$ | n/a | $0.000 \pm 0.000$ | n/a |
| uniform | CP-CPA (ours) | $0.017 \pm 0.003$ | 0.00 | $\mathbf{0.096 \pm 0.000}$ | 0.47 |

Table 5: **The price of safety in the binding regime.** Mean cumulative pseudo-regret of CP-CPA and of the unconstrained $\widehat{K}$-only (same pricing trajectory, conformal layer removed) against the unconstrained oracle, their gap per round, and the safety inventory per round of CP-CPA, as a function of $T$ (cliff / uniform). The gap per round is constant, so term (C) is linear in $T$ as Proposition A.1 requires, while the stockout rate of CP-CPA stays at $\alpha$ minus the Stage-1 dilution. The bound $\gamma_{\max}\mathbb{E}[D_T^+]$ of Theorem 1 over-states the measured gap by a factor of about 5 because safety units are sometimes sold.

| $T$ | 500 | 1000 | 2000 | 4000 | 8000 | 16000 |
|---|---|---|---|---|---|---|
| $R_T$ CP-CPA (cliff / unif.) | 1007 / 680 | 1407 / 1007 | 2194 / 1564 | 3360 / 2525 | 6113 / 4089 | 10722 / 6521 |
| $R_T$ $\widehat{K}$-only | 971 / 638 | 1341 / 939 | 2057 / 1420 | 3097 / 2271 | 5605 / 3559 | 9717 / 5373 |
| gap per round | .072 / .085 | .066 / .068 | .069 / .072 | .066 / .064 | .063 / .066 | .063 / .072 |
| $D_T^+/T$ CP-CPA | .66 / .55 | .56 / .44 | .56 / .49 | .51 / .46 | .48 / .48 | .44 / .47 |
| stockout CP-CPA | .085 / .084 | .089 / .088 | .091 / .091 | .093 / .093 | .095 / .095 | .096 / .096 |

### E.3 E3: STEP-SIZE ABLATION

We sweep $\eta \in \{0.01, 0.05, 0.10, 0.20\}$ at $T = 4000$ in the binding regime with 3 seeds (Table 6). The realised stockout rate is $0.091$ to $0.094$ for every $\eta$, that is at the deterministic ceiling $\alpha(1 - T_1/T) + (\alpha+\eta)/(\eta T) \approx 0.094$ of Theorem 2 up to seed-to-seed noise of at most $0.0014$. Regret is not entirely insensitive: a larger $\eta$ makes $\alpha_t^{\text{eff}}$ oscillate more widely, so $D_T^+$ and hence term (C) grow, by $9\%$ on cliff and $15\%$ on uniform noise from $\eta = 0.01$ to $0.20$. This is the trade-off Theorem 1 makes explicit: the coverage guarantee improves with $\eta$ only through the $O(1/(\eta T))$ burn-in term, whereas the holding cost of the oscillation grows with $\eta$. The default $\eta = 0.05$ is a reasonable compromise and is not tuned per environment.

Table 6: **Step-size ablation at** $T = 4000$, binding regime, 3 seeds. Coverage sits at the deterministic ceiling for every $\eta$; the regret cost of a larger $\eta$ is the wider oscillation of the conformal level.

| Noise | $\eta = 0.01$ | $\eta = 0.05$ | $\eta = 0.10$ | $\eta = 0.20$ |
|---|---|---|---|---|
| cliff stockout (ceiling 0.094) | $0.092 \pm 0.001$ | $0.093 \pm 0.001$ | $0.094 \pm 0.000$ | $0.094 \pm 0.000$ |
| cliff $R_T$ | 3666 | 3722 | 3811 | 4001 |
| uniform stockout | $0.091 \pm 0.001$ | $0.093 \pm 0.000$ | $0.093 \pm 0.000$ | $0.094 \pm 0.000$ |
| uniform $R_T$ | 2405 | 2448 | 2531 | 2757 |

### E.4 E4: SEMI-REAL RETAIL SIMULATOR

The semi-real SKU simulator has calendar-driven seasonality, a Markov promotion regime with transition probabilities $0.10$ and $0.20$, and a per-consumer mixture of regular and deal-seeking subgroups whose valuation noises have atom probabilities $0.10$ and $0.55$ at zero. Holding cost rises by $50\%$ during promotion rounds. The seller observes only calendar and promotion features plus per-consumer category and income proxies, an 8-dimensional context. Table 7 reports the realised marginal stockout rate (mean, std, and worst seed), its split over deal-heavy and regular-heavy rounds (rounds with at least $50\%$ deal-seeker share versus fewer), average net profit, cumulative regret, and its log-log slope at $T = 8000$ over 5 seeds with the default holding costs. All residual-grid methods share the slope $\widehat{\beta} \approx 0.69$ on this non-i.i.d. stream. The last row is the anchor-conditional variant of Appendix E.5 on the $G = 4$ partition; it is indistinguishable from marginal CP-CPA on every aggregate.

