LEARNING WHILE PRICING FAIRLY: QUEUE-REGULARIZED EXPLORATION FOR PRICING EXPERIMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pricing experiments generate information about demand, but the prices used for learning are also the offers that different customer groups receive. We study how a seller can reduce differences between group-average prices while retaining the random variation needed for demand estimation. We extend Random Shock Design (RSD) with a fairness penalty scaled by virtual queues. This fairness-regularized RSD (FRSD) uses past disparities to correct the next price, then adds an independent shock for demand estimation. Under bounded demand and a well-conditioned regression design, we bound terminal price disparity and establish vanishing disparity at zero tolerance with diminishing shocks, while demand estimation remains consistent. The revenue analysis separates learning error from price adjustments; a fair benchmark and separate lower bounds identify the cost of terminal fairness. Experiments compare FRSD with RSD, fixed penalties, and current-gap penalties. Synthetic results include a 92% reduction in mean disparity alongside a 2.4% increase in mean revenue relative to RSD with the same estimation and exploration settings. In the Online Retail-calibrated simulation, FRSD reduces disparity by 19% with small costs in revenue and estimation accuracy.

1 INTRODUCTION

A seller conducting a pricing experiment uses price variation to learn demand while earning revenue (Simchi-Levi & Wang, 2023). These experimental prices are also the offers customers actually receive, and their averages can differ across groups. Differences between the group averages can persist even as demand estimates improve. We study how later offers can respond to these differences while continuing to generate information about demand.

Random Shock Design (RSD) provides a starting point: it adds an exogenous perturbation to the price that maximizes estimated revenue, and the resulting variation identifies the effect of price on demand (Nambiar et al., 2019; Simchi-Levi & Wang, 2023). To incorporate fairness, we adjust this revenue-based price in response to past group disparities, then add the shock. Applying the correction in this order preserves the perturbation used for estimation while allowing earlier offers to influence the next pricing decision.

Our fairness-regularized RSD (FRSD) adjusts the correction as price disparities accumulate. For each group, a virtual queue grows while its running average price differs from the market mean (the average of all offers made so far) by more than a specified tolerance, and decreases toward zero when the gap falls below it. With a positive fairness weight, a larger queue places a stronger penalty on pricing away from the market mean. The seller therefore responds not only to the current gap but also to its history: earlier disparities can continue to influence offers after that gap has declined.

Preserving the independent shocks lets us bound demand-estimation error even as the queues change the prices. Under a well-conditioned regression design and diminishing shocks, we also obtain a direct bound on terminal price disparity: at zero tolerance, group-average prices converge to one another while demand learning continues. The proof links accumulated gaps to stronger corrections, establishing fairness through the pricing rule itself. Fairer prices can nevertheless reduce revenue because the true revenue-maximizing prices need not be fair. We therefore separate learning error from the revenue effect of price adjustments and compare the unconstrained optimum with the best

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.

terminally fair price path. Separate lower bounds distinguish the cost of learning from the cost of the fairness requirement itself.

We evaluate these effects in a synthetic environment and an Online Retail-calibrated simulation (Retail), comparing FRSD with RSD, fixed penalties, and current-gap penalties. A bounded linear-demand environment connects these comparisons to the analytical model, with longer runs examining how terminal disparity changes over time.

Our contributions are:

1. We propose FRSD, which uses past price disparities to adjust the strength of a fairness penalty and applies independent shocks after the correction. By correcting the price before drawing the shock, the seller can address group disparities without losing the random variation needed for demand learning.
2. We establish a direct terminal-fairness bound under bounded-model and regression-design conditions. With diminishing shocks and zero tolerance, root-mean-square disparity is $O(\sqrt{\log(eT)/T})$ while slope estimation remains consistent. We also bound the contribution of estimation error to regret and quantify the revenue cost of constraining group-average prices through a fair benchmark.
3. Experiments examine fairness across time horizons and pricing environments. Fixed-parameter runs show decreasing disparity at zero tolerance. Using the same parameter-selection rule, FRSD outperforms fixed and current-gap penalties on disparity in a synthetic setting. In the Retail simulation, it reduces terminal and time-averaged disparity while keeping revenue and estimation error close to RSD.

2 Related Work

Pricing experiments and demand learning. The dynamic pricing literature studies revenue maximization under stochastic demand (Gallego & van Ryzin, 1994) and the learning problem that arises when demand is unknown (Besbes & Zeevi, 2009; Broder & Rusmevichientong, 2012; Harrison et al., 2012; Keskin & Zeevi, 2014; den Boer, 2015). Contextual models bring customer features into the pricing rule (Cohen et al., 2020; Ban & Keskin, 2021), including context-dependent price sensitivity and its implications for regret (Xu & Wang, 2024). We start from random-shock pricing, in which exogenous perturbations of greedy prices identify the effect of price on demand and their magnitude sets the balance between estimation accuracy and revenue loss (Nambiar et al., 2019; Simchi-Levi & Wang, 2023). FRSD keeps this source of variation in a model with a common slope across groups, and changes only the pre-shock decision: estimated revenue is penalized according to the disparities accumulated in each group's price history.

Fair and constrained sequential decisions. Fair bandits impose fairness requirements while learning rewards (Joseph et al., 2016; 2018; Chen et al., 2020; Wang et al., 2021), and other sequential formulations account for decision histories (Alamdari et al., 2024), long-term effects (Xu et al., 2024), or counterfactual comparisons (Chen et al., 2025). Constrained bandit methods study the interplay between regret and cumulative constraint violations (Guo & Liu, 2024; Bernasconi et al., 2025). In FRSD, the relevant history is the sequence of prices actually offered to each group: the algorithm compares the groups' running averages with the market mean and carries the differences into the pricing objective through virtual queues.

Fair pricing. Fair-pricing research examines restrictions on price differences across customers or groups (Cohen et al., 2022; Das et al., 2022; Chen et al., 2023; Yang et al., 2024). Cohen et al. (2025) study price and demand fairness with unknown demand, while Xu et al. (2023) distinguish equality of expected offered prices from equality of expected accepted prices. We study realized group-average offers over a finite experiment, using their accumulated differences to adjust later prices while preserving random-shock identification. Following the efficiency–fairness perspective (Bertsimas et al., 2011; 2012; Bateni et al., 2022), a known-demand benchmark separates the cost of constraining these averages from the cost of learning demand.

3 PROBLEM SETUP

A seller conducts a pricing experiment over T customer arrivals. At arrival t, the seller observes the customer's context $x_t \in [0,1]^d$ and group $g_t \in G$, posts a price $p_t \in [\underline{p}, \overline{p}]$, and observes demand $D_t = D_t(p_t)$. The set G contains the groups assessed for fairness, and demand follows the contextual linear model

$$D_t(p) = bp + \theta^\top x_t + \eta_{g_t} + \varepsilon_t. \tag{1}$$

The common slope $b < 0$ captures the effect of price on demand, while θ and the group-specific intercept η_g account for context and group membership; ε_t is demand noise. We assume bounded coefficients, with b in a known interval $[\underline{b}, \overline{b}] \subset (-\infty, 0)$. Learning this slope is the main estimation objective, and exogenous price variation is what separates its effect from the context and group effects that also enter demand.

These prices also provide the basis for assessing fairness: we compare the average prices offered to different groups over the experiment. Averaging over each group's arrivals lets individual offers vary for learning while making persistent differences in the groups' average offers visible. Let $N_g(T) = \sum_{t=1}^{T} \mathbf{1}\{g_t = g\}$ denote the number of arrivals from group g, where $\mathbf{1}$ is the indicator function; the group-coverage assumption used below requires $N_g(T) \geq \pi_{\min} T$ for every group, with a fixed $\pi_{\min} > 0$. For a realized price path $\mathbf{p} = (p_1, \ldots, p_T)$, group-price disparity is

$$\Delta(\mathbf{p}) := \max_{g,g' \in G} \left| \frac{1}{N_g(T)} \sum_{t:g_t=g} p_t - \frac{1}{N_{g'}(T)} \sum_{t:g_t=g'} p_t \right|. \tag{2}$$

We call the path terminally τ-price-fair when $\Delta(\mathbf{p}) \leq \tau$. The tolerance applies to group averages over the whole experiment; individual prices remain free to differ.

To isolate the revenue cost of the fairness requirement, we compare the highest attainable revenues with and without it under the same known demand model. Without the requirement, the seller maximizes expected revenue separately at each arrival; with conditionally mean-zero demand noise, that revenue is

$$r_t(p) = p\big(bp + \theta^\top x_t + \eta_{g_t}\big).$$

Since $b < 0$, this quadratic is maximized without price constraints at

$$p_t^\star = -\frac{\theta^\top x_t + \eta_{g_t}}{2b}, \tag{3}$$

which we assume lies in the feasible interval $[\underline{p}, \overline{p}]$ at every arrival. This price path maximizes total expected revenue, but its group averages need not satisfy the fairness requirement. Imposing the requirement means maximizing revenue over price paths with $\Delta(\mathbf{p}) \leq \tau$, which gives the fair benchmark

$$\mathbf{p}^F(\tau) \in \arg\max_{\mathbf{p} \in [\underline{p}, \overline{p}]^T} \left\{ \sum_{t=1}^{T} r_t(p_t) : \Delta(\mathbf{p}) \leq \tau \right\}. \tag{4}$$

The price of fairness is the difference in optimal revenue induced by this constraint:

$$\mathrm{PoF}_T(\tau) = \sum_{t=1}^{T} r_t(p_t^\star) - \sum_{t=1}^{T} r_t(p_t^F(\tau)). \tag{5}$$

Both benchmarks use known demand, so their revenue difference isolates the cost of terminal price fairness from learning. Appendix F characterizes the fair benchmark through adjustments to group-average prices, including the case of active price bounds. The seller in a pricing experiment must instead act on estimated demand and respond to the group-price differences left by earlier offers. FRSD incorporates this price history into the revenue objective while retaining independent shocks for demand estimation.

4 FRSD: DESIGN AND ANALYSIS

FRSD adjusts the strength of the fairness penalty according to past differences between each group's average price and the market mean. Let $\bar{p}_{g,t-1}$ and $\bar{p}_{t-1}$ denote the group and market average prices

through arrival $t-1$. We track their difference with a virtual queue,

$$Q_{g,t} = \left[Q_{g,t-1} + |\bar{p}_{g,t-1} - \bar{p}_{t-1}| - \tau\right]^+, \qquad Q_{g,1} = 0. \tag{6}$$

for $t \geq 2$, where $[z]^+ = \max\{z, 0\}$ and $\tau \geq 0$. For a group that has not yet appeared, we use the market mean in place of its group mean. The queue grows when the gap exceeds τ and decreases toward zero when it falls below τ. A positive queue can therefore persist after the current gap has declined, carrying earlier disparities into the next pricing decision. Since the update uses only past prices, the queue is available before the next offer is chosen.

For the arriving group g_t, we use this queue to penalize departures from the market mean. The queue scales the penalty in the revenue objective, giving the queue-regularized pricing rule

$$u_t \in \arg\max_{p \in [\underline{p}, \overline{p}]} \left\{ p(\widehat{b}_t p + \widehat{\mu}_t(g_t, x_t)) - \lambda_f Q_{g_t,t}(p - \bar{p}_{t-1})^2 \right\}, \tag{7}$$

where $\widehat{\mu}_t(g, x) = \widehat{\theta}_t^\top x + \widehat{\eta}_{g,t}$ combines the current estimates of the context and group effects, $\lambda_f \geq 0$ scales the penalty, and $\bar{p}_0 \in [\underline{p}, \overline{p}]$ is the initial market reference. A zero queue leaves the seller maximizing estimated revenue alone; as the queue grows, departures from the market mean become more costly, so past disparities weigh more heavily on the next price.

The corrected price u_t determines where the seller explores. To preserve the random variation used for demand estimation, we first project it onto $[\underline{p} + \delta_t, \overline{p} - \delta_t]$ and then add an independent perturbation. This leaves room for either shock sign without violating the price constraints. Set $\delta_t = t^{-\alpha}$ with $\alpha \in [0, 1/4]$ and assume the interior interval is nonempty. The posted price is

$$\widetilde{p}_t = \mathrm{Proj}\big(u_t, [\underline{p} + \delta_t, \overline{p} - \delta_t]\big), \qquad p_t = \widetilde{p}_t + \Delta_t, \quad \Delta_t \in \{-\delta_t, +\delta_t\} \tag{8}$$

Here Proj denotes Euclidean projection. Conditional on the history $\mathcal{F}_{t-1}$ before arrival t and current (x_t, g_t), $\widetilde{p}_t$ is fixed before the equiprobable shock sign is drawn, and the shock is independent of demand noise. Under the linear model, these properties give $\mathbb{E}[\Delta_t D_t \mid \mathcal{F}_{t-1}, x_t, g_t] = b\delta_t^2$. The shock–demand products therefore identify the slope even though the pre-shock price depends on past observations. Accumulating these products and projecting onto the known slope interval gives the estimator

$$\widehat{b}_{t+1} = \mathrm{Proj}\left(\frac{\sum_{s=1}^t \Delta_s D_s}{\sum_{s=1}^t \Delta_s^2}, [\underline{b}, \overline{b}] \right). \tag{9}$$

After updating the slope, we refit the context and group coefficients by ridge regression. The posted price also updates the group and market averages, which determine the next queues. The next offer therefore uses the updated demand estimates while responding to past group-price disparities through the queues. Algorithm 1 summarizes this process.

Algorithm 1 FRSD: overview of pricing and queue updates

1: Initialize estimates with $\widehat{b}_1 \in [\underline{b}, \overline{b}]$ and the market reference $\bar{p}_0$; set $Q_{g,1} = 0$ for all g.
2: **for** $t = 1, \ldots, T$ **do**
3: Observe (x_t, g_t) and set $\delta_t = t^{-\alpha}$.
4: Compute u_t using $Q_{g_t,t}$ and $\bar{p}_{t-1}$ in equation (7).
5: Project the center; draw an equiprobable sign and post p_t as in equation (8).
6: Observe demand; update $\widehat{b}_{t+1}$ by equation (9) and refit the remaining coefficients.
7: Update group and market price averages; set $Q_{g,t+1} = [Q_{g,t} + |\bar{p}_{g,t} - \bar{p}_t| - \tau]^+$ for all g.
8: **end for**

4.1 Price correction

The quadratic penalty yields an explicit correction to the greedy price. Write $v_t = -\widehat{\mu}_t(g_t, x_t)/(2\widehat{b}_t)$ for the unconstrained revenue-maximizing price computed from the current demand estimates without the penalty.

Proposition 1 (Price correction)**.** *If the projection in equation* (8) *is inactive, then*

$$u_t = \frac{-\widehat{b}_t}{-\widehat{b}_t + \lambda_f Q_{g_t,t}} v_t + \frac{\lambda_f Q_{g_t,t}}{-\widehat{b}_t + \lambda_f Q_{g_t,t}} \bar{p}_{t-1}. \tag{10}$$

When the projection is inactive, the correction interpolates between the greedy price and the market mean. Holding the estimates and market mean fixed, increasing $\lambda_f Q_{g_t,t}$ shifts weight toward the market mean, while a zero queue recovers the greedy price. This changes the pre-shock price without changing the magnitude of the perturbation that follows. Appendix C gives the correction with active price constraints and extends the regret analysis to projected prices through a cumulative error bound.

4.2 Estimation, regret, and price differences

The same ordering of correction and randomization lets us analyze estimation error and revenue loss separately. For the upper bounds, we retain the shock independence specified above and assume $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}, x_t, g_t] = 0$ and $|\varepsilon_t| \le \bar{\varepsilon}$ almost surely.

Write $S_T = \sum_{t=1}^{T} \delta_t^2$ for the sum of squared shock magnitudes, and measure expected regret relative to the feasible oracle prices as $\bar{R}(T) = \sum_{t=1}^{T} \mathbb{E}[r_t(p_t^\star) - r_t(p_t)]$. Realized regret, $\sum_{t=1}^{T}[p_t^\star D_t(p_t^\star) - p_t D_t(p_t)]$, also includes demand noise; under the linear model it differs from the noise-free revenue loss by $\sum_{t=1}^{T}(p_t^\star - p_t)\varepsilon_t$.

Theorem 1 (Estimation and price-path bounds)**.** *Under bounded parameters and noise, feasible oracle prices, and exogenous shock signs, there is a constant $C > 0$ such that*

$$\mathbb{E}[(\widehat{b}_{T+1} - b)^2] \le \frac{C}{S_T}, \tag{11}$$

and

$$\bar{R}(T) = (-b) \sum_{t=1}^{T} \mathbb{E}\left[(\widetilde{p}_t - p_t^\star)^2 + \delta_t^2\right]. \tag{12}$$

Moreover, for $T \ge 2$, every $g \in G$ satisfies

$$\frac{1}{T-1} \sum_{t=2}^{T} |\bar{p}_{g,t-1} - \bar{p}_{t-1}| \le \tau + \frac{Q_{g,T}}{T-1}. \tag{13}$$

The quantity S_T captures the direct trade-off introduced by exploration: a larger value tightens the estimation bound but increases the shock contribution $(-b)S_T$ to regret. The remaining regret comes from deviations of the pre-shock prices from the oracle. Quadratic revenue and the shock's zero conditional mean make this separation exact. Appendix D further derives finite-sample and simultaneous estimation-error bounds and extends the analysis to conditionally sub-Gaussian demand noise.

To control the learning contribution to the remaining price error, let ϕ_t stack the context and non-reference group indicators, and let $V_n = \Lambda + \sum_{s=1}^{n} \phi_s \phi_s^\top$ be the ridge design matrix for a fixed positive definite penalty Λ. Define

$$E_T = \sum_{t=1}^{T} \mathbb{E}[(v_t - p_t^\star)^2], \qquad A_T = \sum_{t=1}^{T} \mathbb{E}[(\widetilde{p}_t - v_t)^2].$$

The first quantity measures error in the greedy prices along the FRSD trajectory; the second measures their subsequent adjustment, including projection onto the allowed interval.

Theorem 2 (Learning error and price adjustments)**.** *Under the assumptions of Theorem 1, suppose $\|\phi_t\| \le L$ and either (i) $\lambda_{\min}(V_n) \ge \kappa n$ almost surely for all $n \ge n_0$, with fixed $\kappa > 0$ and n_0, or (ii) the regressors are i.i.d. with $\mathbb{E}[\phi_t \phi_t^\top] \succeq \kappa I$. Then*

$$E_T = O\left(1 + \log T + \sum_{n=1}^{T-1} S_n^{-1}\right), \qquad \left|\frac{\bar{R}(T)}{-b} - A_T - S_T\right| \le E_T + 2\sqrt{E_T A_T}. \tag{14}$$

For $\alpha \in [0, 1/4]$, $E_T = O(S_T)$; hence $A_T = O(S_T)$ implies $\bar{R}(T) = O(S_T)$ together with slope MSE $O(S_T^{-1})$.

The design condition ensures that the context and group coefficients can be estimated from the observed arrivals. The proof tracks error through the two estimation updates: slope error enters the

ridge fit, and both errors enter the greedy price. The resulting bound accounts for arbitrary price adjustments, including active projection, through A_T. Appendix A.5 gives explicit constants and the proof; Corollary 1 covers corrections that do not increase the distance to the oracle price.

The price correction also gives a direct guarantee for terminal fairness. Accumulated gaps strengthen the penalty on departures from the market mean, which controls how far the group averages can remain apart.

Theorem 3 (Terminal price fairness)**.** *Fix $\lambda_f > 0$ and $0 < \alpha \leq 1/4$. Under the assumptions of Theorem 1, the pathwise design condition (i) of Theorem 2, bounded initialization, and group coverage, there are constants C_1, C_2 independent of T and τ such that FRSD satisfies*

$$\mathbb{E}[\Delta(\mathbf{p})^2] \leq C_1\tau + C_2\frac{\log(eT)}{T}. \tag{15}$$

At $\tau = 0$, root-mean-square disparity is $O(\sqrt{\log(eT)/T})$, and terminal disparity converges to zero almost surely when coverage holds along the growing horizon.

The proof tracks variation between group means through a logarithmic queue potential (Appendix E.7). In a two-group known-demand instance, a current-gap penalty at any fixed finite strength cannot attain vanishing disparity, whereas zero-tolerance FRSD does (Appendix E.8). This distinguishes persistent correction from a penalty that vanishes with the current gap. Appendix E.6 relates terminal disparity to queue increments over finite windows through pathwise bounds.

4.3 Regret lower bound

The rates so far concern learning around the oracle prices. Terminal fairness raises a further issue: the oracle prices themselves may put group averages outside the permitted tolerance. Two instance classes in the linear-demand model separate the resulting costs. In the first, all groups share an oracle price, demand noise is Gaussian with zero conditional mean, and the class contains the fixed-parameter family from Appendix A.4. In the second, the common slope is fixed at $b_0 < 0$, the oracle prices of two groups lie in the interior and differ by $\Delta^\star > \tau$, deterministic arrival counts satisfy the group-coverage assumption, and only conditionally mean-zero noise is required.

For a fixed policy–estimator pair, let L_T denote its supremum slope-estimation MSE over the first class and R_T its worst-case expected regret over the union of both classes. For the second class, write $B_T(\tau, \Delta^\star) = c_f T(\Delta^\star - \tau)^2$, where $c_f > 0$ depends on revenue curvature and the group-coverage bound.

Theorem 4 (Regret lower bound)**.** *There exists $c_L > 0$ such that every policy–estimator pair that is terminally τ-price-fair almost surely on the separated subclass satisfies*

$$R_T \geq \max\left\{\frac{c_L}{L_T},\, B_T(\tau, \Delta^\star)\right\}. \tag{16}$$

The two terms reflect different requirements. Learning the slope incurs the first cost, while satisfying terminal fairness on the separated class forces departures from the oracle prices even when demand is known. That second cost is the fairness cost introduced in Section 3. Appendix A.4 derives its exact value and constructs prices that attain it by charging a constant price within each group; the group-coverage bound then yields $B_T(\tau, \Delta^\star)$.

Because the worst case over the union includes both constructions, regret is bounded below by the larger cost. For fixed τ and $\Delta^\star > \tau$, the fairness term grows linearly with the horizon, ruling out sublinear worst-case regret under the stated fairness requirement on this class. This cost comes from the separation of the oracle prices, not from uncertainty about demand. Appendix A gives the full proof.

5 Experiments

Environments and protocol. We begin with two environments to measure how the fairness correction affects price disparity, revenue, and demand estimation. The Synthetic environment combines Gaussian context features with a group effect that separates the oracle prices. In the

Retail simulation, we use prices and contexts from UCI Online Retail (Chen, 2015) to calibrate demand intercepts. We then set the price slope and simulate demand at the offered prices for UK and non-UK groups. Both environments add Gaussian demand noise with standard deviation 0.10 and clip negative demand to zero. A third, bounded linear-demand environment tests the mechanism under the analytical model. Appendix B gives the details.

To isolate the correction, each FRSD configuration is compared with RSD using the same estimation and exploration settings; we call this the matched RSD baseline. Within each paired run, all policies use the same arrivals, demand noise, and shock signs; the signs are drawn independently across arrivals with equal probabilities. For the Synthetic and Retail comparisons, each run samples $T = 2000$ arrivals without replacement. We evaluate two fixed Synthetic configurations, C1 and C2, over 100 paired runs. For Retail, we select a configuration, C3, on separate training runs and evaluate it on 500 test runs. Table 4 and Appendix B.3 give the parameters and selection protocol.

Metrics. Fairness is measured by terminal group-price disparity, defined in equation (2). For Retail, we also report time-averaged disparity: the absolute difference between the two groups' cumulative mean prices at each arrival, averaged over the experiment. Before both groups appear, the unobserved group's mean equals the market mean. Alongside these measures, we report realized revenue per arrival, $T^{-1}\sum_t p_t D_t$, and terminal squared slope error, whose across-run average estimates the slope MSE. Table 1 gives means and standard deviations; Figure 1 shows the Retail means with 95% confidence intervals.

Table 1: Independent-shock results over 100 paired Synthetic runs and 500 paired Retail runs. Entries are mean $\pm$ across-seed standard deviation. Higher revenue is better; lower slope MSE and disparity are better. Squared slope error is displayed in units of 10^{-3}. RSD is reported separately for each configuration.

Environment	Policy	Revenue/arrival	Slope MSE (10^{-3})	Disparity
Synthetic C1	RSD	8.846 ± 0.600	4.980 ± 5.164	1.088 ± 0.231
	FRSD	9.057 ± 0.541	5.885 ± 5.540	0.089 ± 0.030
Synthetic C2	RSD	9.254 ± 0.413	2.125 ± 3.113	1.196 ± 0.188
	FRSD	9.028 ± 0.601	2.830 ± 3.867	0.040 ± 0.021
Retail C3	RSD	0.912 ± 0.027	0.473 ± 0.519	0.116 ± 0.080
	FRSD	0.911 ± 0.028	0.494 ± 0.543	0.094 ± 0.065

Fairness and revenue in Synthetic. FRSD substantially reduces disparity in both Synthetic configurations. Under C1, mean disparity falls by 92% relative to RSD and mean revenue rises by 2.4%. The disparity reduction occurs in all 100 paired runs, while revenue improves in 61 of them.

Under C2, FRSD reduces mean disparity by 97% but earns less than RSD. FRSD's slope MSE is higher than RSD's in both configurations (Table 1).

Retail. Under Retail C3, FRSD reduces mean terminal disparity by 19.0% relative to RSD, with a paired 95% confidence interval of 14.5–23.5% for the reduction. Mean revenue decreases by 0.17% and slope MSE increases by 4.3% (Figure 1), leaving revenue and estimation performance close to the baseline. The fairness gain is also reflected in time-averaged disparity, which is 14.1% lower than RSD's. Appendices B.3 and B.4 report alternative parameter settings and examine the signed differences and queue activity behind these outcomes.

Does retaining past disparities help? We replace the queue-weighted penalty by either a fixed penalty (Fixed) or a current-gap penalty (Current), whose coefficient is proportional to the current excess group-to-market gap. The remaining pricing and estimation steps are unchanged. Each correction method is evaluated on the same 17-point strength grid and 80 training runs. On the training runs, we select the setting with the lowest mean disparity among those whose mean revenue decreases by at most 1% and slope MSE increases by at most 10% relative to matched RSD. Evaluation uses 200 separate paired test runs. We also apply a pre-specified 5% revenue-cost target. Appendix B.8 gives the full protocol and results.

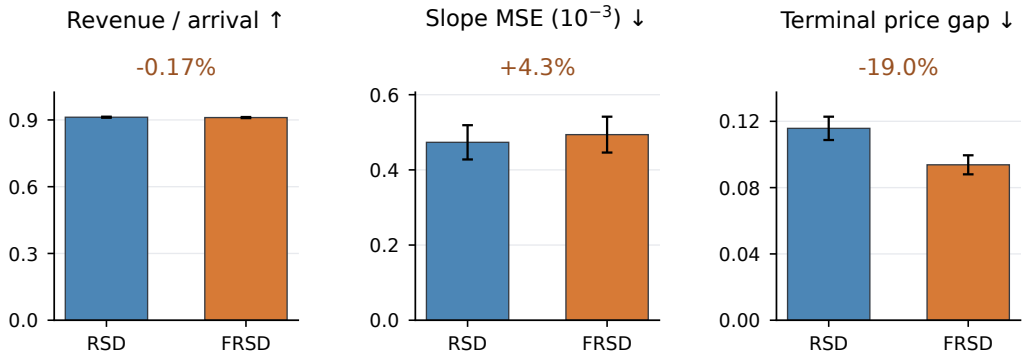

Figure 1: Retail outcomes under C3. Bars show means over 500 paired runs; whiskers are 95% Student-t intervals for each mean. The labels above the panels give FRSD's percentage changes relative to RSD. The interval for the percentage reduction in the text uses a paired linearization of the ratio of means (Appendix B.3).

Table 2: FRSD relative to the training-selected fixed and current-gap penalties on 200 paired test runs. Entries are percentage changes in test means under the primary training rule. Lower disparity and MSE, and higher revenue, are better. C1/C2/C3 specify the exploration and regression settings; correction strengths are selected separately for this comparison.

Environment	Comparator	Revenue	Slope MSE	Disparity
Synthetic C1	Fixed	$+0.18\%$	-1.2%	-70.8%
	Current	$+1.12\%$	-1.0%	-75.7%
Synthetic C2	Fixed	-1.19%	-9.3%	$+13.6\%$
	Current	-1.60%	-15.4%	$+40.6\%$
Retail C3	Fixed	$+0.07\%$	$+0.2\%$	-16.3%
	Current	$+0.10\%$	$+1.5\%$	-3.8%
Bounded linear	Fixed	$+0.70\%$	$+7.6\%$	-72.4%
	Current	$+0.23\%$	$+2.7\%$	-17.5%

Under C1, FRSD reduces mean disparity by 70.8% relative to the fixed penalty and by 75.7% relative to the current-gap penalty, with higher mean revenue and slightly lower slope MSE (Table 2). Retaining past disparities thus improves on both alternatives in this setting. Under C2, the simpler penalties yield smaller disparities under the primary rule; with the 5% training target, FRSD instead yields smaller disparity, at a 1.64% test revenue cost relative to RSD. In the Retail simulation, FRSD reduces mean disparity relative to the fixed penalty. Its difference from the current-gap penalty is -0.0036, with a paired 95% confidence interval $[-0.0077, 0.0005]$ that includes zero. The benefit of accumulated disparities therefore depends on the exploration setting and the permitted revenue cost.

A bounded linear-demand check. The bounded linear-demand environment uses bounded contexts and noise, feasible oracle prices, and untruncated linear demand. Relative to matched RSD, FRSD reduces mean disparity by 90.8% and attains higher mean revenue; slope MSE is 13.3% higher. The price and regret identities agree numerically to 1.2×10^{-8}; Appendix B.8 reports their components and the fair benchmark. Two fixed-parameter experiments examine Theorem 3 over longer horizons. Under a repeated-block design satisfying its regression and coverage conditions, mean disparity under zero-tolerance FRSD falls from 0.0309 at 500 arrivals to 0.0024 at 32,000. Mean disparity also decreases with the horizon when arrivals are randomly sampled (Appendix B.9).

Parameter sensitivity and perturbation scheme. To examine the role of the tolerance, we test five values with the other C1 parameters fixed. In the Synthetic environment, tighter tolerances give smaller mean disparities and lower mean revenue, tracing the fairness–revenue trade-off within a single sweep. In the Retail simulation, mean disparity stays above that of RSD at every tested

tolerance, so changing the tolerance alone does not reproduce the gain under C3. Figure 4 and Table 8 give the full sweep, and Appendix B.3 reports the joint parameter choice used for C3.

We next examine the role of the perturbation scheme in demand estimation. Holding arrivals and noise fixed, we compare the independent shocks of Section 4 with deterministic alternating signs under C1 and C2. In the Synthetic environment, FRSD's slope MSE is lower than RSD's under alternating signs but higher under independent shocks (Table 7), showing that the estimation ranking depends on the perturbation scheme even when the correction rule is unchanged.

6 Discussion

The tolerance and fairness weight play different roles in FRSD. The tolerance τ determines which running gaps accumulate in the queue, whereas λ_f determines how strongly that queue affects the next offer. Their joint effect enters the price correction through $\lambda_f Q_{g_t,t}$, linking the history of group disparities to the current revenue objective.

The experiments show that the fairness correction can reduce disparities while preserving much of RSD's revenue performance. FRSD achieves this with higher mean revenue under Synthetic C1; under Retail C3, it reduces both terminal and time-averaged disparity with a small revenue decrease.

Our criterion concerns the prices offered to the chosen groups. Extending the analysis to individual welfare and downstream outcomes would connect these price differences to their consequences for customers. Nonlinear demand, group-specific slopes, strategic customers, and inventory constraints are further directions; field evaluation would require causal demand validation, an agreed fairness objective, and customer protections.

7 Conclusion

FRSD uses accumulated price disparities to guide later offers while preserving the independent variation needed to learn demand. The terminal-fairness bound quantifies how tolerance and experiment length control disparity under the stated bounded-model and design conditions. The regret analysis accounts for estimation error and subsequent price adjustments, while the fair benchmark identifies the revenue cost of the fairness requirement itself. Longer-horizon experiments show the decline in disparity predicted at zero tolerance. At finite horizons, FRSD improves fairness over simpler penalties under Synthetic C1 and over matched RSD under Retail C3, where revenue and slope MSE remain close to the baseline.

AI use statement

Generative AI tools assisted with writing and editing, literature and citation checks, and LaTeX preparation. They also supported the implementation and execution of the experiments and consistency checks of the reported results and mathematical arguments. The authors are responsible for reviewing the final manuscript, mathematical arguments, code, and experimental claims.

Ethics statement

This work uses public transaction data to calibrate simulated interventions; it collects no new human-subject data and deploys no pricing policy. Poor choices of audited groups, tolerances, or welfare criteria could nevertheless legitimize harmful pricing. Any deployment would require legal and policy review, ongoing monitoring, appropriate transparency, and meaningful recourse.

Reproducibility statement

The proofs are in Appendices A and E.7; Appendix B specifies the environments, configurations, seeds, metrics, and comparison protocols. Per-run summaries, saved inputs and trajectories, data checksums, and generation scripts link these specifications to the results. Mechanism comparisons fix selections using training data and report every evaluated strength on separate test seeds. Both horizon experiments use fixed parameters and report all runs.

REFERENCES

Parand Alamdari, Toryn Q. Klassen, Elliot Creager, and Sheila A. McIlraith. Remembering to be fair: Non-markovian fairness in sequential decision making. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 906–920. PMLR, 2024.

Gah-Yi Ban and N. Bora Keskin. Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science*, 67(9):5549–5568, 2021. doi: 10.1287/mnsc.2020.3680.

MohammadHossein Bateni, Yiwei Chen, Dragos Florin Ciocan, and Vahab Mirrokni. Fair resource allocation in a volatile marketplace. *Operations Research*, 70(1):288–308, 2022. doi: 10.1287/opre.2020.2049.

Martino Bernasconi, Matteo Castiglioni, and Andrea Celli. No-regret is not enough! bandits with general constraints through adaptive regret minimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 3877–3898. PMLR, 2025.

Dimitris Bertsimas, Vivek F. Farias, and Nikolaos Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011. doi: 10.1287/opre.1100.0865.

Dimitris Bertsimas, Vivek F. Farias, and Nikolaos Trichakis. On the efficiency–fairness trade-off. *Management Science*, 58(12):2234–2250, 2012. doi: 10.1287/mnsc.1120.1549.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009. doi: 10.1287/opre.1080.0640.

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

Daqing Chen. Online retail. UCI Machine Learning Repository, 2015.

Jiajun Chen, Jin Tian, and Christopher John Quinn. Causal logistic bandits with counterfactual fairness constraints. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9146–9175. PMLR, 2025.

Xin Chen, Zexing Xu, Zishuo Zhao, and Yuan Zhou. Personalized pricing with group fairness constraint. In *Proceedings of the 6th ACM Conference on Fairness, Accountability, and Transparency*, pp. 1520–1530, 2023. doi: 10.1145/3593013.3594097.

Yifang Chen, Alex Cuellar, Haipeng Luo, Jignesh Modi, Heramb Nemlekar, and Stefanos Nikolaidis. Fair contextual multi-armed bandits: Theory and experiments. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pp. 181–190. PMLR, 2020.

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

Maxime C. Cohen, Adam N. Elmachtoub, and Xiao Lei. Price discrimination with fairness constraints. *Management Science*, 68(12):8536–8552, 2022. doi: 10.1287/mnsc.2022.4317.

Maxime C. Cohen, Sentao Miao, and Yining Wang. Dynamic pricing with fairness constraints. *Operations Research*, 73(6):3027–3043, 2025. doi: 10.1287/opre.2023.0123.

Shantanu Das, Swapnil Dhamal, Ganesh Ghalme, Shweta Jain, and Sujit Gujar. Individual fairness in feature-based pricing for monopoly markets. In *Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 486–495. PMLR, 2022.

Arnoud V. den Boer. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1):1–18, 2015. doi: 10.1016/j.sorms.2015.03.001.

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

Hengquan Guo and Xin Liu. Stochastic constrained contextual bandits via lyapunov optimization based estimation to decision framework. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 2204–2231. PMLR, 2024.

J. Michael Harrison, N. Bora Keskin, and Assaf Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586, 2012. doi: 10.1287/mnsc.1110.1426.

Matthew Joseph, Michael Kearns, Jamie H. Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Matthew Joseph, Michael J. Kearns, Jamie H. Morgenstern, Seth Neel, and Aaron Roth. Meritocratic fairness for infinite and contextual bandits. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 158–163, 2018. doi: 10.1145/3278721.3278764.

N. Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014. doi: 10.1287/opre.2014.1294.

Mila Nambiar, David Simchi-Levi, and He Wang. Dynamic learning and pricing with model misspecification. *Management Science*, 65(11):4980–5000, 2019. doi: 10.1287/mnsc.2018.3194.

David Simchi-Levi and Chonghuan Wang. Pricing experimental design: Causal effect, expected revenue and tail risk. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 31788–31799. PMLR, 2023.

Lequn Wang, Yiwei Bai, Wen Sun, and Thorsten Joachims. Fairness of exposure in stochastic bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10686–10696. PMLR, 2021.

Jianyu Xu and Yu-Xiang Wang. Pricing with contextual elasticity and heteroscedastic valuation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55286–55304. PMLR, 2024.

Jianyu Xu, Dan Qiao, and Yu-Xiang Wang. Doubly fair dynamic pricing. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 9941–9975. PMLR, 2023. URL `https://proceedings.mlr.press/v206/xu23i.html`.

Yuancheng Xu, Chenghao Deng, Yanchao Sun, Ruijie Zheng, Xiyao Wang, Jieyu Zhao, and Furong Huang. Adapting static fairness to sequential decision-making: Bias mitigation strategies towards equal long-term benefit rate. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54962–54982. PMLR, 2024.

Zongsen Yang, Xingyu Fu, Pin Gao, and Ying-Ju Chen. Fairness regulation of prices in competitive markets. *Manufacturing & Service Operations Management*, 26(5):1897–1917, 2024. doi: 10.1287/msom.2022.0552.

A Proofs and Additional Results

The proofs follow the pricing rule and the quantities it controls. We first derive the corrected price. We then use the independent shock to prove the estimation bound and regret decomposition, and the queue recursion to bound running gaps. The lower-bound constructions separate learning costs from the cost of terminal fairness. We next derive the design-based greedy-price error bound and its relation to price adjustments. The final proof returns to the regret decomposition to establish the conditional non-worsening rates.

Guide to the mathematical extensions. The proofs in this section establish the main-text statements. The later mathematical sections develop their consequences for the implemented prices, estimation updates, and fairness benchmark:

- Appendix C completes the price-correction formula when the constraints are active. It then gives a geometric condition for improvement in oracle error and an accumulated-error condition that preserves the regret order with projected prices. These results extend the analysis of Section 4 using the same pricing rule.
- Appendix D derives explicit moments and probability bounds for the shock-product estimator. A conditional sub-Gaussian extension includes Gaussian demand noise. The final calculations trace slope and prediction errors through the ridge update and into the greedy-price term of the regret decomposition.
- Appendix E expresses each virtual queue through the observed price history. Window identities, group-weighting relations, and mean updates describe how earlier offers determine the next correction. Appendix E.7 then combines the queue recursion with the implemented prices to prove Theorem 3, including active projection.
- Appendix F characterizes the benchmark in Section 3. It reduces fair pricing to adjustments of group-average oracle prices, handles active price bounds through clipped shifts, and provides certificates for the optimal revenue cost. The two-group formula connects this construction to the fairness term of the lower bound.
- Appendix G expands the adaptive likelihood and testing arguments behind Theorem 4. It records the fixed instance neighborhood and constants, and links the lower-bound coefficient to the attained fairness cost.