### E.5 E5: ANCHOR-CONDITIONAL COVERAGE

The $G = 4$ anchor grid is the product of the promotion regime ($\rho_t \in \{0, 1\}$) and the deal-seeker share (at least $0.4$ or not), two axes that modulate different sources of conditional miscoverage (operating point versus atom mass) and are each well populated. Holding costs are multiplied by 10 so that the target binds in as many anchors as possible; horizons are $T \in \{2, 4, 8, 16, 32\} \times 10^3$ with 10 seeds. Figure 3 reports two statistics per run. Panel (b) is the worst-anchor overshoot $\max_g(\widehat{\alpha}_T^{(g)} - \alpha)_+$ bounded by Theorem 3: for CP-CPA-WACI it is zero in 47 of 50 runs and at most

Table 7: **Semi-real retail simulator at** $T = 8000$, 5 seeds, $\alpha = 0.10$. "worst seed" is the maximum stockout rate over seeds: JOINT-NO-CP and $\widehat{K}$-only have one catastrophic seed, CP-CPA none.

| Method | Stockout (marginal) | worst seed | deal-heavy | reg.-heavy | Avg. net profit | $R_T$ (mean) | $\widehat{\beta}$ |
|---|---|---|---|---|---|---|---|
| CMRUP-ONLY | $0.302 \pm 0.221$ | 0.644 | 0.322 | 0.290 | $9.72 \pm 1.01$ | 17708 | 0.74 |
| $\widehat{K}$-only | $0.092 \pm 0.124$ | 0.334 | 0.105 | 0.085 | $10.15 \pm 1.17$ | 14286 | 0.69 |
| LCB-OCO-ONLY | $0.223 \pm 0.057$ | 0.284 | 0.197 | 0.237 | $9.39 \pm 0.16$ | 20382 | 0.93 |
| JOINT-NO-CP | $0.091 \pm 0.123$ | 0.330 | 0.104 | 0.084 | $10.15 \pm 1.17$ | 14268 | 0.69 |
| D2-EXP4 | $0.000 \pm 0.000$ | 0.000 | 0.000 | 0.000 | $6.66 \pm 0.11$ | 42211 | 0.88 |
| CP-CPA (ours) | $0.044 \pm 0.038$ | **0.095** | 0.054 | 0.038 | $10.21 \pm 1.18$ | 13757 | 0.69 |
| CP-CPA-WACI (ours) | $0.044 \pm 0.038$ | **0.096** | 0.050 | 0.040 | $10.21 \pm 1.18$ | 13754 | 0.69 |

0.002 in the other three (all at $T \leq 4000$), below the deterministic ceiling $(\alpha + \eta)/(\eta \min_g V_g(T))$ at every horizon; marginal CP-CPA overshoots in 30 of 50 runs, by up to 0.156 in a single anchor and by 0.015 to 0.042 on average across seeds, with no downward trend in $T$. Panel (a) is the two-sided gap $\max_g |\widehat{\alpha}_T^{(g)} - \alpha|$, which also charges for under-coverage. Both methods under-cover in anchors where the profit optimum already meets the target (the slack regime, in which no safe policy should add inventory), so this gap is not driven to zero by either; it nevertheless falls from 0.056 to 0.0035 for CP-CPA-WACI (log-log slope $-0.86$) as the per-anchor estimates sharpen, versus 0.083 to 0.047 for marginal CP-CPA.

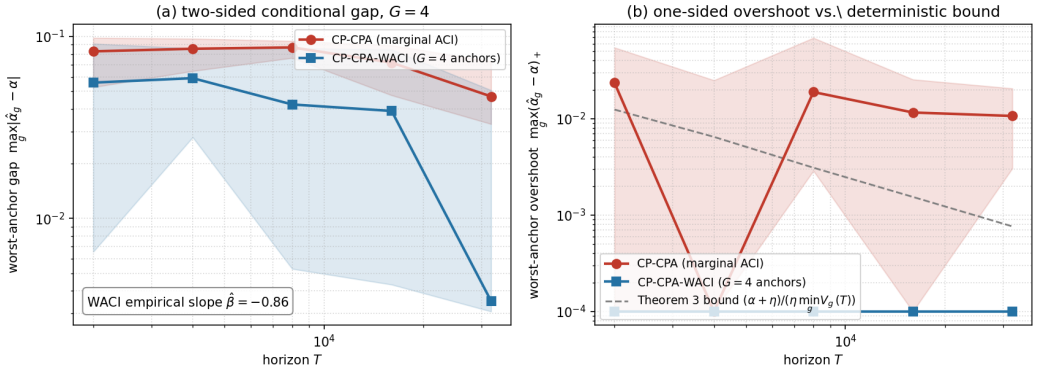



Figure 3: **Anchor-conditional coverage over a** $G = 4$ **anchor grid** (semi-real simulator, holding costs times 10, $\alpha = 0.10$, median of 10 seeds with interquartile band). (a) two-sided worst-anchor gap; (b) one-sided worst-anchor overshoot, the quantity bounded by Theorem 3, with the deterministic ceiling $(\alpha + \eta)/(\eta \min_g V_g(T))$ (dashed). CP-CPA-WACI (blue) never overshoots (plotted at the $10^{-4}$ floor); marginal CP-CPA (red) leaves a persistent per-anchor overshoot that a single conformal state cannot remove.

### E.6 E6: REAL-DATA-CALIBRATED EPISODE

**Calibration.** We take the highest-volume SKU of the UCI Online Retail II log (stock code 85123A, a decorative t-light holder), aggregate it to a daily panel of $D \approx 589$ trading days, and fit a weighted logistic demand model $q = \sigma(a + b^\top x - \lambda p)$ whose linear index $\hat{\mu}(x) = (a + b^\top x)/\lambda$ defines the calibrated valuation parameter $\theta^\star$. The daily contexts are the SKU's real features: annual seasonality (sine and cosine of the day of year), a December and holiday flag, a weekend flag, a causal 7-day demand-momentum term, and a long-run trend, augmented with a per-consumer income covariate drawn from the dataset's own customer log-spend distribution. The idiosyncratic noise $\xi$ is sampled from the bounded mean-zero survival curve $S(w) = \mathbb{P}(\xi \geq w)$ implied by the fitted elasticity $\lambda$ (calibrated-survival pseudo-$R^2 = 0.86$ against the observed daily purchase rate); we also retain the non-parametric isotonic fit of $S$ for the demand-curve diagnostic in Figure 4(a). Because the valuation mean stays linear in $x$, the Stage-1 censored estimator remains unbiased and

all agents run without modification. One episode is a chronological pass over the trading year; horizons $T \in \{2356, 4712, 9424, 14136\}$ correspond to 4, 8, 16, and 24 episodes.

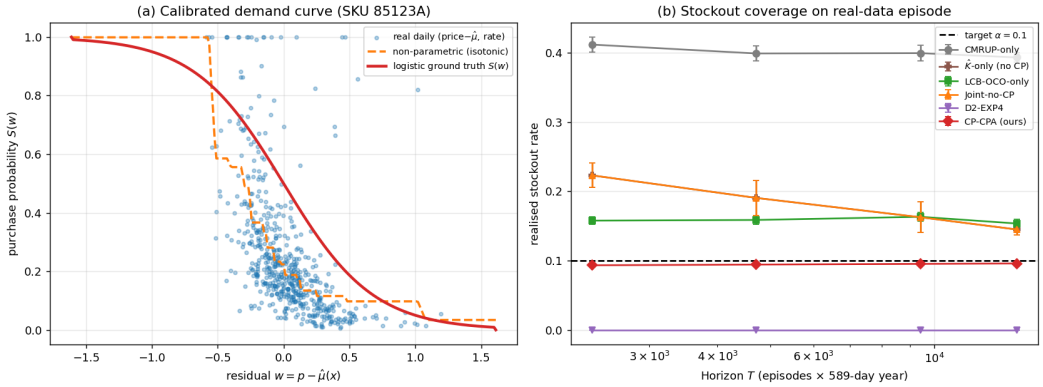



Figure 4: **Real-data-calibrated episode (UCI Online Retail II, SKU 85123A).** (a) The calibrated demand curve: real daily (price minus $\hat{\mu}$, purchase rate) scatter, the non-parametric isotonic fit, and the logistic ground-truth survival $S(w)$ used by the environment. (b) Realised stockout rate against horizon (5 seeds, one standard deviation). CP-CPA (red) sits at 0.093 to 0.096 for every horizon with zero seed-to-seed variance, while the plug-in JOINT-NO-CP (orange, coinciding with $\widehat{K}$-only) starts at 0.22 and is still 45% above target after 24 episodes, and D2-EXP4 sits at 0 by over-provisioning.