The price, queue, and benchmark identities hold pathwise, while the estimation and concentration results use the stated conditional noise and randomization assumptions. The known-demand benchmark defines the cost of a fairness requirement; the adaptive policy learns demand as prices are offered.

A.1 Notation and assumptions

We collect the notation used below. Let $\mathcal{F}_{t-1}$ denote the history before arrival t and $\mathcal{H}_t = \sigma(\mathcal{F}_{t-1}, x_t, g_t)$. Shock magnitudes are $\delta_t = t^{-\alpha}$ with $\alpha \in [0, 1/4]$, and $S_T = \sum_{t=1}^{T} \delta_t^2$. Group and market averages through arrival t are $\bar{p}_{g,t}$ and $\bar{p}_t$, and the oracle price $p_t^\star$ is defined in equation (3). In the lower-bound constructions, $\mathbb{P}_h$ denotes the joint law of the observed experiment under instance h and $\bar{R}_h(T)$ the corresponding expected regret; the quantities L_T and R_T are defined in Section 4.

Table 3: Notation used in the appendix. Symbols follow the main text.

Symbol	Meaning
$[\underline{p}, \overline{p}]$	feasible price interval
b, $\widehat{b}_t$	demand slope and its estimate at arrival t
θ, η_g, $\mu_t = \theta^\top x_t + \eta_{g_t}$	context, group, and combined effects
ε_t, $\bar{\varepsilon}$	demand noise and its bound
$\delta_t = t^{-\alpha}$, $S_T = \sum_{t=1}^{T} \delta_t^2$	shock magnitude and squared-shock sum
$\Delta_t \in \{-\delta_t, +\delta_t\}$	exogenous price shock
u_t, $\widetilde{p}_t$, p_t	corrected, projected, and posted prices
$Q_{g,t}$, τ, λ_f	virtual queue, tolerance, and fairness weight
$\Delta(\mathbf{p})$, $p_t^\star$	group-price disparity and oracle price
$\bar{R}(T)$, $\mathrm{PoF}_T(\tau)$	expected regret and price-of-fairness benchmark

Assumptions. We collect the conditions used in the main text. Theorem 1 uses (A1)–(A4); the lower-bound constructions use the instance-specific conditions in Section 4.3, with Gaussian noise for the learning family and (A5) for group coverage.

(A1) *Bounded model.* Contexts satisfy $x_t \in [0,1]^d$, as in Section 3. There are constants $\kappa_b, \kappa_\theta, \kappa_\eta, \bar{\varepsilon} < \infty$ such that $|b| \le \kappa_b$, $\|\theta\|_1 \le \kappa_\theta$, $|\eta_g| \le \kappa_\eta$ for every g, and $|\varepsilon_t| \le \bar{\varepsilon}$ almost surely. Write $M_\mu := \kappa_\theta + \kappa_\eta$ and $P := \max\{|\underline{p}|, |\overline{p}|\}$.

(A2) *Feasible oracle.* The oracle price $p_t^\star$ lies in $[\underline{p}, \overline{p}]$ for every t.

(A3) *Noise.* $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}, x_t, g_t] = 0$ for every t.

(A4) *Exogenous shocks.* Conditional on $\mathcal{H}_t$, the shock Δ_t is uniform on $\{-\delta_t, +\delta_t\}$ and independent of ε_t.

(A5) *Group coverage.* $N_g(T) \ge \pi_{\min} T$ for every group g, with a fixed $\pi_{\min} > 0$.

A.2 Proof of Proposition 1

Write $c_t = -\widehat{b}_t > 0$ and $q_t = \lambda_f Q_{g_t,t} \ge 0$. The unconstrained greedy price is

$$v_t = -\frac{\widehat{\mu}_t(g_t, x_t)}{2\widehat{b}_t} = \frac{\widehat{\mu}_t(g_t, x_t)}{2c_t}, \tag{17}$$

and hence $\widehat{\mu}_t(g_t, x_t) = 2c_t v_t$. The objective in equation (7) has second derivative $-2(c_t + q_t) < 0$, so its maximizer is unique. If the projection in equation (8) is inactive, this maximizer lies in $[\underline{p} + \delta_t, \overline{p} - \delta_t]$, strictly inside the original price interval because $\delta_t > 0$. Its first-order condition is therefore

$$-2c_t u_t + 2c_t v_t - 2q_t(u_t - \bar{p}_{t-1}) = 0.$$

Solving for u_t gives

$$u_t = \frac{2c_t v_t + 2q_t \bar{p}_{t-1}}{2(c_t + q_t)} = \frac{c_t}{c_t + q_t} v_t + \frac{q_t}{c_t + q_t} \bar{p}_{t-1}, \tag{18}$$

which is equation (10). The weights are nonnegative and sum to one. Holding the estimates and market mean fixed, increasing $\lambda_f Q_{g_t,t}$ places more weight on the market mean.

A.3 Proof of Theorem 1

Estimation error. Let $\mathcal{H}_t = \sigma(\mathcal{F}_{t-1}, x_t, g_t)$ denote the information available before drawing the shock, and write $\mu_t = \theta^\top x_t + \eta_{g_t}$. Both $\widetilde{p}_t$ and μ_t are $\mathcal{H}_t$-measurable. Since $\Delta_t^2 = \delta_t^2$, the centered shock–demand product is

$$M_t := \Delta_t D_t - b\delta_t^2 = \Delta_t(b\widetilde{p}_t + \mu_t + \varepsilon_t).$$

The zero conditional mean of Δ_t and its conditional independence from ε_t give $\mathbb{E}[M_t \mid \mathcal{H}_t] = 0$. Bounded parameters, contexts, prices, and noise also give a constant $C > 0$, independent of t and T,

such that $\mathbb{E}[M_t^2 \mid \mathcal{H}_t] \le C\delta_t^2$. For $s < t$, M_s is measurable with respect to $\mathcal{F}_{t-1}$, so $\mathbb{E}[M_s M_t] = 0$. Consequently,

$$\mathbb{E}\left[\left(\sum_{t=1}^{T} M_t\right)^2\right] = \sum_{t=1}^{T} \mathbb{E}[M_t^2] \le C S_T.$$

Before projection, the estimator is $b + S_T^{-1} \sum_{t=1}^{T} M_t$. Projection onto an interval containing b cannot increase its distance from b, hence

$$\mathbb{E}[(\widehat{b}_{T+1} - b)^2] \le C/S_T. \tag{19}$$

Regret decomposition. Completing the square in the concave quadratic revenue from equation (1) around the feasible oracle price gives

$$r_t(p_t^\star) - r_t(p) = (-b)(p - p_t^\star)^2. \tag{20}$$

Substituting $p_t = \widetilde{p}_t + \Delta_t$ and conditioning on $\mathcal{H}_t$ yields

$$\begin{aligned}\mathbb{E}[(p_t - p_t^\star)^2 \mid \mathcal{H}_t] &= (\widetilde{p}_t - p_t^\star)^2 + 2(\widetilde{p}_t - p_t^\star)\mathbb{E}[\Delta_t \mid \mathcal{H}_t] + \mathbb{E}[\Delta_t^2 \mid \mathcal{H}_t] \\ &= (\widetilde{p}_t - p_t^\star)^2 + \delta_t^2.\end{aligned}$$

Taking expectations and summing over t proves equation (12).

Running price differences. The queue recursion implies

$$Q_{g,t} - Q_{g,t-1} \ge |\bar{p}_{g,t-1} - \bar{p}_{t-1}| - \tau. \tag{21}$$

Summing from $t = 2$ to T and using $Q_{g,1} = 0$ gives

$$\sum_{t=2}^{T} |\bar{p}_{g,t-1} - \bar{p}_{t-1}| \le (T-1)\tau + Q_{g,T}.$$

Division by $T - 1$ proves equation (13) for $T \ge 2$.

A.4 Proof of Theorem 4

The two instance classes in Section 4 isolate different sources of regret. We first construct a family in which learning the demand slope requires departures from a revenue-optimal price, then derive the revenue cost of terminal fairness on the separated class. Since L_T and R_T are suprema over classes containing the learning family, they dominate its estimation risks and expected regrets. The policy and estimator are fixed functions of the observed history and any parameter-independent randomization; they do not know which instance generates demand.

Learning instances. Choose b_0 in the interior of $[\underline{b}, \overline{b}]$ and a nonzero price p_0 in the interior of $[\underline{p}, \overline{p}]$. Set $x_t = 0$ and $\theta = 0$, and fix a deterministic group sequence satisfying the coverage assumption. For $h \in [0, h_0]$, define

$$b(h) = b_0 + h, \qquad \eta_g(h) = -(2b_0 + h)p_0 \quad \text{for every } g, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2),$$

where $\sigma > 0$ is fixed. Choose the fixed constant $h_0 > 0$ sufficiently small that the slopes remain in their known negative interval and the oracle prices

$$p^\star(h) = \frac{(2b_0 + h)p_0}{2(b_0 + h)} = p_0 - \frac{h p_0}{2(b_0 + h)}$$

remain interior. All coefficients are bounded on this fixed interval. Each instance has the same oracle price across groups, although this price varies with h. Its mean demand, denoted by $m_h(p)$, satisfies

$$m_h(p) - m_0(p) = h(p - p_0), \qquad p^\star(0) = p_0.$$

Thus pricing at p_0 provides no information for distinguishing these instances. This construction makes explicit the learning–revenue trade-off studied in random-shock pricing (Simchi-Levi & Wang, 2023).

Information and regret. Write $\mathbb{P}_h$ for the joint law of the observed experiment and the policy–estimator randomization under instance h, and $\bar{R}_h(T)$ for its expected regret. First, R_T has a positive lower bound independent of T. Indeed, the first price has the same distribution under $h = 0$ and $h = h_0$. Let $d_0 = |p^\star(h_0) - p_0| > 0$ and $c_b = -(b_0 + h_0) > 0$. The inequality

$$(p - p_0)^2 + (p - p^\star(h_0))^2 \geq d_0^2/2$$

and the nonnegativity of every subsequent oracle loss imply

$$R_T \geq \tfrac{1}{2}\{\bar{R}_0(T) + \bar{R}_{h_0}(T)\} \geq c_b d_0^2/4 =: r_0 > 0.$$

Set $\kappa = \min\{r_0 h_0^2, \sigma^2(-b_0)/4\} > 0$ and choose $h = \sqrt{\kappa/R_T}$, which lies in $(0, h_0]$. The conditional action rule is the same under both instances. The chain rule for relative entropy and the Gaussian conditional demand distributions therefore give

$$\begin{aligned}\mathrm{KL}(\mathbb{P}_0 \| \mathbb{P}_h) &= \frac{h^2}{2\sigma^2}\mathbb{E}_0\left[\sum_{t=1}^{T}(p_t - p_0)^2\right] \\ &= \frac{h^2 \bar{R}_0(T)}{2\sigma^2(-b_0)} \leq \frac{\kappa}{2\sigma^2(-b_0)} \leq \frac{1}{8}.\end{aligned}$$

Pinsker's inequality bounds the total variation distance by $1/4$. For any final slope estimator $\widehat{b}$, use the test that selects instance h when $\widehat{b} \geq b_0 + h/2$ and instance 0 otherwise. An incorrect decision entails squared estimation error at least $h^2/4$. The sum of the two testing error probabilities is at least one minus total variation, so

$$\begin{aligned}L_T &\geq \tfrac{1}{2}\{\mathbb{E}_0[(\widehat{b} - b_0)^2] + \mathbb{E}_h[(\widehat{b} - (b_0 + h))^2]\} \\ &\geq \frac{h^2}{8}\left(1 - \|\mathbb{P}_0 - \mathbb{P}_h\|_{\mathrm{TV}}\right) \geq \frac{3\kappa}{32 R_T}.\end{aligned}$$

It follows that $R_T \geq c_L / L_T$ with $c_L = 3\kappa/32 > 0$, independent of the policy, estimator, and horizon.

Fairness cost. Consider the two-group class with deterministic counts $N_A + N_B = T$, common slope $b_0 < 0$, and fixed interior oracle prices $p_A^\star > p_B^\star$ separated by $\Delta^\star > \tau$. Let $\bar{p}_A$ and $\bar{p}_B$ be the realized terminal group-average prices, and write $d_A = \bar{p}_A - p_A^\star$ and $d_B = \bar{p}_B - p_B^\star$. On every terminally τ-price-fair path,

$$d_B - d_A = \Delta^\star - (\bar{p}_A - \bar{p}_B) \geq \Delta^\star - \tau.$$

Jensen's inequality within each group, followed by weighted Cauchy–Schwarz, gives

$$\begin{aligned}\sum_{t=1}^{T}(p_t - p_{g_t}^\star)^2 &\geq N_A d_A^2 + N_B d_B^2 \\ &\geq \frac{(d_B - d_A)^2}{1/N_A + 1/N_B} \geq \frac{N_A N_B}{T}(\Delta^\star - \tau)^2.\end{aligned}$$

Multiplying by $-b_0$ and taking expectations proves the finite-sample revenue lower bound. Since $N_A, N_B \geq \pi_{\min} T$, it is at least

$$(-b_0)\pi_{\min}^2 T(\Delta^\star - \tau)^2 = B_T(\tau, \Delta^\star),$$

with $c_f = (-b_0)\pi_{\min}^2$. The finite-sample bound is also attainable when demand is known: use constant prices within groups with shifts $d_A = -N_B(\Delta^\star - \tau)/T$ and $d_B = N_A(\Delta^\star - \tau)/T$. These prices remain between the two interior oracle prices and have gap τ. Thus, on this class, the benchmark in equation (5) satisfies $\mathrm{PoF}_T(\tau) = (-b_0)N_A N_B(\Delta^\star - \tau)^2/T$.

Worst-case regret over the union is at least as large as either subclass bound. Taking their maximum proves equation (16).

A.5 LEARNING ERROR AND PRICE ADJUSTMENTS

We derive Theorem 2 from the ridge update under its stated design condition. This condition controls the greedy-price error; it is not required for Theorem 1 or the lower bound.

Write $\mu_t = \phi_t^\top \gamma$, with fixed true context and group coefficients γ, and let P bound the absolute feasible prices. After n observations the update is

$$\widehat{\gamma}_{n+1} = V_n^{-1} \sum_{s=1}^{n} \phi_s (D_s - \widehat{b}_{n+1} p_s), \qquad V_n = \Lambda + \sum_{s=1}^{n} \phi_s \phi_s^\top.$$

For $n \geq n_0$, substitution of the linear demand model gives

$$\widehat{\gamma}_{n+1} - \gamma = V_n^{-1} \left\{ \sum_{s=1}^{n} \phi_s \varepsilon_s + (b - \widehat{b}_{n+1}) \sum_{s=1}^{n} \phi_s p_s - \Lambda \gamma \right\}.$$

The noise sum is a vector martingale: ϕ_s is observed before the demand noise and its conditional mean is zero. Cross terms at different arrivals therefore have zero expectation, and

$$\mathbb{E} \left\| \sum_{s=1}^{n} \phi_s \varepsilon_s \right\|^2 \leq n L^2 \bar{\varepsilon}^2, \qquad \left\| \sum_{s=1}^{n} \phi_s p_s \right\| \leq nLP.$$

The design condition gives $\|V_n^{-1}\| \leq 1/(\kappa n)$ pathwise. Using the squared-norm inequality for three summands and the slope bound with constant C_b yields

$$\mathbb{E}\|\widehat{\gamma}_{n+1} - \gamma\|^2 \leq \frac{3L^2\bar{\varepsilon}^2}{\kappa^2 n} + \frac{3L^2 P^2 C_b}{\kappa^2 S_n} + \frac{3\|\Lambda\gamma\|^2}{\kappa^2 n^2}. \tag{22}$$

No independence between the slope estimate, the price history, and the ridge fit is used: the price-history factor is bounded pathwise before taking expectations.

For the current query, $\|\phi_{n+1}\| \leq L$ implies $\mathbb{E}[(\widehat{\mu}_{n+1} - \mu_{n+1})^2] \leq L^2 \mathbb{E}\|\widehat{\gamma}_{n+1} - \gamma\|^2$, even when that query depends on past observations. With $\beta_b = -\bar{b} > 0$, feasible oracle prices give

$$\mathbb{E}[(v_t - p_t^\star)^2] \leq \frac{\mathbb{E}[(\widehat{\mu}_t - \mu_t)^2] + 4P^2 \mathbb{E}[(\widehat{b}_t - b)^2]}{2\beta_b^2}.$$

Summing equation (22), the slope bound, and the finite initial terms proves

$$E_T \leq K_0 + K_1 \log T + K_2 \sum_{n=1}^{T-1} S_n^{-1}$$

for constants independent of T. The finite prefix is bounded because the ridge matrix is positive definite, demand and regressors are bounded, and the estimated slope stays in its fixed negative interval. For $\alpha = 0$ the sum is $O(\log T)$; for $0 < \alpha \leq 1/4$ it is $O(T^{2\alpha})$. Since $S_T = \Theta(T^{1-2\alpha})$, both cases imply $E_T = O(S_T)$.

For alternative (ii), the random design yields the same rate. Let q be the regressor dimension and $\Sigma = \mathbb{E}[\phi_t \phi_t^\top] \succeq \kappa I$. A $1/4$-net of the unit sphere can be chosen with at most 9^q points. For a fixed unit vector u, the independent variables $(u^\top \phi_s)^2$ lie in $[0, L^2]$. The bounded-variable exponential inequality gives

$$\Pr\left(\left| \frac{1}{n} \sum_{s=1}^{n} (u^\top \phi_s)^2 - u^\top \Sigma u \right| > \frac{\kappa}{4} \right) \leq 2 \exp\left(-\frac{n\kappa^2}{8L^4} \right).$$

For completeness, the inequality follows from $\mathbb{E}e^{s(X - \mathbb{E}X)} \leq e^{s^2 L^4/8}$ for $X \in [0, L^2]$, independence, exponential Markov, and optimization over s; applying the same argument to $-X$ gives the two tails. A maximal $1/4$-separated subset of the sphere is a net and has cardinality at most 9^q by disjoint-ball volume comparison. For a symmetric matrix M, approximating a maximizing unit vector by this net gives $\|M\| \leq 2 \max_u |u^\top M u|$. A union bound therefore yields

$$\Pr\left\{ \lambda_{\min}(V_n) < \frac{\kappa n}{2} \right\} \leq 2\, 9^q \exp\left(-\frac{n\kappa^2}{8L^4} \right).$$

On the complementary event, the preceding ridge argument holds with $\kappa/2$. The squared noise sum and slope error multiplied by the event indicator are bounded in expectation by their unrestricted second moments, so conditioning on the whole design is unnecessary. On the exceptional event, $V_n \succeq \Lambda \succ 0$, bounded demands and prices, and bounded slope estimates give $\|\widehat{\gamma}_{n+1}-\gamma\| \le K(n+1)$. Its contribution to mean squared error is consequently at most $K'(n+1)^2 e^{-cn}$ for fixed constants. This sequence is summable. It only changes the constant in the bound for E_T, proving alternative (ii).

To prove the revenue statement, expand $\widetilde{p}_t - p_t^\star = (v_t - p_t^\star) + (\widetilde{p}_t - v_t)$. The exact decomposition in Theorem 1 gives

$$\frac{\bar{R}(T)}{-b} = E_T + A_T + S_T + 2\sum_{t=1}^{T} \mathbb{E}[(v_t - p_t^\star)(\widetilde{p}_t - v_t)].$$

Cauchy–Schwarz on the product of time and probability spaces gives $|\sum_t \mathbb{E}[(v_t - p_t^\star)(\widetilde{p}_t - v_t)]| \le \sqrt{E_T A_T}$, proving equation (14). This last identity and bound require only Theorem 1; the design condition is used to bound E_T. In particular,

$$\bar{R}(T) \le (-b)\{(\sqrt{E_T} + \sqrt{A_T})^2 + S_T\}.$$

If $A_T = O(S_T)$ the displayed rate follows; larger sustained adjustments enter through the same bound. Here A_T measures the algorithm's price adjustments, including projection, whereas $\mathrm{PoF}_T(\tau)$ measures the optimal cost of the fairness requirement.

A.6 Conditional learning and regret rates

Corollary 1 (Conditional learning and regret rates)**.** *Under the assumptions of Theorem 1, suppose the price projections are inactive, the greedy prices satisfy $\sum_{t=1}^{T} \mathbb{E}[(v_t - p_t^\star)^2] \le C_R S_T$, and the correction satisfies $|u_t - p_t^\star| \le |v_t - p_t^\star|$ almost surely for all t. Then*

$$\mathbb{E}[(\widehat{b}_{T+1} - b)^2] = O(S_T^{-1}), \qquad \bar{R}(T) = O(S_T). \tag{23}$$

Consequently, the product of these learning and regret quantities is $O(1)$ under the stated conditions.

Proof. The learning rate follows directly from Theorem 1. The assumed comparison between the corrected and greedy prices gives

$$(u_t - p_t^\star)^2 \le (v_t - p_t^\star)^2. \tag{24}$$

The price projection in equation (8) is inactive by assumption, so $\widetilde{p}_t = u_t$. Substitution into equation (12) gives

$$\bar{R}(T) \le (-b)\sum_{t=1}^{T} \mathbb{E}\left[(v_t - p_t^\star)^2 + \delta_t^2\right] = O(S_T). \tag{25}$$

More explicitly, the assumed greedy-price error bound gives $\bar{R}(T) \le (-b)(C_R + 1)S_T$. Combining this with equation (11) yields

$$\bar{R}(T)\mathbb{E}[(\widehat{b}_{T+1} - b)^2] \le C(-b)(C_R + 1),$$

which is the stated $O(1)$ product bound.

B Experimental Details

We describe the environments and paired runs used in Section 5, followed by Retail parameter selection, price-path diagnostics, and sensitivity analyses. Implementation checks precede the mechanism comparisons and longer-horizon experiments. For these checks, the reference implementation is a direct transcription of the authors' simulation code, recorded in `journal_loop.py` and `retail_data.py`; Appendix B.7 gives the comparison with its recorded outputs.

B.1 Data construction and simulation

Synthetic. We generate a population of 50,000 records once using seed 23, assigning each record to the second group with probability 0.48. Each record contains an intercept and three independent standard-normal context features, denoted here by z_1, z_2, z_3. We set

$$q = 8.4 + 0.85z_1 - 0.45z_2 + 0.35z_3 + 1.25\mathbf{1}\{g = B\}, \qquad b = -0.12, \qquad \mu = -2bq.$$

The simulated demand at price p is

$$D(p) = \max\{\mu + bp + \varepsilon, 0\}, \qquad \varepsilon \sim \mathcal{N}(0, 0.10^2).$$

Prices lie in $[4, 14]$. The reference price is $q^c = \mathrm{Proj}(q, [6, 12])$, whose population median initializes the market reference. The generator also draws a logged price by adding normal noise with standard deviation 0.55 to q^c and clipping to $[4, 14]$. This draw is unused by the pricing policies but is included when reproducing the random-number sequence.

Retail. The public UCI Online Retail dataset (Chen, 2015) contains 541,909 transactions. We remove records missing customer ID, stock code, country, or invoice date, retain positive quantities and prices, and restrict unit prices to the interval from 0.50 to the 95th percentile, 8.50. This leaves 341,167 transaction records: 304,754 from UK customers and 36,413 from non-UK customers. The unit of sampling is a transaction, rather than a distinct customer.

The context contains an intercept, log product-median price, log-one-plus product popularity, sine and cosine terms for month and day of week, and invoice hour divided by 24. Non-intercept features are standardized using the retained population's means and sample standard deviations. We fit unit price by ordinary least squares on these features and a non-UK indicator using a sample of 120,000 records with random state 11. If $q = x^\top \beta + \gamma \mathbf{1}\{g = \text{non-UK}\}$ denotes the fitted price, the simulation sets $b = -0.10$ and $\mu = -2bq$, with the same clipped-demand equation and noise distribution as above. This calibrates a demand environment from observed prices; it does not estimate causal price sensitivity from the transaction logs. Posted prices lie in $[0.50, 5.00]$.

The reference price is $q^c = \mathrm{Proj}(q, [1, 4])$; this clipping affects about 41.8% of retained records, and the median of q^c initializes the market reference. The untruncated quadratic revenue has unconstrained maximizer q; about 25.8% of these values fall outside the allowed interval $[0.5, 5]$. Projection keeps all experimental prices feasible. We compare policies using directly simulated revenue; Appendix B.7 also reports the revenue difference from q^c used in the implementation checks.

These simulations evaluate finite-horizon policy outcomes with Gaussian contexts or standardized empirical features, Gaussian noise, and nonnegative demand clipping. The main-text upper bounds use bounded contexts and noise with untruncated linear demand; Appendix D.5 extends the noise condition to conditional sub-Gaussian tails, including Gaussian noise. Clipping affects about 17–19% of Retail arrivals under C1 and C2 and changes both the quadratic revenue identity and the shock–demand moment used for slope estimation. The reported slope MSE measures finite-horizon error relative to the prescribed structural coefficient b, alongside directly simulated revenue.

B.2 Policies, configurations, and pairing

Both policies initialize $\widehat{b}_1 = -0.08$, the context-intercept estimate at $-2\widehat{b}_1\bar{p}_0$, and the other context and group coefficients at zero. The slope estimate is projected onto $[-0.25, -0.02]$. Context and group ridge penalties are both 1 in C1/C2 and both 100 in C3; group A in the Synthetic environment and UK in the Retail simulation serve as the reference groups without dummy variables. The price correction, interior projection, estimator, and queue timing follow Algorithm 1. RSD sets $\lambda_f = 0$, so its prices are unaffected by the queue recorded for diagnostics.

C1 uses the reference implementation's defaults. C2 is a second fixed parameter set, introduced as a diagnostic after examining C1; its values were not fitted to the evaluation runs. Within each configuration, matched RSD shares FRSD's exploration exponent and ridge penalties.

For each evaluation seed, NumPy's default random generator samples 2,000 population indices without replacement and then draws the demand noise. An independent generator initialized by `SeedSequence([seed, 915731])` draws equiprobable shock signs. All policies on the same

Table 4: FRSD configurations. RSD uses the same exploration exponent and ridge penalties and sets $\lambda_f = 0$. C1 and C2 are fixed configurations; C3 is selected on separate training runs for the main Retail comparison. Both ridge penalties are 1 in C1/C2 and 100 in C3.

Environment	Configuration	α	λ_f	τ
Synthetic	C1	0.20	1.25	0.05
	C2	0.10	0.25	0.02
Retail	C1	0.20	1.25	0.15
	C2	0.10	0.10	0.10
	C3	0.05	0.0003	0.001

seed receive the same arrivals, noise, and signs. Alternating-sign comparisons replace only the signs by $+1, -1, +1, -1, \ldots$; they keep the shock magnitudes and all other inputs fixed.

Each C1/C2 configuration and sign scheme uses all 100 seeds, 101–200; the tolerance grid uses all 20 seeds, 201–220, at each of its five values. C3 uses the separate training and test seeds specified below. Seeds index simulation replications from the same population, and reported standard deviations measure between-run variability. Paired intervals use the empirical standard deviation of within-seed differences and the Student-t quantile. Trajectory bands are pointwise intervals rather than simultaneous confidence bands.

B.3 Retail parameter selection and additional baselines

We select configuration C3 by tuning FRSD's parameters in the Retail simulation. The queue update, quadratic price objective, interior projection, projected slope estimator, and independent shocks follow Algorithm 1; population construction, calibration, price range, and demand noise follow Appendix B.1.

After exploratory searches over the fairness weight, tolerance, exploration exponent, and ridge penalties, a joint search evaluates 480 FRSD configurations on 200 training seeds, 14001–14200. It crosses $\alpha \in \{0, 0.05, 0.10, 0.15, 0.20\}$, a common context/group ridge penalty in $\{1, 10, 30, 100\}$, $\lambda_f \in \{0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03\}$, and $\tau \in \{0.001, 0.03, 0.08, 0.15\}$. Each combination of exploration exponent and ridge penalties has a matched RSD baseline. Two additional RSD settings from the exploratory searches give 22 RSD configurations in total. The best-revenue RSD on these training runs has $\alpha = 0.15$ and ridge penalties 0.1. None of the 480 FRSD settings has both higher mean revenue and lower mean disparity than this reference on the training runs.

Before testing, the matched-comparison selection requires at least a 5% reduction in mean disparity, at most a 5% revenue decrease, and at most a 10% increase in slope MSE. Among eligible configurations, it selects the largest disparity reduction, giving C3. Two stricter comparisons require these improvements against matched RSD, C1 RSD, and training-selected RSD simultaneously, with cost caps of either 5%/10% or 10%/25% for revenue/MSE; neither yields an eligible configuration. A further candidate minimizes the maximum normalized revenue and estimation costs while reducing disparity by at least 5% against all three references. Its parameters are $\alpha = 0.15$, ridge penalties 1, $\lambda_f = 0.03$, and $\tau = 0.001$.

We fix the candidates before evaluating them on 500 test seeds, 15001–15500. Table 5 reports all 22 RSD settings, both selected FRSD candidates, and an exploratory-search candidate with $\alpha = 0.10$, ridge penalties 0.01, $\lambda_f = 0.03$, and $\tau = 0.001$. Against matched RSD, FRSD under C3 reduces disparity by 19.0%, decreases revenue by 0.17%, and increases slope MSE by 4.3%.

The additional RSD settings show how exploration and regularization affect all three metrics. Relative to C1 RSD, FRSD under C3 has 9.1% higher disparity, 4.2% lower revenue, and 65.9% lower slope MSE. Relative to training-selected RSD, the corresponding changes are $+10.8\%$, -4.6%, and -48.7%. The matched comparison measures the effect of adding the correction at fixed exploration and estimation settings. Across settings, revenue, disparity, and slope MSE vary together, so the full comparison in Table 5 shows the trade-offs associated with choosing a pricing configuration.

The other FRSD candidates obtain fairness gains at larger costs relative to their respective matched baselines. The all-baseline joint-search candidate reduces disparity by 19.4%, with an 8.3% revenue decrease and a 65.6% MSE increase. The exploratory-search candidate reduces disparity by 24.8%, with a 7.8% revenue decrease and a 90.9% MSE increase. All test trajectories are saved, and the parameters stay fixed throughout testing.

Table 5: All settings evaluated on the 500 Retail test seeds. Ridge denotes both the context and group penalties. Entries are mean ± across-seed standard deviation; slope MSE is in units of 10^{-3}. C3 is the main matched-comparison setting. The other FRSD rows are the exploratory-search candidate and the all-baseline joint-search candidate.

Policy	α	Ridge	λ_f	τ	Revenue/ arrival	Slope MSE (10^{-3})	Disparity
RSD	0	1	0	0	0.885 ± 0.022	0.497 ± 0.427	0.086 ± 0.065
RSD	0	10	0	0	0.884 ± 0.022	0.485 ± 0.422	0.088 ± 0.065
RSD	0	30	0	0	0.881 ± 0.022	0.465 ± 0.408	0.095 ± 0.070
RSD	0	100	0	0	0.871 ± 0.024	0.429 ± 0.388	0.113 ± 0.079
RSD	0.05	1	0	0	0.929 ± 0.025	0.529 ± 0.564	0.083 ± 0.064
RSD	0.05	10	0	0	0.927 ± 0.025	0.520 ± 0.558	0.086 ± 0.065
RSD	0.05	30	0	0	0.924 ± 0.024	0.502 ± 0.541	0.095 ± 0.070
RSD†	0.05	100	0	0	0.912 ± 0.027	0.473 ± 0.519	0.116 ± 0.080
RSD	0.1	1	0	0	0.948 ± 0.029	0.668 ± 0.812	0.083 ± 0.064
RSD	0.1	10	0	0	0.946 ± 0.028	0.665 ± 0.810	0.086 ± 0.065
RSD	0.1	30	0	0	0.943 ± 0.028	0.652 ± 0.792	0.096 ± 0.071
RSD	0.1	100	0	0	0.928 ± 0.032	0.635 ± 0.771	0.117 ± 0.081
RSD	0.15	1	0	0	0.954 ± 0.035	0.964 ± 1.227	0.084 ± 0.065
RSD	0.15	10	0	0	0.951 ± 0.034	0.974 ± 1.236	0.087 ± 0.065
RSD	0.15	30	0	0	0.946 ± 0.034	0.976 ± 1.230	0.096 ± 0.071
RSD	0.15	100	0	0	0.929 ± 0.040	0.986 ± 1.236	0.117 ± 0.082
RSD	0.2	1	0	0	0.951 ± 0.041	1.450 ± 1.754	0.086 ± 0.065
RSD	0.2	10	0	0	0.947 ± 0.042	1.486 ± 1.796	0.089 ± 0.066
RSD	0.2	30	0	0	0.940 ± 0.042	1.524 ± 1.835	0.097 ± 0.072
RSD	0.2	100	0	0	0.920 ± 0.050	1.591 ± 1.906	0.118 ± 0.082
RSD	0.15	0.1	0	0	0.954 ± 0.035	0.963 ± 1.226	0.085 ± 0.065
RSD	0.1	0.01	0	0	0.949 ± 0.029	0.669 ± 0.815	0.083 ± 0.064
FRSD	0.1	0.01	0.03	0.001	0.875 ± 0.040	1.278 ± 1.268	0.063 ± 0.043
FRSD (C3)	0.05	100	0.0003	0.001	0.911 ± 0.028	0.494 ± 0.543	0.094 ± 0.065
FRSD	0.15	1	0.03	0.001	0.875 ± 0.045	1.596 ± 1.666	0.068 ± 0.042

†RSD baseline matched to FRSD (C3).

Percentage changes are ratios of means, rather than averages of per-run ratios. Their confidence intervals use a first-order linearization retaining the paired baseline observations and a Student-t critical value. These intervals describe individual comparisons and are not adjusted for multiple testing.

The test suite contains 12,500 policy trajectories. We check the batched implementation against the per-arrival reference implementation over full 2,000-arrival runs, reconstruct all terminal metrics from saved trajectories, and verify price bounds and shock magnitudes. Data and reference-code checksums confirm that the same population and reference implementation are used throughout.

B.4 Retail price differences and queue activity

To interpret the absolute disparities above, we look at the direction of the price gap and how it varies across runs. Table 6 reports the terminal signed difference, UK minus non-UK, alongside center contact and queue values from the saved trajectories. The signed difference shows which group receives higher average prices; fairness itself is measured by the absolute disparity in equation (2).

Table 6: Diagnostics from independent-shock trajectories: 100 paired runs for C1/C2 and 500 for C3. Signed differences are mean ± across-run standard deviation. Center contact is the mean fraction of arrivals at an endpoint of the interior price interval, a diagnostic relevant to the inactive-projection conditions in Proposition 1 and Corollary 1; endpoint contact alone does not establish whether projection changed the price. The last two columns report FRSD's mean queues after T observations, $Q_{g,T+1}$. FRSD uses tolerances 0.15, 0.10, and 0.001 under C1–C3. RSD's diagnostic queues use a different tolerance, $\tau = 0$, and are not shown.

Config.	Policy	Signed difference	Center contact	UK queue	Non-UK queue
C1	RSD	-0.039 ± 0.093	40.3%	–	–
C1	FRSD	-0.011 ± 0.176	37.0%	0.00	81.65
C2	RSD	-0.040 ± 0.092	41.8%	–	–
C2	FRSD	-0.015 ± 0.164	37.8%	0.00	140.79
C3	RSD	-0.098 ± 0.101	39.9%	–	–
C3	FRSD	-0.048 ± 0.104	36.3%	28.96	257.99

For C1 and C2, FRSD brings the mean signed difference closer to zero, but the differences vary more across runs. Mean absolute disparity is 78.1% and 67.7% higher than RSD's, respectively. Under C3, the mean signed difference is -0.048 for FRSD versus -0.098 for RSD, and the corresponding standard deviations are 0.104 and 0.101. The 19.0% reduction in mean absolute disparity thus comes with a shift toward zero and across-run dispersion close to matched RSD.

To see whether the improvement extends along the price path, we average the absolute difference between the two groups' cumulative mean prices over all 2,000 arrivals in each C3 trajectory. Until both groups have appeared, we set the unobserved group's mean to the market mean, as in the pricing rule. On the same 500 paired test runs, this time-averaged disparity is 0.1694 for matched RSD and 0.1455 for FRSD, a 14.1% reduction.

Group proportions affect how price differences enter the queues. UK transactions make up about 89.3% of the Retail population, so at the same group-price disparity their group mean sits closer to the market mean than the non-UK mean does. In C1 and C2, the UK queue is positive at only 0.4% and 1.3% of UK arrivals, respectively, and ends at zero on every run. In C3, both groups have positive queues before pricing at about 99.5% of their arrivals. These diagnostics show how the same queue rule responds under different parameter settings; the resulting prices reflect the joint changes in exploration, estimation, and fairness correction.

B.5 Independent and alternating shocks

Changing only the shock signs isolates the role of the exploration scheme. Table 7 shows that, in the Synthetic environment, FRSD has lower slope MSE than RSD under alternating signs and higher slope MSE under independent signs. The independent scheme follows equation (8); deterministic alternation falls outside its conditional randomization assumption. In the 100-run Retail C1/C2 comparisons, neither scheme yields a mean fairness gain over matched RSD.

B.6 Full tolerance sweep

The preceding comparison changes the perturbation signs; this sweep varies only the tolerance. We fix $\alpha = 0.20$ and $\lambda_f = 1.25$, share arrivals, noise, and independent signs within each seed, and evaluate RSD once as the common reference. The grid is used for sensitivity analysis, not to select the C1/C2 configurations in Table 4.

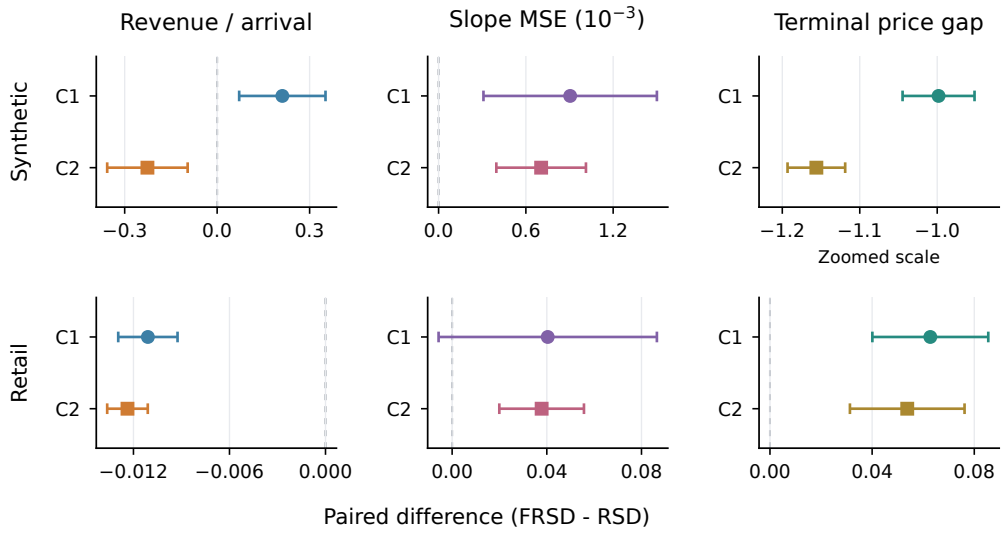

Figure 2: C1/C2 comparisons: mean paired differences, FRSD minus matched RSD, under independent shocks. Circles denote C1 and squares C2; error bars are 95% Student-t confidence intervals across 100 paired runs. Positive revenue differences favor FRSD; negative slope-MSE and disparity differences favor FRSD. The Synthetic disparity panel uses a zoomed horizontal scale that excludes zero. Slope MSE differences are displayed in units of 10^{-3}. Figure 1 reports the separate Retail C3 comparison. Intervals are descriptive and not adjusted for multiple comparisons.