Table 8: **Real-data-calibrated episode at $T = 14136$ (24 episodes)**, 5 seeds, $\alpha = 0.10$. "Dec." is the stockout rate on December and holiday rounds. CP-CPA is the only method whose marginal rate meets the target; it matches its ablation JOINT-NO-CP on profit and regret. LCB-OCO-ONLY attains the lowest raw regret by running a 0.154 stockout rate, that is by violating the constraint (see "Reading the regret column" below).

| Method | Stockout | Dec. | Avg. profit | $R_T$ | $D_T^+/T$ |
|---|---|---|---|---|---|
| CMRUP-ONLY | $0.393 \pm 0.015$ | 0.716 | 26.95 | 39389 | n/a |
| $\widehat{K}$-only | $0.145 \pm 0.018$ | 0.454 | 27.91 | 25760 | n/a |
| LCB-OCO-ONLY | $0.154 \pm 0.013$ | 0.625 | 28.75 | 13750 | n/a |
| JOINT-NO-CP | $0.145 \pm 0.018$ | 0.454 | 27.91 | 25760 | 0.00 |
| D2-EXP4 | $0.000 \pm 0.000$ | 0.000 | 12.03 | 250434 | n/a |
| CP-CPA (ours) | $\mathbf{0.096 \pm 0.000}$ | 0.329 | 27.88 | 26202 | 2.02 |

**Findings in detail.** The realised stockout rate of CP-CPA is 0.093, 0.094, 0.095, 0.096 at $T = 2356, 4712, 9424, 14136$ with standard deviation 0.000 across seeds at every horizon: the Stage-2 rate is pinned to $\alpha$ up to the $(\alpha + \eta)/(\eta(T - T_1)) < 0.003$ correction and the residual $\alpha - \widehat{\alpha}_T$ is the Stage-1 dilution $\alpha T_1/T$. The plug-in layer of JOINT-NO-CP is inactive ($D_T^+ = 0$ in every run, so it coincides with $\widehat{K}$-only), and its realised rate is 0.223 after 4 episodes, 0.145 after 24, and 0.454 in December. CMRUP-ONLY (0.393) and LCB-OCO-ONLY (0.154) also overshoot, and D2-EXP4 avoids stockouts by over-provisioning at a 57% profit penalty (12.03 versus 27.88). CP-CPA provisions on average $D_T^+/T = 2.02$ safety units per round on top of the estimated profit optimum, shrinking with $T$ (5.9, 4.3, 2.8, 2.0) as the demand estimate sharpens; the price is 0.1% of average profit and 1.7% of cumulative regret relative to JOINT-NO-CP. The off-season stockout rate of CP-CPA is 0.077 against 0.329 in December.

**Reading the regret column.** LCB-OCO-ONLY attains the lowest raw regret (13750) in Table 8, below the 26202 of CP-CPA. This is not evidence against the method. First, LCB-OCO-ONLY achieves it by violating the safety constraint, with a 0.154 stockout rate (54% above the target) and 0.625 in December, so it is not an admissible competitor; among methods that meet the target, CP-CPA is the only entry. Second, the effect is specific to this calibrated curve: the fitted demand for

a single high-volume SKU is close to non-contextual, so a price-only learner has little contextual structure to exploit, whereas the contextual and inventory machinery of CP-CPA and JOINT-NO-CP pays a fixed $\widetilde{O}(T^{2/3})$ Stage-1 cost that is not amortised at this horizon. The controlled comparison is therefore CP-CPA against its ablation JOINT-NO-CP, which share the identical pricing and allocation stack and differ only in the conformal layer: they match on profit ($27.88$ versus $27.91$) and regret ($26202$ versus $25760$), while only CP-CPA attains the target coverage. The clean regret-slope separation is the job of the synthetic benchmark, where oracle access is exact.

### E.7 LIMITATIONS OF THE EMPIRICAL EVALUATION

Three caveats remain. First, we report pseudo-regret against the unconstrained oracle's expected profit rather than realised regret, which is standard in contextual pricing benchmarks but understates the variance contribution of the $O(T^{2/3})$ exploration rounds. Second, the binding regime is created by scaling holding costs; we did not search over other ways of making the target bind, and at $T \leq 16000$ the linear term (C) is still small relative to the $T^{2/3}$ pricing regret, so the slopes in Table 3 do not yet show the asymptotic linear behaviour that Proposition A.1 predicts. Third, on the semi-real simulator the target is slack in most anchors and seeds, so the two-sided conditional gap of Figure 3(a) is a conservative statistic; the one-sided overshoot in panel (b) is the quantity our theory controls.