Table 7: Full sign-scheme comparison on seeds 101–200, with $T = 2000$. Entries are mean $\pm$ across-seed standard deviation. Revenue is higher-is-better; the other metrics are lower-is-better.

Environment	Config.	Signs	Policy	Revenue/arrival	MSE (10^{-3})	Disparity
Synthetic	C1	Alternating	RSD	8.846 ± 0.304	2.415 ± 0.952	1.442 ± 0.095
Synthetic	C1	Alternating	FRSD	9.374 ± 0.278	0.629 ± 0.813	0.098 ± 0.020
Synthetic	C1	Independent	RSD	8.846 ± 0.600	4.980 ± 5.164	1.088 ± 0.231
Synthetic	C1	Independent	FRSD	9.057 ± 0.541	5.885 ± 5.540	0.089 ± 0.030
Synthetic	C2	Alternating	RSD	9.506 ± 0.121	0.475 ± 0.272	1.419 ± 0.074
Synthetic	C2	Alternating	FRSD	9.360 ± 0.289	0.180 ± 0.238	0.042 ± 0.022
Synthetic	C2	Independent	RSD	9.254 ± 0.413	2.125 ± 3.113	1.196 ± 0.188
Synthetic	C2	Independent	FRSD	9.028 ± 0.601	2.830 ± 3.867	0.040 ± 0.021
Retail	C1	Alternating	RSD	0.957 ± 0.036	0.997 ± 1.098	0.089 ± 0.062
Retail	C1	Alternating	FRSD	0.946 ± 0.034	1.076 ± 1.211	0.120 ± 0.074
Retail	C1	Independent	RSD	0.949 ± 0.036	1.299 ± 1.461	0.080 ± 0.061
Retail	C1	Independent	FRSD	0.938 ± 0.034	1.340 ± 1.508	0.143 ± 0.101
Retail	C2	Alternating	RSD	0.951 ± 0.025	0.489 ± 0.447	0.087 ± 0.063
Retail	C2	Alternating	FRSD	0.940 ± 0.024	0.544 ± 0.483	0.106 ± 0.067
Retail	C2	Independent	RSD	0.947 ± 0.025	0.616 ± 0.816	0.079 ± 0.061
Retail	C2	Independent	FRSD	0.935 ± 0.024	0.653 ± 0.847	0.133 ± 0.096

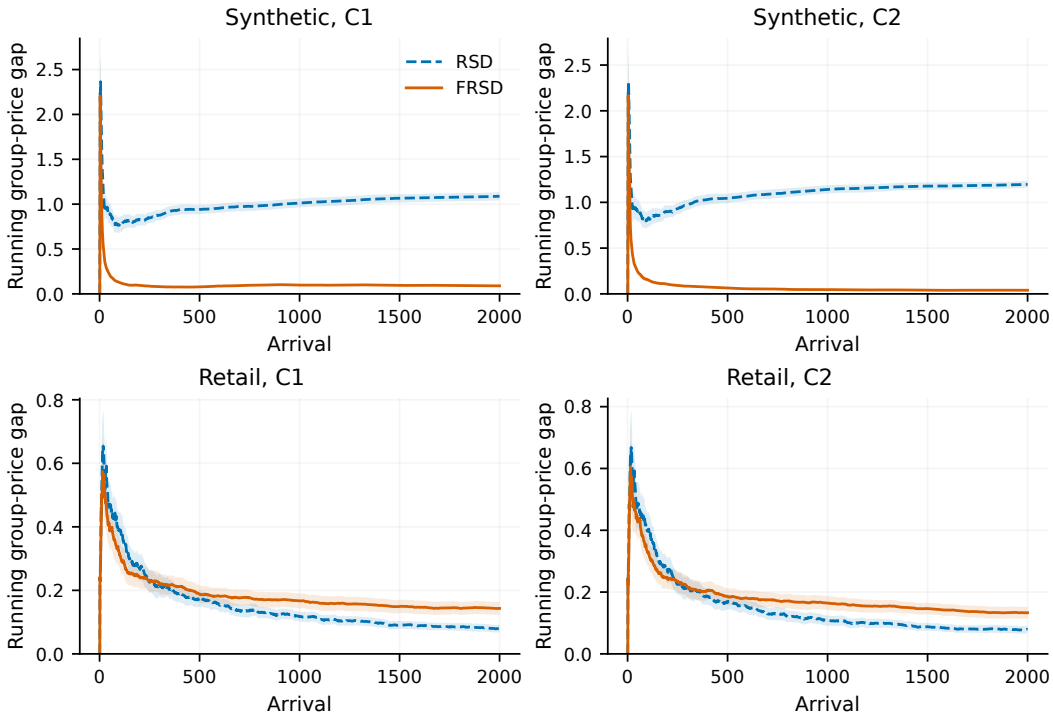

Figure 3: Disparity between cumulative group-average prices under independent shocks for C1 and C2. Lines show means across all 100 runs; shaded regions are pointwise 95% Student-t intervals. The Retail panels use C1/C2; C3 is evaluated in Figure 1. Disparity is zero before both groups appear because the unobserved group's mean is initialized at the market mean.

Table 8: All tolerance-grid results, mean $\pm$ across-seed standard deviation over seeds 201–220. C1's other parameters are held fixed.

Environment	Policy	τ	Revenue/arrival	MSE (10^{-3})	Disparity
Synthetic	RSD	–	9.041 ± 0.521	2.238 ± 3.744	1.202 ± 0.176
Synthetic	FRSD	0.03	9.078 ± 0.579	3.116 ± 4.059	0.056 ± 0.021
Synthetic	FRSD	0.05	9.114 ± 0.549	3.013 ± 4.062	0.100 ± 0.025
Synthetic	FRSD	0.10	9.147 ± 0.580	3.185 ± 4.210	0.192 ± 0.047
Synthetic	FRSD	0.15	9.274 ± 0.510	3.187 ± 4.267	0.284 ± 0.073
Synthetic	FRSD	0.20	9.276 ± 0.485	3.037 ± 4.082	0.393 ± 0.057
Retail	RSD	–	0.952 ± 0.029	0.749 ± 1.132	0.086 ± 0.083
Retail	FRSD	0.03	0.897 ± 0.058	0.941 ± 1.386	0.157 ± 0.108
Retail	FRSD	0.05	0.927 ± 0.031	0.949 ± 1.359	0.191 ± 0.148
Retail	FRSD	0.10	0.937 ± 0.026	0.841 ± 1.302	0.129 ± 0.073
Retail	FRSD	0.15	0.940 ± 0.027	0.831 ± 1.251	0.120 ± 0.090
Retail	FRSD	0.20	0.942 ± 0.028	0.841 ± 1.303	0.130 ± 0.102

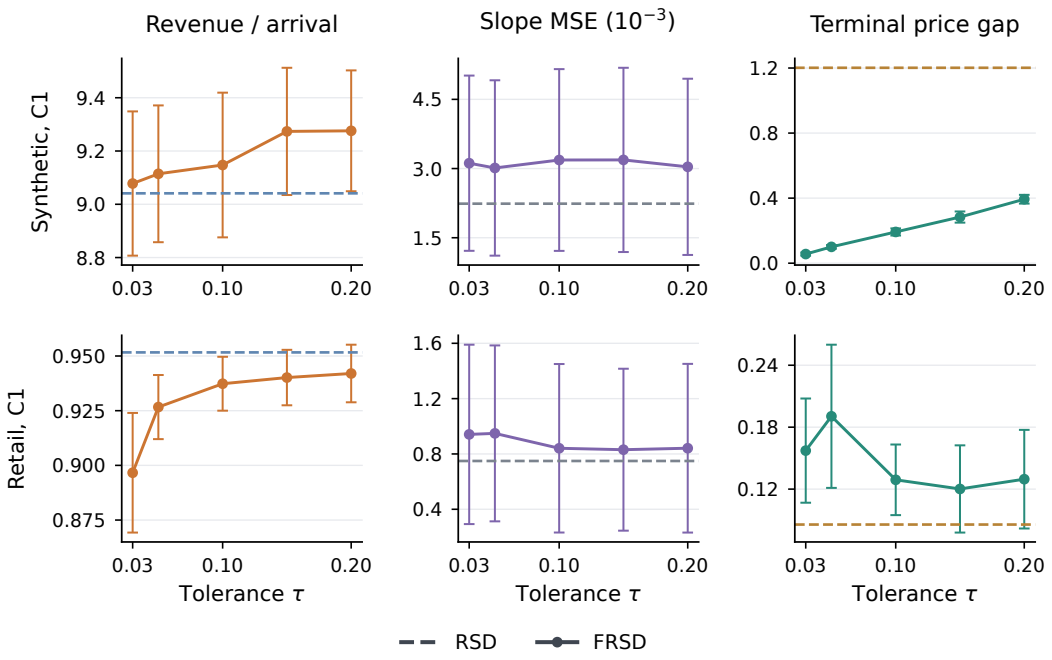

Figure 4: Tolerance sweep on 20 paired runs. FRSD points are means with 95% Student-t intervals; dashed lines are the matched RSD means. In the Synthetic environment, tighter tolerances yield smaller mean disparities and lower mean revenues across the tested values. In the Retail simulation, disparity is not monotone in tolerance and exceeds RSD's throughout this grid.

B.7 Implementation agreement and numerical checks

We verify the reference implementation against the authors' recorded C1 results with alternating signs on seeds 1–20. All 16 rounded means for revenue, reference-revenue loss, slope MSE, and disparity agree across both environments and policies. Table 9 reports three of these metrics, with reference-revenue loss defined as

$$\frac{1}{T}\sum_{t=1}^{T}\{q_t^c D_t(q_t^c) - p_t D_t(p_t)\},$$

where q_t^c is the environment-specific reference price defined above and both demand evaluations use the same noise. This diagnostic compares revenue with q_t^c; theoretical regret instead uses the feasible oracle price $p_t^\star$.

Table 9: Reference-implementation results on 20 alternating-sign runs using C1. Reference loss is measured relative to the reference price q_t^c.

Environment	Policy	Reference loss/arrival	MSE (10^{-3})	Disparity
Synthetic	RSD	0.894 ± 0.299	1.857 ± 1.029	1.465 ± 0.072
Synthetic	FRSD	0.459 ± 0.198	0.624 ± 0.604	0.103 ± 0.016
Retail	RSD	0.015 ± 0.034	1.307 ± 1.597	0.087 ± 0.063
Retail	FRSD	0.025 ± 0.032	1.371 ± 1.672	0.081 ± 0.056

The C1/C2, tolerance, and implementation-check suite contains 1,920 saved policy trajectories of 2,000 arrivals: 80 implementation-check runs, 1,600 configuration/sign comparisons, and 240 tolerance-grid runs. Automated checks verify finite saved values, price feasibility, shock magnitudes, terminal squared slope error and disparity, and revenue reconstructed from saved prices and demands.

The queue certificate uses $Q_{g,T}$, the queue after $T-1$ observations. Its two-group consequence,

$$\frac{1}{T-1}\sum_{t=2}^{T}|\bar{p}_{A,t-1}-\bar{p}_{B,t-1}| \le 2\tau + \frac{Q_{A,T}+Q_{B,T}}{T-1},$$

follows by adding the group-to-market inequalities and bounds the average pairwise disparity through arrival $T-1$.

B.8 Mechanism comparisons and a bounded linear experiment

Three ways to weight the same penalty. We compare FRSD's queue correction with two simpler alternatives, holding the revenue objective, price projection, ridge fit, slope estimator, initialization, and independent shocks fixed. At arrival t, write $h_{g,t-1} = |\bar{p}_{g,t-1} - \bar{p}_{t-1}|$. The coefficient multiplying $(p-\bar{p}_{t-1})^2$ in equation (7) is respectively

$$k, \qquad k[h_{g_t,t-1}-\tau]^+, \qquad kQ_{g_t,t}$$

for the fixed penalty, current-gap penalty, and FRSD. The current-gap penalty uses only information available before the offer; its input is zero for a group that has not appeared. RSD uses coefficient zero. These are mechanism comparisons within the same pricing architecture, rather than implementations of separately published algorithms.

Training and held-out evaluation. Each correction family uses the same 17 candidates,

$k \in \{10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}, 0.003, 0.01, 0.03, 0.1, 0.25, 0.3, 1, 1.25, 3, 10, 30, 100\}$.

The range spans several orders of magnitude because current and accumulated gaps have different scales. Exploration, both ridge penalties, and tolerance follow the Synthetic C1, Synthetic C2, and Retail C3 settings in Table 4. Correction strength is selected separately in each family for this mechanism comparison; Table 1 reports FRSD at the fixed strengths listed in Table 4.

We fix the grids, targets, seeds, and reporting rules before running these comparisons. All 52 policies, including RSD, are evaluated on 80 training seeds, 31001–31080. Within each family, the primary selection minimizes mean terminal disparity among candidates with mean revenue at least 99% of matched RSD and slope MSE at most 110% of matched RSD. A pre-specified secondary selection changes only the revenue threshold to 95%. RSD is an available zero-correction fallback in every family. Ties favor higher revenue and then smaller strength. Selected settings remain fixed throughout evaluation of all candidates on 200 separate test seeds, 32001–32200. Pairing and the horizon $T = 2000$ follow Section 5.

The primary selected strengths for the fixed penalty, current-gap penalty, and FRSD are $0.3/1.25/0.03$ for C1, $0.3/3/0.001$ for C2, $0.01/0.3/0.0003$ for Retail C3, and $0.3/100/1.25$ for the bounded linear environment. Under the secondary target, only the C2 FRSD selection changes, to $k = 0.03$. Table 10 reports all primary selections. The secondary C2 FRSD has mean revenue 9.1309, slope MSE 1.855×10^{-3}, and disparity 0.0802; its matched RSD has 9.2827, 1.781×10^{-3}, and 1.2052, respectively.

The targets concern training means. On test runs, primary C2 FRSD has a 1.15% revenue decrease relative to RSD. In the bounded linear-demand environment, the current-gap penalty and FRSD have 10.3% and 13.3% higher slope MSE than RSD. Figure 6 shows every tested strength, including the selected settings. The complete suite contains 16,640 training and 41,600 test policy trajectories. Metrics and inputs are saved for every run, with full selected-policy paths on the first three test seeds and code to reproduce the remaining paths.

Paired comparisons. All differences and ratios use paired observations. For C1, the disparity difference between FRSD and the fixed penalty is -0.2568, with a 95% confidence interval $[-0.2648, -0.2487]$; the difference from the current-gap penalty is -0.3307, with interval $[-0.3374, -0.3241]$. The corresponding revenue difference from the fixed penalty is 0.0165, with interval $[-0.0578, 0.0908]$, which includes zero despite the higher sample mean. In the Retail simulation, the disparity difference from the fixed penalty is -0.0176, with interval $[-0.0265, -0.0086]$; the difference from the current-gap penalty is -0.0036, with interval $[-0.0077, 0.0005]$. These are pointwise paired Student-t intervals, not simultaneous intervals across the grid. Percentage intervals use the paired ratio linearization in Appendix B.3.

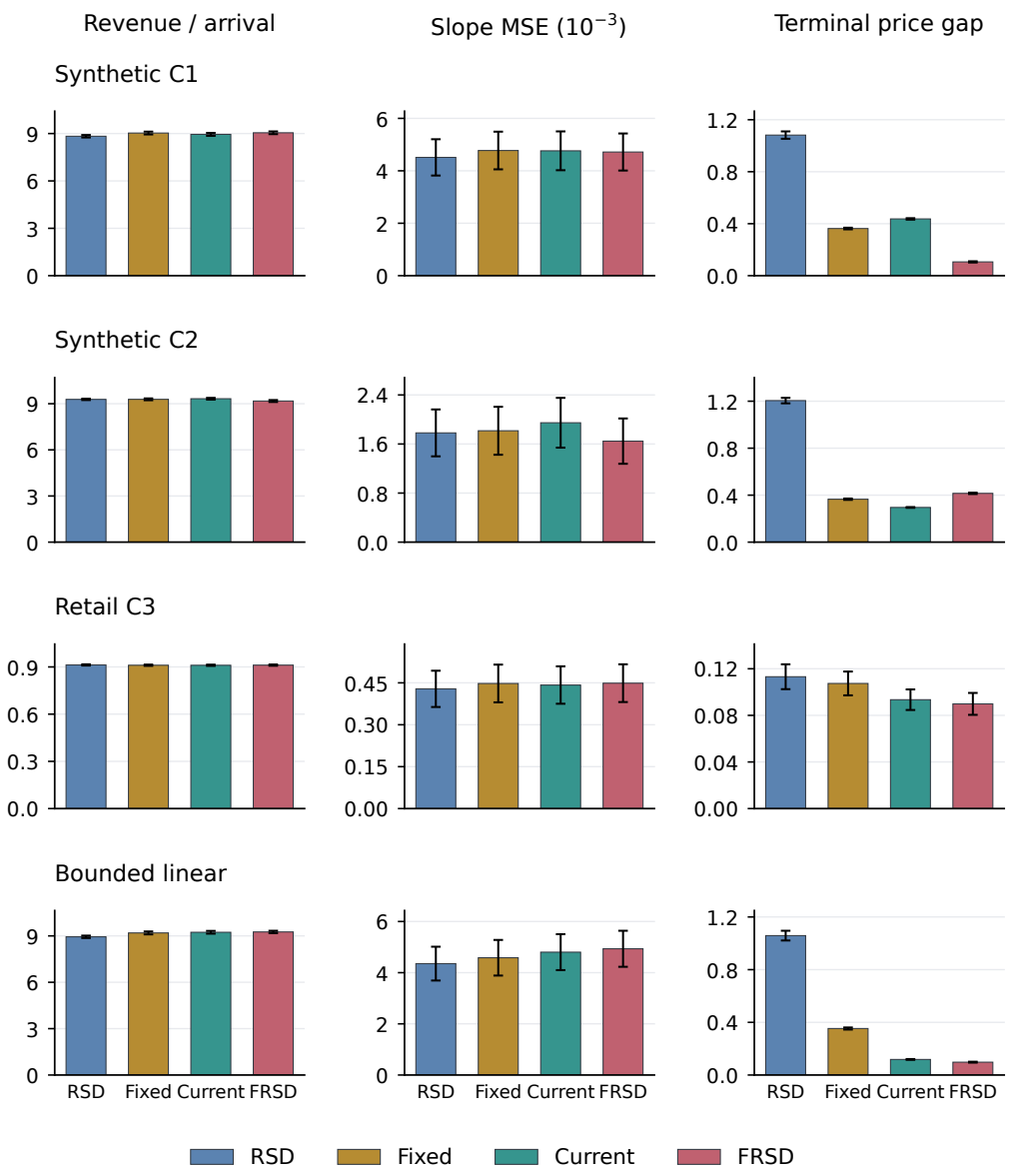

Figure 5: Primary mechanism selections in all four environments. Bars are means over 200 paired test runs; whiskers are 95% Student-t intervals for individual means. FRSD yields substantially smaller disparities under C1; C2, Retail, and bounded linear demand show the accompanying differences in revenue and estimation. All bar axes begin at zero.

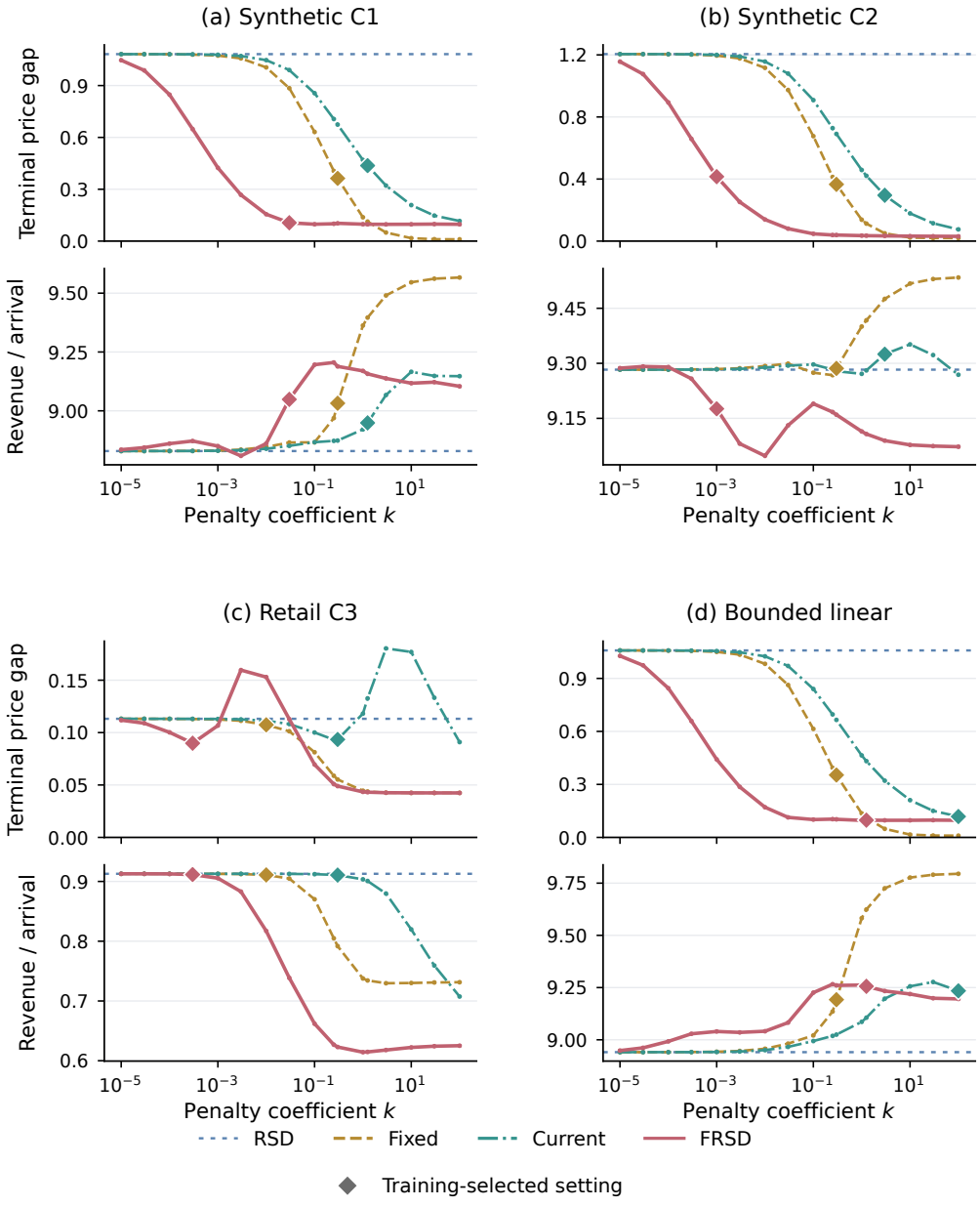

Figure 6: Disparity and revenue across all 17 penalty coefficients per correction method, averaged over 200 test runs. Each environment has a disparity panel above its revenue panel, with coefficient k on a logarithmic axis. Diamonds mark the training-selected primary settings; horizontal dotted lines give the RSD means. The paired panels show how increasing the coefficient changes both outcomes within each method. Slope MSE also enters selection and is reported for the selected policies in Table 10.

Table 10: Primary mechanism selections on 200 paired test runs. Entries are mean ± across-run standard deviation; MSE is in units of 10^{-3}. Fixed and Current denote the fixed and current-gap penalties; both share FRSD's estimation and exploration settings. Correction strength k is selected on the separate training runs.

Environment	Policy	k	Revenue/arrival	MSE	Disparity
Synthetic C1	RSD	0	8.8291 ± 0.5696	4.512 ± 4.965	1.0818 ± 0.2036
Synthetic C1	Fixed	0.3	9.0322 ± 0.6386	4.774 ± 5.152	0.3629 ± 0.0456
Synthetic C1	Current	1.25	8.9487 ± 0.6713	4.762 ± 5.309	0.4369 ± 0.0406
Synthetic C1	FRSD	0.03	9.0486 ± 0.6237	4.717 ± 5.069	0.1061 ± 0.0345
Synthetic C2	RSD	0	9.2827 ± 0.3218	1.781 ± 2.742	1.2052 ± 0.1673
Synthetic C2	Fixed	0.3	9.2862 ± 0.4308	1.816 ± 2.805	0.3656 ± 0.0414
Synthetic C2	Current	3	9.3249 ± 0.4044	1.947 ± 2.912	0.2954 ± 0.0240
Synthetic C2	FRSD	0.001	9.1761 ± 0.4603	1.647 ± 2.650	0.4153 ± 0.0426
Retail C3	RSD	0	0.9128 ± 0.0285	0.428 ± 0.466	0.1132 ± 0.0765
Retail C3	Fixed	0.01	0.9109 ± 0.0294	0.448 ± 0.484	0.1074 ± 0.0733
Retail C3	Current	0.3	0.9106 ± 0.0290	0.442 ± 0.480	0.0934 ± 0.0633
Retail C3	FRSD	0.0003	0.9115 ± 0.0291	0.449 ± 0.484	0.0898 ± 0.0675
Bounded linear	RSD	0	8.9407 ± 0.5590	4.354 ± 4.738	1.0584 ± 0.2652
Bounded linear	Fixed	0.3	9.1921 ± 0.6804	4.583 ± 4.992	0.3531 ± 0.0569
Bounded linear	Current	100	9.2344 ± 0.6423	4.802 ± 5.035	0.1182 ± 0.0230
Bounded linear	FRSD	1.25	9.2560 ± 0.5516	4.934 ± 5.053	0.0975 ± 0.0254

Bounded linear-demand environment. We generate a separate population of 50,000 observations using seed 602026. Three independent features x_1, x_2, x_3 are uniform on $[0, 1]$, and group B membership is Bernoulli with probability 0.48. With an intercept included in the regression, set

$$p^\star = 8.5 + (x_1 - 0.5) - 0.6(x_2 - 0.5) + 0.4(x_3 - 0.5) + 1.25\mathbf{1}\{g = B\}, \qquad b = -0.12.$$

Demand is $D(p) = -2bp^\star + bp + \varepsilon$, with independent noise uniform on $[-\sqrt{3}\,0.1, \sqrt{3}\,0.1]$. No truncation is applied, so rare negative demand realizations are retained. Prices lie in $[4, 14]$, and the oracle lies in $[7.5, 10.75]$. Thus contexts and noise are bounded and the oracle is feasible, as required for the upper-bound model. Exploration, ridge penalties, and tolerance match C1. The initial market reference is the population median oracle price.

Learning, adjustment, and benchmark diagnostics. For every selected trajectory in the bounded linear-demand environment, we reconstruct the greedy prices from the preceding demand observations and ridge fit. The sample sums corresponding to E_T and A_T, the cross term $2\sum_t (v_t - p_t^\star)(\widetilde{p}_t - v_t)$, and the shock term $2\sum_t \Delta_t(\widetilde{p}_t - p_t^\star)$ recover the noise-free oracle regret. The largest identity residual is 1.2×10^{-8}. This checks the pathwise identities numerically; the expectations in Theorems 1 and 2 average over random paths.

Table 11: Bounded linear-demand diagnostics averaged over 200 test runs. All losses and squared-error sums are divided by $T = 2000$. "Fair cost" denotes the price of fairness $\mathrm{PoF}_T(d)/T$ at each path's attained disparity d. The cross term records whether the adjustment offsets or amplifies greedy-price error.

Policy	E/T	A/T	Cross$/T$	Oracle loss	Fair cost	Excess
RSD	31.432	11.856	-34.505	1.0632	0.0031	1.0601
Fixed	60.502	42.315	-96.132	0.8115	0.0243	0.7872
Current	92.219	85.501	-171.393	0.7688	0.0385	0.7303
FRSD	114.852	120.508	-229.215	0.7468	0.0399	0.7069

The negative cross terms show that the implemented adjustments offset part of the greedy-price error; A_T alone is therefore not a realized revenue loss. Under FRSD, the mean total shock cross term is -0.867, with a 95% interval $[-8.602, 6.867]$, consistent with its zero conditional mean. The minimum observed eigenvalue ratio $\lambda_{\min}(V_n)/n$ over $50 \le n \le 2000$ is also recorded for every run;

its across-run mean is 0.0442. This finite-horizon diagnostic describes the fitted design, rather than establishing a uniform asymptotic design condition.

For the benchmark calculation, let $d = \Delta(\mathbf{p})$ be the realized disparity and let ν_A, ν_B be the group means of the oracle prices. The group shifts in Appendix F remain inside the price bounds on every tested path. Hence the exact optimal cost at that same disparity is

$$\mathrm{PoF}_T(d) = (-b)\frac{N_A N_B}{T}(|\nu_A - \nu_B| - d)_+^2.$$

Subtracting it from oracle regret separates the cost of the attained fairness level from the remaining policy loss (Table 11). The latter is not attributed solely to statistical learning: it includes how the policy allocates prices within and between groups. All tested paths satisfy the fair-projection inequality numerically.

Implementation checks. A separately written per-arrival loop reproduces the batched RSD, fixed, current-gap, and FRSD trajectories in the Synthetic and Retail environments, including slope clipping, ridge updates, queue timing, and price projection. Full-horizon FRSD runs also agree with the reference implementation. Data and reference-code hashes are checked, and selection decisions are reconstructed from the saved training arrays. Every scheduled run is included.

B.9 Fixed-parameter horizon check

To examine the zero-tolerance result in Theorem 3, we use the bounded linear-demand population with $\alpha = 0.2$, $\lambda_f = 1.25$, and ridge penalties 1. We compare $\tau = 0.05$ with $\tau = 0$ and matched RSD, fixing all parameters before evaluation. The protocol specifies 100 separate seeds, 33001–33100, and prefixes of 500, 2,000, 8,000, and 32,000 arrivals. Each seed samples 32,000 arrivals without replacement from the 50,000-observation population and shares arrivals, noise, and independent shock signs across the three policies.

Table 12: Fixed-parameter horizon check over 100 paired runs. All entries are means; disparity intervals are individual 95% Student-t intervals. MSE is in units of 10^{-3}. The queue column is $\max_g Q_{g,T+1}/T$ for FRSD and is not reported for RSD. All four prefixes and both FRSD tolerances are shown.

T	Policy	Disparity	95% interval	Revenue	MSE	Queue$/T$
500	RSD	0.8297	[0.7563, 0.9031]	8.6180	7.795	–
500	FRSD, $\tau = 0.05$	0.0952	[0.0854, 0.1049]	9.1722	7.883	0.0246
500	FRSD, $\tau = 0$	0.0298	[0.0256, 0.0341]	9.1473	7.886	0.0555
2,000	RSD	1.0316	[0.9783, 1.0849]	8.9757	4.612	–
2,000	FRSD, $\tau = 0.05$	0.0985	[0.0940, 0.1029]	9.2221	5.041	0.0064
2,000	FRSD, $\tau = 0$	0.0125	[0.0107, 0.0143]	9.1673	5.049	0.0208
8,000	RSD	1.1688	[1.1296, 1.2081]	9.3608	2.482	–
8,000	FRSD, $\tau = 0.05$	0.0986	[0.0963, 0.1010]	9.2896	2.681	0.0029
8,000	FRSD, $\tau = 0$	0.0052	[0.0044, 0.0059]	9.1775	2.725	0.0081
32,000	RSD	1.2476	[1.2133, 1.2820]	9.6008	1.336	–
32,000	FRSD, $\tau = 0.05$	0.1017	[0.1009, 0.1025]	9.3960	1.448	0.0025
32,000	FRSD, $\tau = 0$	0.0024	[0.0021, 0.0027]	9.1806	1.570	0.0033

At zero tolerance, mean terminal disparity decreases from 0.0298 at 500 arrivals to 0.0024 at 32,000, while the normalized maximum queue decreases from 0.0555 to 0.0033. With $\tau = 0.05$, longer-horizon disparity remains near 0.10. The zero-tolerance runs illustrate the approach toward equal group-average prices predicted by Theorem 3, while the revenue and estimation columns show the accompanying finite-horizon trade-offs.

The saved outputs contain all 1,200 policy-prefix records and all 100 shared input sequences, together with the first three full sets of trajectories and the deterministic runner needed to reproduce the rest. Replay of a bounded linear-demand FRSD/RSD run verifies agreement with the mechanism-comparison engine before evaluation. Checks cover price feasibility, shock magnitudes, terminal disparity reconstructed from group prices, and engine hashes.

A design satisfying the theorem's conditions. We also run a structured synthetic check with the same demand coefficients, noise, price bounds, and policy parameters. Its arrivals repeat a 16-row block: each of the eight binary triples in $\{0,1\}^3$ appears once in group A and once in group B, in alternating group order. If Z is the block's regression matrix, including the intercept and group indicator, then $Z^\top Z \succ 0$ and, for every $n \geq 32$,

$$\lambda_{\min}(V_n) \geq \left\lfloor \frac{n}{16} \right\rfloor \lambda_{\min}(Z^\top Z) \geq \frac{n}{32}\lambda_{\min}(Z^\top Z).$$

Both groups have at least $n/3$ arrivals for every $n \geq 2$, so the pathwise design and coverage conditions hold by construction. We fix this design before running 100 separate noise and shock seeds, 34001–34100. Table 13 reports all three policies, four prefixes, and outcome metrics, complementing the randomly sampled population experiment with a direct check under the theorem's conditions.

Table 13: The structured-design check with analytically verified design and coverage conditions. Means and individual 95% disparity intervals over 100 paired runs; notation and MSE units follow Table 12. Parameters are fixed before evaluation.

T	Policy	Disparity	95% interval	Revenue	MSE	Queue/T
500	RSD	0.8430	[0.7874, 0.8987]	8.7236	6.916	–
500	FRSD, $\tau = 0.05$	0.0840	[0.0746, 0.0935]	9.2813	7.319	0.0120
500	FRSD, $\tau = 0$	0.0309	[0.0267, 0.0351]	9.2847	7.400	0.0483
2,000	RSD	1.0276	[0.9804, 1.0748]	9.0897	4.870	–
2,000	FRSD, $\tau = 0.05$	0.0959	[0.0917, 0.1001]	9.2778	5.745	0.0017
2,000	FRSD, $\tau = 0$	0.0121	[0.0104, 0.0139]	9.2992	5.841	0.0190
8,000	RSD	1.1434	[1.1022, 1.1846]	9.3560	3.088	–
8,000	FRSD, $\tau = 0.05$	0.0987	[0.0970, 0.1004]	9.3047	3.702	0.0004
8,000	FRSD, $\tau = 0$	0.0050	[0.0043, 0.0057]	9.3073	3.919	0.0074
32,000	RSD	1.2330	[1.1998, 1.2662]	9.6931	1.325	–
32,000	FRSD, $\tau = 0.05$	0.0996	[0.0981, 0.1010]	9.4096	1.570	0.0001
32,000	FRSD, $\tau = 0$	0.0024	[0.0021, 0.0027]	9.3120	1.742	0.0031

C Price Correction and Revenue

We begin with the price correction in Section 4, including the case where price constraints are active. Its explicit form shows how the queue weight changes the chosen price and when that change reduces oracle error. We then use a cumulative error bound to extend the regret-rate calculation to projected prices. The later sections turn to estimation, price histories, and the fair benchmark.

C.1 The constrained quadratic problem

At one arrival, suppress the time index and write

$$c = -\widehat{b} > 0, \qquad q = \lambda_f Q \geq 0, \qquad m = \bar{p}_{t-1}, \qquad v = \widehat{\mu}/(2c).$$

Let $I = [\underline{p}, \overline{p}]$ and $J = [\underline{p} + \delta, \overline{p} - \delta]$ be nonempty. The queue and estimates are held fixed during this pricing decision. Completing the square in its objective gives

$$f(p) = -cp^2 + 2cvp - q(p-m)^2 \tag{26}$$

$$= cv^2 - \frac{cq}{c+q}(v-m)^2 - (c+q)(p-z)^2, \qquad z = \frac{cv+qm}{c+q}. \tag{27}$$

Thus the constrained optimization is a scalar projection of z. This representation describes the implemented decision whether or not a price bound is active.

Proposition 2 (Correction with price constraints)**.** *For the objective in equation* (7)*, the unique maximizer and the pre-shock price satisfy*

$$u = \mathrm{Proj}(z, I), \qquad \widetilde{p} = \mathrm{Proj}(z, J). \tag{28}$$

Both posted prices $\widetilde{p} - \delta$ and $\widetilde{p} + \delta$ belong to I. The formulas include the case where J consists of a single point.

Proof. Since $c + q > 0$, maximizing equation (27) over I is equivalent to minimizing $(p - z)^2$ there. The minimizer is unique and is the clipped value of z. For nested intervals $J \subseteq I$, scalar clipping satisfies

$$\mathrm{Proj}(\mathrm{Proj}(z, I), J) = \mathrm{Proj}(z, J).$$

To verify this, if z lies below I, both sides equal the lower endpoint of J; if it lies above I, both equal the upper endpoint of J; if $z \in I$, the inner projection leaves it unchanged. These cases also cover equality at the endpoints. The implementation projects u onto J, giving the second formula. Finally, $\widetilde{p} \in J$ implies $\underline{p} \leq \widetilde{p} \pm \delta \leq \overline{p}$. □

When $z \in J$, both projections leave z unchanged and Proposition 1 follows. When z lies outside J, equation (28) gives the corresponding endpoint directly. In particular, the independent shock keeps its original magnitude in either case.

C.2 Correction strength and the market reference

Define $w = c/(c + q)$ and $a = 1 - w = q/(c + q)$. Before clipping,

$$z = wv + am, \qquad z - m = w(v - m), \qquad z - v = a(m - v). \tag{29}$$

These identities distinguish the remaining distance from the market reference from the displacement away from the greedy price.

Corollary 2 (Limiting behavior of the correction weights)**.** *The weight w belongs to $(0, 1]$, equals one exactly when $\lambda_f Q = 0$, and tends to zero as $\lambda_f Q \to \infty$. Holding the estimates fixed, w is nonincreasing in each factor separately, and strictly decreasing in either factor when the other is positive. If projection is inactive, $u = wv + (1 - w)m$.*

Proof. Differentiating gives

$$\frac{\partial w}{\partial q} = -\frac{c}{(c+q)^2}, \qquad \frac{\partial w}{\partial Q} = -\frac{c\lambda_f}{(c+q)^2}, \qquad \frac{\partial w}{\partial \lambda_f} = -\frac{cQ}{(c+q)^2}.$$

The signs, endpoint value, and limit follow because $c > 0$ and both factors in q are nonnegative. The final representation follows from equation (28). □

Projection preserves a useful version of the contraction. Put $m_J = \mathrm{Proj}(m, J)$. Scalar projection is nonexpansive, so

$$|\widetilde{p} - m_J| = |\,\mathrm{Proj}(z, J) - \mathrm{Proj}(m, J)| \leq w|v - m|. \tag{30}$$

Moreover, $z(q)$ moves monotonically along the segment from v toward m. Clipping is monotone, so its image moves toward m_J, possibly remaining at an endpoint for an interval of weights. Thus stronger correction has a direct local interpretation even at a price boundary.

For two pricing decisions with the same c, q, J but different inputs (v, m) and (v', m'), the same argument yields

$$|\widetilde{p}(v, m) - \widetilde{p}(v', m')| \leq w|v - v'| + (1 - w)|m - m'|. \tag{31}$$

The two coefficients quantify how uncertainty in the greedy price and changes in the historical reference enter a single decision. These are fixed-state comparisons: the queue and estimates are inputs to the displayed map.

The objective also gives an exact accounting of the local adjustment. On the unconstrained quadratic, the estimated revenue forgone in moving from v to z is

$$\widehat{r}(v) - \widehat{r}(z) = ca^2(v - m)^2,$$

whereas the improvement in the penalized objective is

$$f(z) - f(v) = \frac{q^2}{c + q}(v - m)^2. \tag{32}$$

Indeed, equation (27) gives $f(z) - f(v) = (c + q)(v - z)^2$, and substituting $a = q/(c + q)$ proves the formula. At the constrained optimum u, the corresponding comparison $f(u) \geq f(p)$ holds for every feasible alternative $p \in I$.

C.3 When correction reduces oracle error

The rate condition in Corollary 1 concerns distance from the oracle rather than distance from the market reference. The distinction can be expressed exactly. For a fixed feasible oracle price o, let

$$e = v - o, \qquad d = m - v, \qquad z = v + ad.$$

Then

$$(z - o)^2 - (v - o)^2 = 2aed + a^2 d^2. \tag{33}$$

For $a > 0$, the unprojected correction does not increase squared oracle error precisely when

$$ed \le -\frac{a}{2} d^2. \tag{34}$$

This follows by subtracting the two squares; for $a = 0$ the errors are identical. A particularly transparent sufficient condition is that m lie on the segment between v and o. Then z also lies on that segment and cannot be farther from o than v is. If $o \in J$, projecting z onto J cannot increase its distance from o, so the same comparison holds for the pre-shock price.

More generally, when $ed < 0$ and $d \neq 0$, condition (34) is equivalent to $a \le -2e/d$. The formula separates the direction of the correction from its size: a correction directed toward the oracle can pass it and still reduce squared error if its displacement is less than twice the initial distance. Equality at twice that distance preserves the original error. This gives an explicit geometric interpretation of the condition used in the main-text corollary.

C.4 A cumulative error condition with active projection

The exact regret decomposition also permits accumulated-error conditions. Restore the time index and define

$$a_t = \frac{\lambda_f Q_{g_t,t}}{-\widehat{b}_t + \lambda_f Q_{g_t,t}}, \quad e_t = v_t - p_t^\star, \quad d_t = \bar{p}_{t-1} - v_t, \quad z_t = v_t + a_t d_t.$$

Use the positive part of the change in squared error to define

$$E_T = \sum_{t=1}^{T} \mathbb{E}[e_t^2], \qquad H_T = \sum_{t=1}^{T} \mathbb{E}\big[(2a_t e_t d_t + a_t^2 d_t^2)^+\big]. \tag{35}$$

The term H_T counts increases in oracle error; decreases remain beneficial without having to offset an increase at every arrival. Equation (33) implies

$$\sum_{t=1}^{T} \mathbb{E}[(z_t - p_t^\star)^2] \le E_T + H_T. \tag{36}$$

Proposition 3 (Regret with projected corrections)**.** *Under the assumptions of Theorem 1, including nonempty interior price intervals, the implemented FRSD prices satisfy*

$$\bar{R}(T) \le (-b)\{2E_T + 2H_T + 3S_T\}. \tag{37}$$

If every oracle price belongs to its interior interval J_t, the sharper bound is

$$\bar{R}(T) \le (-b)\{E_T + H_T + S_T\}. \tag{38}$$

Consequently, $E_T + H_T \le K_R S_T$ implies $\bar{R}(T) \le (-b)(2K_R + 3)S_T$, together with the estimation bound C/S_T from Theorem 1.

Proof. Let $o_{J,t} = \mathrm{Proj}(p_t^\star, J_t)$. Feasibility of $p_t^\star$ in I gives $|o_{J,t} - p_t^\star| \le \delta_t$. Nonexpansiveness and the triangle inequality give

$$|\widetilde{p}_t - p_t^\star| \le |\,\mathrm{Proj}(z_t, J_t) - \mathrm{Proj}(p_t^\star, J_t)| + \delta_t \le |z_t - p_t^\star| + \delta_t.$$

Squaring, using $(x + y)^2 \le 2x^2 + 2y^2$, and applying equation (36) yields

$$\sum_t \mathbb{E}[(\widetilde{p}_t - p_t^\star)^2] \le 2(E_T + H_T) + 2S_T.$$

Adding the shock contribution in equation (12) proves (37). If $p_t^\star \in J_t$, projection directly gives $|\widetilde{p}_t - p_t^\star| \le |z_t - p_t^\star|$, proving (38). Substitution of the accumulated-error bound proves the final statement. □

This extends the main rate calculation in two directions: individual arrivals can increase oracle error if their total contribution is controlled, and the implemented projection can remain active. When $H_T = 0$ and all projections are inactive, it recovers the mechanism behind Corollary 1. The exploration sum still determines both the estimation bound and the direct shock cost.

D Demand Estimation and Revenue Concentration

Theorem 1 bounds mean squared error. The same shock construction yields probability bounds along the learning process. We derive them directly from the conditional distribution of the shock, so the adaptive choice of the price center remains part of the argument. Throughout this section, the model and bounded-noise assumptions are those of (A1)–(A4), and the known slope interval contains b.

D.1 Conditional moments under adaptive pricing

Let $\mathcal{H}_t = \sigma(\mathcal{F}_{t-1}, x_t, g_t)$, $\mu_t = \theta^\top x_t + \eta_{g_t}$, and $A_t = b\widetilde{p}_t + \mu_t$. Both A_t and the pre-shock center are $\mathcal{H}_t$-measurable. The centered product from the main proof is

$$M_t = \Delta_t D_t - b\delta_t^2 = \Delta_t(A_t + \varepsilon_t). \tag{39}$$

Conditional independence in (A4) gives

$$\mathbb{E}[M_t \mid \mathcal{H}_t] = A_t\mathbb{E}[\Delta_t \mid \mathcal{H}_t] + \mathbb{E}[\Delta_t \mid \mathcal{H}_t]\mathbb{E}[\varepsilon_t \mid \mathcal{H}_t] = 0.$$

Taking another conditional expectation shows $\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] = 0$. Thus $\sum_{s\le t} M_s$ is a martingale in the observed-history filtration, even though its predictable center changes with past observations.

Proposition 4 (Moment bound with explicit constant)**.** *Under (A1)–(A4),*

$$\mathbb{E}[M_t^2 \mid \mathcal{H}_t] = \delta_t^2\{A_t^2 + \mathbb{E}[\varepsilon_t^2 \mid \mathcal{H}_t]\} \le C\delta_t^2, \quad C = (\kappa_b P + M_\mu)^2 + \bar{\varepsilon}^2. \tag{40}$$

Before slope projection, the estimator $\check{b}_{T+1} = S_T^{-1}\sum_{t\le T}\Delta_t D_t$ is unbiased and satisfies the exact identity

$$\mathbb{E}[(\check{b}_{T+1} - b)^2] = \frac{1}{S_T^2}\sum_{t=1}^{T}\delta_t^2\mathbb{E}\{A_t^2 + \mathbb{E}[\varepsilon_t^2 \mid \mathcal{H}_t]\}. \tag{41}$$

Projection onto the known slope interval gives $\mathbb{E}[(\widehat{b}_{T+1} - b)^2] \le C/S_T$.

Proof. Since $\Delta_t^2 = \delta_t^2$, squaring (39) eliminates the shock sign. Expanding $(A_t + \varepsilon_t)^2$ and applying (A3) gives the equality in (40). Assumption (A1) bounds $|A_t|$ by $\kappa_b P + M_\mu$, which proves the inequality. For $s < t$, M_s is $\mathcal{F}_{t-1}$-measurable. Therefore

$$\mathbb{E}[M_s M_t] = \mathbb{E}\{M_s\mathbb{E}[M_t \mid \mathcal{F}_{t-1}]\} = 0.$$

The deterministic denominator S_T and the identity $\check{b}_{T+1} - b = S_T^{-1}\sum_t M_t$ give unbiasedness and (41). Finally, scalar projection onto an interval containing b does not increase distance from b. □

In equation (41), the corrected centers affect finite-sample estimation error through A_t^2, while the deterministic denominator remains S_T. The identity therefore links the pricing decision to estimation precision even though successive centers may depend on one another through the observed history.

D.2 A conditional exponential bound

Write $K = \kappa_b P + M_\mu + \bar{\varepsilon}$. We have $|M_t| \le K\delta_t$ almost surely. For a mean-zero random variable $X \in [-r, r]$, convexity of the exponential gives

$$e^{sX} \le \frac{r + X}{2r}e^{sr} + \frac{r - X}{2r}e^{-sr}, \qquad \mathbb{E}[e^{sX}] \le \cosh(sr) \le e^{s^2r^2/2}.$$

The last inequality follows from $\log\cosh(u) \le u^2/2$: both sides vanish at zero, and for $u \ge 0$ their derivatives obey $\tanh(u) \le u$; symmetry covers negative u. The case $r = 0$ is immediate. Applying this argument conditionally yields

$$\mathbb{E}[e^{sM_t} \mid \mathcal{H}_t] \le \exp(s^2K^2\delta_t^2/2). \tag{42}$$

Iterated conditioning then gives

$$\mathbb{E}\exp\left(s\sum_{t=1}^{T} M_t\right) \le \exp(s^2K^2S_T/2). \tag{43}$$

No independence between successive centers is used: the preceding partial sum is measurable before the next shock, which is exactly what the iteration needs.

Proposition 5 (Finite-sample slope concentration)**.** *For $0 < \beta < 1$, with probability at least $1-\beta$,*

$$|\widehat{b}_{T+1} - b| \le K\sqrt{\frac{2\log(2/\beta)}{S_T}}. \tag{44}$$

More generally, for $x > 0$,

$$\Pr\{|\widehat{b}_{T+1} - b| \ge x\} \le 2\exp\left(-\frac{S_T x^2}{2K^2}\right). \tag{45}$$

Proof. For $y > 0$ and $s > 0$, Markov's inequality and (43) give

$$\Pr\{\textstyle\sum_t M_t \ge y\} \le \exp(-sy + s^2K^2S_T/2).$$

The minimizing value is $s = y/(K^2S_T)$, giving $\exp[-y^2/(2K^2S_T)]$. Apply the same argument to $-M_t$, add the two tails, and put $y = S_T x$. The projected error is no larger than the unprojected error on every path. Choosing x to make the last bound equal to β proves (44). □

The resulting interval can be intersected with $[\underline{b}, \overline{b}]$. On the event in (44), that intersection still contains the true slope. The bound concerns the same estimator as Algorithm 1; no alternative fitting rule is introduced.

D.3 Bounds throughout the experiment

For each $n \ge 1$, allocate error probability $\beta_n = \beta/[n(n+1)]$. Since $\sum_{n\ge1} 1/[n(n+1)] = 1$, applying (44) at every n and taking a union bound proves the following statement.

Corollary 3 (Simultaneous slope bounds)**.** *With probability at least $1-\beta$, for all integers $n \ge 1$ on which the experiment is defined,*

$$|\widehat{b}_{n+1} - b| \le K\sqrt{\frac{2\log(2n(n+1)/\beta)}{S_n}}. \tag{46}$$

The same event gives the bound at any almost surely finite, history-dependent observation count, with that count substituted for n.

Proof. The probability that at least one fixed-n inequality fails is at most $\sum_n \beta_n = \beta$. On its complement every inequality holds at once. Selecting one of those integer indices, including by a rule based on observations, preserves the inequality on that event. □

The exploration exponent determines the scale of these bounds. With $r = 2\alpha \in [0, 1/2]$, monotonicity of x^{-r} gives

$$\frac{(T+1)^{1-r} - 1}{1-r} \le S_T \le 1 + \frac{T^{1-r} - 1}{1-r}. \tag{47}$$

The lower bound sums integrals over $[t, t+1]$; the upper bound treats $t = 1$ separately and sums over $[t-1, t]$. Thus S_T has order $T^{1-2\alpha}$, and the fixed-horizon confidence radius has order $T^{-1/2+\alpha}$ for fixed β. At $\alpha = 0$, $S_T = T$ exactly. At $\alpha = 1/4$, it has order $\sqrt{T}$.

A simple finite-sample condition follows from $S_T \ge T^{1-2\alpha}$, since every term $t^{-2\alpha}$ is at least $T^{-2\alpha}$. To make the radius in (44) at most $\epsilon > 0$, it suffices that

$$T \ge \left(\frac{2K^2\log(2/\beta)}{\epsilon^2}\right)^{1/(1-2\alpha)}. \tag{48}$$

These formulas put the exploration–precision relationship in finite-sample terms; the main regret decomposition assigns the same S_T its revenue cost.

D.4 Pathwise revenue and the expected decomposition

Let $W = \overline{p} - \underline{p}$ and define the noise-free loss along the realized price path by

$$G_T = \sum_{t=1}^{T} \{r_t(p_t^\star) - r_t(p_t)\}, \qquad V_T = (-b) \sum_{t=1}^{T} \{(\widetilde{p}_t - p_t^\star)^2 + \delta_t^2\}.$$

Both quantities are random because the price centers depend on the history. Completing the square before taking expectations gives the exact identity

$$G_T - V_T = 2(-b) \sum_{t=1}^{T} (\widetilde{p}_t - p_t^\star)\Delta_t. \tag{49}$$

The summands on the right have conditional mean zero and magnitude at most $2(-b)W\delta_t$. Repeating the exponential argument above therefore gives, with probability at least $1-\beta$,

$$|G_T - V_T| \le 2(-b)W\sqrt{2S_T \log(2/\beta)}. \tag{50}$$

Taking expectations in (49) recovers Theorem 1: $\mathbb{E}[G_T] = \mathbb{E}[V_T] = \bar{R}(T)$.

Under the additive-noise model, evaluate the oracle and posted-price revenues with the same ε_t. The realized revenue difference $\mathcal{R}_T$ then satisfies

$$\mathcal{R}_T - G_T = \sum_{t=1}^{T} (p_t^\star - p_t)\varepsilon_t. \tag{51}$$

Conditioning first on $\mathcal{H}_t$ and Δ_t makes the price difference fixed. Assumptions (A3)–(A4) give $\mathbb{E}[\varepsilon_t \mid \mathcal{H}_t, \Delta_t] = 0$, and the summand has magnitude at most $W\bar{\varepsilon}$. Another application of the same argument gives

$$\Pr\{|\mathcal{R}_T - G_T| > W\bar{\varepsilon}\sqrt{2T\log(2/\beta)}\} \le \beta.$$

Combining the last two bounds with error probabilities $\beta/2$ gives

$$|\mathcal{R}_T - V_T| \le 2(-b)W\sqrt{2S_T\log(4/\beta)} + W\bar{\varepsilon}\sqrt{2T\log(4/\beta)} \tag{52}$$

with probability at least $1-\beta$. The two terms separately account for random exploration and demand noise. The predictable quantity V_T retains the center-error term of the main decomposition, rather than replacing it by an unconditional rate.

D.5 Conditionally sub-Gaussian demand noise

The identification argument also applies to unbounded noise with controlled conditional tails. This includes Gaussian noise, as used in the simulation construction. Retain bounded model parameters, feasible posted prices, and exogenous shock signs. In place of the almost-sure noise bound, suppose that, for a constant $\sigma \ge 0$,

$$\mathbb{E}[e^{s\varepsilon_t} \mid \mathcal{H}_t] \le e^{s^2\sigma^2/2} \quad \text{for all } s \in \mathbb{R}, \qquad \Delta_t \text{ is independent of } \varepsilon_t \text{ conditional on } \mathcal{H}_t. \tag{53}$$

The exponential bound entails zero conditional mean and conditional variance at most σ^2. Define

$$K_\sigma^2 = (\kappa_b P + M_\mu)^2 + \sigma^2.$$

Proposition 6 (Slope guarantees with conditional noise tails)**.** *Under* (53) *and the stated parameter, price, and shock conditions, for every $x > 0$,*

$$\mathbb{E}[(\widehat{b}_{T+1} - b)^2] \le \frac{K_\sigma^2}{S_T}, \qquad \Pr\{|\widehat{b}_{T+1} - b| \ge x\} \le 2\exp\left(-\frac{S_T x^2}{2K_\sigma^2}\right). \tag{54}$$

The simultaneous bound (46) *holds with K_σ in place of K.*

Proof. The identity $M_t = \Delta_t(A_t + \varepsilon_t)$ and its zero conditional mean are unchanged. Because $\Delta_t^2 = \delta_t^2$,

$$\mathbb{E}[M_t^2 \mid \mathcal{H}_t] = \delta_t^2\{A_t^2 + \mathbb{E}[\varepsilon_t^2 \mid \mathcal{H}_t]\} \le \delta_t^2 K_\sigma^2.$$

Orthogonality of the martingale differences and slope projection give the mean squared error bound as before. For the tail bound, conditional independence allows conditioning on the two shock signs:

$$\mathbb{E}[e^{sM_t} \mid \mathcal{H}_t] = \tfrac{1}{2}e^{s\delta_t A_t}\mathbb{E}[e^{s\delta_t \varepsilon_t} \mid \mathcal{H}_t] + \tfrac{1}{2}e^{-s\delta_t A_t}\mathbb{E}[e^{-s\delta_t \varepsilon_t} \mid \mathcal{H}_t]$$
$$\leq e^{s^2\delta_t^2\sigma^2/2}\cosh(s\delta_t A_t) \leq e^{s^2\delta_t^2 K_\sigma^2/2}.$$

The iterated exponential and tail calculations used in (43)–(45) now apply with K_σ. The same summable allocation of error probabilities proves the simultaneous statement. □

For conditionally Gaussian noise with mean zero and variance at most σ^2, completing the square in its density gives exactly the exponential bound in (53). Thus boundedness of the noise itself is not needed for the slope results in this extension. The mechanism remains the independent, zero-mean price shock and the deterministic squared-shock sum.

The revenue identities also retain their form when the oracle is feasible. The exploration term in (49) does not involve demand noise. For the additional noise term in (51), conditioning on $\mathcal{H}_t, \Delta_t$ gives

$$\mathbb{E}[e^{s(p_t^\star - p_t)\varepsilon_t} \mid \mathcal{H}_t, \Delta_t] \leq \exp\{s^2\sigma^2(p_t^\star - p_t)^2/2\} \leq e^{s^2\sigma^2 W^2/2}.$$

Consequently (52) holds with σ in place of $\bar{\varepsilon}$. Taking conditional expectations still yields the exact expected regret decomposition. This extension concerns the same linear demand model and preserves the separation between predictable pricing decisions and independent exploration.

D.6 How the slope update enters the remaining fit

After updating the common slope, the algorithm uses it to refit the context and group effects. To see how slope error enters this fit, stack the context and group regressors through time n into a matrix Z, their true coefficients into a vector γ, prices into $\mathbf{p}$, and noise into $\boldsymbol{\varepsilon}$. Thus

$$\mathbf{D} = b\mathbf{p} + Z\gamma + \boldsymbol{\varepsilon}.$$

For a positive definite ridge matrix Λ, the update using $\widehat{b}_{n+1}$ is

$$V = Z^\top Z + \Lambda, \qquad \widehat{\gamma} = V^{-1}Z^\top(\mathbf{D} - \widehat{b}_{n+1}\mathbf{p}).$$

Substitution yields the exact finite-sample identity

$$\widehat{\gamma} - \gamma = V^{-1}\{Z^\top\boldsymbol{\varepsilon} + (b - \widehat{b}_{n+1})Z^\top\mathbf{p} - \Lambda\gamma\}. \tag{55}$$

It holds for the realized adaptive design; no conditioning on the entire design is needed. For any query regressor z, Cauchy–Schwarz in the V^{-1} norm gives

$$|z^\top(\widehat{\gamma} - \gamma)| \leq \|z\|_{V^{-1}}\big(\|Z^\top\boldsymbol{\varepsilon}\|_{V^{-1}} + |b - \widehat{b}_{n+1}|\|Z^\top\mathbf{p}\|_{V^{-1}} + \|\Lambda\gamma\|_{V^{-1}}\big). \tag{56}$$

On the event in (44), its radius can be substituted for the middle coefficient. This identifies exactly how slope uncertainty propagates into the next demand prediction, alongside observation noise and regularization. It complements the slope guarantee without requiring a separate rate assumption for the context design.

D.7 From prediction error to greedy-price error

The preceding identities can be connected to the greedy-price term in Corollary 1. Let $\beta_b = -\bar{b} > 0$ be the smallest absolute slope allowed by the known negative interval. At arrival t, write $\widehat{\mu}_t = \widehat{\mu}_t(g_t, x_t)$ and $\mu_t = \theta^\top x_t + \eta_{g_t}$. Using $\mu_t = -2bp_t^\star$ gives

$$v_t - p_t^\star = -\frac{\widehat{\mu}_t}{2\widehat{b}_t} + \frac{\mu_t}{2b} \tag{57}$$

$$= -\frac{(\widehat{\mu}_t - \mu_t) + 2p_t^\star(\widehat{b}_t - b)}{2\widehat{b}_t}. \tag{58}$$

Since $|\widehat{b}_t| \geq \beta_b$ and $|p_t^\star| \leq P$,

$$|v_t - p_t^\star| \leq \frac{|\widehat{\mu}_t - \mu_t| + 2P|\widehat{b}_t - b|}{2\beta_b}. \tag{59}$$

The denominator shows why projecting the estimated slope onto a known negative interval also stabilizes the pricing step: errors in the estimated intercept and slope enter with a controlled scale.

Proposition 7 (Prediction errors sufficient for the greedy-price budget)**.** *Under the feasible-oracle model and slope projection, the accumulated greedy-price error satisfies*

$$E_T \le \frac{1}{2\beta_b^2}\left\{\sum_{t=1}^{T}\mathbb{E}[(\widehat{\mu}_t-\mu_t)^2]+4P^2\sum_{t=1}^{T}\mathbb{E}[(\widehat{b}_t-b)^2]\right\}. \tag{60}$$

Under the slope-estimation assumptions above, its second sum is bounded by

$$\sum_{t=1}^{T}\mathbb{E}[(\widehat{b}_t-b)^2] \le (\bar{b}-\underline{b})^2 + C\sum_{n=1}^{T-1}\frac{1}{S_n}, \tag{61}$$

with $C = K_\sigma^2$ for the conditional sub-Gaussian extension. Thus, for the exploration exponents in the main text, a bound $\sum_t \mathbb{E}[(\widehat{\mu}_t-\mu_t)^2] = O(S_T)$ implies $E_T = O(S_T)$.

Proof. Square (59) and use $(x+y)^2 \le 2x^2+2y^2$, then take expectations and sum. The initial slope estimate and true slope lie in the same interval, so their squared difference is at most its squared width. At arrival $t \ge 2$, the estimate uses $t-1$ completed observations and has mean squared error at most C/S_{t-1}, proving (61). For $\alpha = 0$, $S_n = n$ and $\sum_{n=1}^{T-1} 1/S_n \le 1+\log T$. For $0 < \alpha \le 1/4$, use $S_n \ge n^{1-2\alpha}$ and the integral comparison to obtain

$$\sum_{n=1}^{T-1}\frac{1}{S_n} \le 1+\frac{T^{2\alpha}-1}{2\alpha}. \tag{62}$$

For each fixed $\alpha \in (0, 1/4]$, this is $O(T^{2\alpha})$, which is $O(S_T)$ because $2\alpha \le 1-2\alpha$. The logarithmic bound at zero is also $O(S_T)$. Substitution into (60) proves the last assertion. □

The price-error condition can thus be traced to the algorithm's two estimation updates. Equation (56) identifies the components of $\widehat{\mu}_t - \mu_t$, while (61) bounds the accumulated slope error. If the cumulative prediction error and the correction-increase budget H_T are both of order S_T, Proposition 3 gives that regret order for the implemented projected prices.

There is a corresponding pathwise implication. Given error radii $r_t^\mu, r_t^b \ge 0$, on any event where $|\widehat{\mu}_t-\mu_t| \le r_t^\mu$ and $|\widehat{b}_t - b| \le r_t^b$, equation (59) gives

$$|v_t - p_t^\star| \le (r_t^\mu + 2Pr_t^b)/(2\beta_b).$$

The simultaneous slope bound provides the second radius for all arrivals after initialization; a prediction-error bound for the fitted context and group effects supplies the first. This separates what is controlled by independent shocks from what is controlled by the observed covariate design.

E Queues and Group Price Histories

The virtual queue carries earlier price differences into later decisions. To make this dependence explicit, write

$$h_{g,s} = |\bar{p}_{g,s} - \bar{p}_s|, \qquad a_{g,s} = h_{g,s} - \tau, \qquad Q_{g,s+1} = [Q_{g,s} + a_{g,s}]^+.$$

Here s counts completed arrivals, so $Q_{g,s+1}$ is the queue available for the next price. The initialization is $Q_{g,1} = 0$. Group averages follow the unobserved-group convention in Section 4.

E.1 An exact representation of the price history

Proposition 8 (Queue as accumulated excess price difference)**.** *For every realized price path and $n \ge 1$,*

$$Q_{g,n+1} = \max\left\{0, \max_{1\le k\le n}\sum_{s=k}^{n}(h_{g,s}-\tau)\right\}. \tag{63}$$

Thus the queue records the largest positive excess over the tolerance among suffixes of the observed history.

Proof. For $n = 1$, the formula is the defining update $Q_{g,2} = [a_{g,1}]^+$. Suppose it holds at $n-1$. Adding $a_{g,n}$ and taking the positive part gives

$$Q_{g,n+1} = \max\{0, a_{g,n}, a_{g,n} + a_{g,n-1}, \dots, a_{g,n} + \cdots + a_{g,1}\}.$$

These are exactly the empty suffix and all suffix sums ending at n. Induction proves the claim. □

This representation explains why a queue can remain positive after the current gap drops below τ. A positive suffix sum can still contain earlier excess differences. A run of below-tolerance gaps clears the queue once the accumulated shortfall reaches its value at the start of the run.

The same formula supplies quantitative statements. If, for a block of k updates beginning with $Q_{g,\ell}$, every gap satisfies $h_{g,s} \le \tau - \eta$ for some $\eta > 0$, then

$$Q_{g,\ell+k} \le [Q_{g,\ell} - k\eta]^+. \tag{64}$$

This follows by applying the monotone map $x \mapsto [x - \eta]^+$ repeatedly, using $[[x-a]^+ - b]^+ = [x - a - b]^+$ for $a, b \ge 0$. If instead every gap on the block is at least $\tau + \eta$, then no reflection is needed and

$$Q_{g,\ell+k} \ge Q_{g,\ell} + k\eta. \tag{65}$$

Together, these formulas make the accumulation and release of correction strength explicit on any observed path.

E.2 Gap bounds over arbitrary time windows

Proposition 9 (Pointwise queue–gap bound)**.** *For $t \ge 2$,*

$$|\bar{p}_{g,t-1} - \bar{p}_{t-1}| \le \tau + Q_{g,t} - Q_{g,t-1}. \tag{66}$$

More generally, for $1 \le \ell \le r$,

$$\sum_{s=\ell}^{r} h_{g,s} \le (r - \ell + 1)\tau + Q_{g,r+1} - Q_{g,\ell}. \tag{67}$$

Proof. The defining update is at least its unreflected argument, so $Q_{g,s+1} - Q_{g,s} \ge h_{g,s} - \tau$. This gives the pointwise result with $s = t - 1$; summation over the window telescopes to (67). □

The role of reflection can also be displayed exactly. Define

$$k_{g,s} = [-Q_{g,s} - h_{g,s} + \tau]^+.$$

Then

$$Q_{g,s+1} = Q_{g,s} + h_{g,s} - \tau + k_{g,s}, \qquad k_{g,s} Q_{g,s+1} = 0. \tag{68}$$

The second identity follows because $k_{g,s}$ is positive only when the unreflected update is negative. Summing the first identity gives

$$\sum_{s=\ell}^{r} h_{g,s} = (r - \ell + 1)\tau + Q_{g,r+1} - Q_{g,\ell} - \sum_{s=\ell}^{r} k_{g,s}. \tag{69}$$

The queue bound therefore discards only nonnegative reflection terms. Setting $\ell = 1$, $r = T - 1$, and $Q_{g,1} = 0$ gives the running-gap statement of Theorem 1. Setting $r = T$ instead uses $Q_{g,T+1}$ and includes the gap after the final observation, as in an average over all T observed price histories.

E.3 Group-to-market and pairwise differences

At a completed arrival count s, let $\pi_{g,s} = N_g(s)/s$ for the groups observed so far. Their means satisfy

$$\bar{p}_s = \sum_g \pi_{g,s} \bar{p}_{g,s}, \qquad \bar{p}_{g,s} - \bar{p}_s = \sum_h \pi_{h,s} (\bar{p}_{g,s} - \bar{p}_{h,s}). \tag{70}$$

The identity follows by grouping the sum of all posted prices by customer group. Terms with zero counts have zero weight. For any two groups with defined means, the triangle inequality gives

$$|\bar{p}_{g,s} - \bar{p}_{h,s}| \le h_{g,s} + h_{h,s}. \tag{71}$$

Combining this with (67) gives a window bound for pairwise group differences:

$$\sum_{s=\ell}^{r} |\bar{p}_{g,s} - \bar{p}_{h,s}| \le 2(r-\ell+1)\tau + Q_{g,r+1} + Q_{h,r+1} - Q_{g,\ell} - Q_{h,\ell}. \tag{72}$$

With the manuscript's convention for an unobserved group, the triangle inequality continues to hold before its first arrival.

For two observed groups A, B, equation (70) is more precise:

$$\begin{aligned} h_{A,s} = \pi_{B,s}|\bar{p}_{A,s} - \bar{p}_{B,s}|, \qquad h_{B,s} = \pi_{A,s}|\bar{p}_{A,s} - \bar{p}_{B,s}|, \\ h_{A,s} + h_{B,s} = |\bar{p}_{A,s} - \bar{p}_{B,s}|. \end{aligned} \tag{73}$$

The group with larger sample share has a mean closer to the market mean at the same pairwise gap. Its queue input is consequently smaller. This is the exact group-weighting relation behind the queue diagnostics in Appendix B.4.

There is also a squared version for any number of observed groups:

$$\sum_{g} \pi_{g,s}(\bar{p}_{g,s} - \bar{p}_s)^2 = \frac{1}{2}\sum_{g,h} \pi_{g,s}\pi_{h,s}(\bar{p}_{g,s} - \bar{p}_{h,s})^2. \tag{74}$$

Expanding the right-hand side gives $\sum_g \pi_{g,s}\bar{p}_{g,s}^2 - \bar{p}_s^2$, which is the left-hand side. Equations (73) and (74) relate the algorithm's market-reference quantities to pairwise fairness measures without changing their definitions.

E.4 The next price and the next group averages

Suppose group g has already appeared n_g times among s arrivals. Posting price p to that group updates its mean and the market mean by

$$\bar{p}_{g,s+1} = \bar{p}_{g,s} + \frac{p - \bar{p}_{g,s}}{n_g + 1}, \qquad \bar{p}_{s+1} = \bar{p}_s + \frac{p - \bar{p}_s}{s+1}. \tag{75}$$

Subtracting gives the signed group-to-market difference after that offer:

$$\bar{p}_{g,s+1} - \bar{p}_{s+1} = \bar{p}_{g,s} - \bar{p}_s + \frac{p - \bar{p}_{g,s}}{n_g + 1} - \frac{p - \bar{p}_s}{s+1}. \tag{76}$$

For an already observed group $h \neq g$, its group mean stays fixed, so

$$\bar{p}_{h,s+1} - \bar{p}_{s+1} = \bar{p}_{h,s} - \bar{p}_s - \frac{p - \bar{p}_s}{s+1}. \tag{77}$$

These are signed identities; the queue takes the absolute value afterward. They show how one offer enters both its own group's history and the shared reference used by the other groups.

For two alternative offers p, p' at the same history, equation (76) gives

$$|e_g(p) - e_g(p')| = \left(\frac{1}{n_g + 1} - \frac{1}{s+1}\right)|p - p'|,$$

where e_g denotes the updated signed gap. The coefficient is nonnegative because $n_g \le s$. The corresponding coefficient for another observed group is $1/(s+1)$. Since both absolute value and positive-part reflection are 1-Lipschitz, the same expressions bound changes in their next queues when the current queues are held fixed. Thus a bounded change in the offer produces a quantitatively controlled change in the next correction inputs.

E.5 Sensitivity to a history or a tolerance

For two nonnegative gap sequences h_s, h'_s and tolerances τ, τ', initialized with zero queues, the queue map obeys

$$|Q_{n+1} - Q'_{n+1}| \le \sum_{s=1}^{n} |h_s - h'_s| + n|\tau - \tau'|. \tag{78}$$

Indeed, the positive-part map is nonexpansive, so each update increases the difference by at most $|h_s - h'_s| + |\tau - \tau'|$. Induction proves the result. On a fixed history, increasing the tolerance decreases every suffix sum in (63); therefore $\tau' \ge \tau$ implies $Q_{n+1}(\tau') \le Q_{n+1}(\tau)$.

The dependence on posted prices can be made explicit as well. Take two price paths with the same group arrivals and write $d_j = |p_j - p'_j|$. Whenever $N_g(s) > 0$,

$$|h_{g,s} - h'_{g,s}| \le \frac{1}{N_g(s)} \sum_{j \le s : g_j = g} d_j + \frac{1}{s} \sum_{j=1}^{s} d_j. \tag{79}$$

This follows from the reverse triangle inequality applied to the two signed gaps, followed by the triangle inequality for their group and market averages. Before the first group arrival both gaps are zero by convention. Substituting (79) into (78) connects perturbations of an entire price history to the resulting queue differences. Together with (31), these identities describe both directions of the feedback: histories change queues, and queues change later offers.

E.6 Terminal disparity from the queue history

The window bounds also control the disparity at the end of the experiment. The link is the limited change in a group average over the last few arrivals: if two terminal averages are far apart, their difference must persist over a preceding window, where it contributes to the queues. This argument uses only bounded posted prices and the queue recursion, so it applies to the implemented prices with active projection and random shocks.

Write $W = \overline{p} - \underline{p} > 0$, and suppose all groups have been observed by T. For distinct groups g, h, put $m_{gh} = \min\{N_g(T), N_h(T)\}$. The queue $Q_{g,T+1}$ is the update after the final observation, as in Algorithm 1; computing it requires no additional offer.

Proposition 10 (Terminal group-price disparity)**.** *For every realized price path and every integer* $1 \le w \le m_{gh}$*,*

$$|\bar{p}_{g,T} - \bar{p}_{h,T}| \le 2\tau + \frac{Q_{g,T+1} + Q_{h,T+1} - Q_{g,T-w+1} - Q_{h,T-w+1}}{w} + \frac{W(w-1)}{2m_{gh}}. \tag{80}$$

The minimum of these bounds over w*, followed by the maximum over group pairs, bounds* $\Delta(\mathbf{p})$*.*

Proof. Let $\ell = T - w + 1$ and $\ell \le s \le T$. Since $w \le m_{gh}$, both groups have appeared by ℓ. The terminal group mean is a convex combination of its mean at s and the prices offered to that group after s. Consequently,

$$|\bar{p}_{g,T} - \bar{p}_{g,s}| \le W \frac{N_g(T) - N_g(s)}{N_g(T)}.$$

The same inequality holds for h. There are at most $T - s$ later arrivals in the two groups combined, so

$$|\bar{p}_{g,T} - \bar{p}_{h,T}| \le |\bar{p}_{g,s} - \bar{p}_{h,s}| + \frac{W(T-s)}{m_{gh}}.$$

Average this inequality over $s = \ell, \dots, T$. Equation (72) bounds the first term, while $w^{-1} \sum_{s=\ell}^{T} (T - s) = (w-1)/2$ gives the second. This proves (80) simultaneously for every eligible window and pair. □

Short windows keep the averaging remainder small; longer windows divide the accumulated queue growth over more arrivals. The realized queue history permits this choice after observing the path, since the inequalities hold pathwise for all windows.

Corollary 4 (Sublinear queues and terminal disparity)**.** *Let* $m_T = \min_g N_g(T)$ *and* $q_T = \max_{g\neq h}(Q_{g,T+1} + Q_{h,T+1})$*, for at least two observed groups. Then, for each integer* $1 \le w \le m_T$*,*

$$\Delta(\mathbf{p}) \le 2\tau + \frac{q_T}{w} + \frac{W(w-1)}{2m_T}. \tag{81}$$

If $q_T \le Wm_T/2$*, this implies*

$$\Delta(\mathbf{p}) \le 2\tau + \sqrt{\frac{2Wq_T}{m_T}}. \tag{82}$$

In particular, along any sequence of paths with $m_T \ge \pi_{\min} T$ *and* $q_T = o(T)$*, the terminal disparity is at most* $2\tau + o(1)$*. If* $q_T = O(T^\beta)$ *for some* $0 \le \beta < 1$*, the remainder is* $O(T^{-(1-\beta)/2})$*.*

Proof. Drop the nonnegative initial queues from (80) and use the common count lower bound m_T. If $q_T = 0$, take $w = 1$. Otherwise, set $a = \sqrt{2m_T q_T/W}$ and $w = \lceil a \rceil$. Under the stated upper bound on q_T, $1 \le w \le m_T$, and

$$\frac{q_T}{w} + \frac{W(w-1)}{2m_T} \le \frac{q_T}{a} + \frac{Wa}{2m_T} = \sqrt{\frac{2Wq_T}{m_T}}.$$

This proves (82). The coverage condition and $q_T = o(T)$ ensure its premise for all sufficiently large T; substitution gives both asymptotic statements. □

The factor two converts the group-to-market tolerance in the queue into a pairwise tolerance. For two groups, their sample shares give a sharper conversion. Let $\pi_{A,T} = N_A(T)/T$ and $\pi_{B,T} = N_B(T)/T$. For $g \in \{A, B\}$, let h be the other group. For every integer $1 \le w \le N_g(T)$,

$$\Delta(\mathbf{p}) \le \frac{1}{\pi_{h,T}} \left\{ \tau + \frac{Q_{g,T+1} - Q_{g,T-w+1}}{w} + \frac{W(w-1)}{2} \left(\frac{1}{N_g(T)} + \frac{1}{T} \right) \right\}. \tag{83}$$

Indeed, the group-mean argument above and its market-mean counterpart give

$$h_{g,T} \le h_{g,s} + W(T-s) \left(\frac{1}{N_g(T)} + \frac{1}{T} \right).$$

Averaging over the window, applying (67), and using $h_{g,T} = \pi_{h,T}\Delta(\mathbf{p})$ proves (83). With positive group coverage and both terminal queues $o(T)$, choose integer windows with $\max_g Q_{g,T+1} = o(w)$ and $w = o(T)$. Minimizing over the two groups then gives

$$\Delta(\mathbf{p}) \le \frac{\tau}{\max\{\pi_{A,T}, \pi_{B,T}\}} + o(1).$$

These results connect observed queue growth to the terminal criterion without changing the pricing rule. The finite-window bounds require no queue-growth assumption; the vanishing remainders use the sublinear growth stated in the corollary.

E.7 Terminal fairness of the implemented algorithm

The window bounds apply to any price path satisfying the queue recursion. We now use the pricing decision itself to control terminal disparity. Define the variation between group means by

$$B_t = \sum_{g \in G} N_g(t)(\bar{p}_{g,t} - \bar{p}_t)^2, \qquad B_0 = 0.$$

Unlike the variance of individual prices, this quantity records only differences between group averages. Its increments connect the queue update with the effect of the next price.

Bounded greedy prices. Put $c_- = -\overline{b} > 0$ and $c_+ = -\underline{b}$. Under the bounded model, $|D_t| \le D_{\max}$ for a fixed constant. The ridge update is

$$\widehat{\gamma}_{n+1} = V_n^{-1} \sum_{s=1}^{n} \phi_s (D_s - \widehat{b}_{n+1} p_s).$$

For $n \ge n_0$, the pathwise design condition of Theorem 2 gives $\|\widehat{\gamma}_{n+1}\| \le L(D_{\max} + c_+ P)/\kappa$. The finite prefix is also bounded because $V_n \succeq \Lambda \succ 0$ and the initialization is bounded. Since $|\widehat{b}_t| \ge c_-$, the greedy prices are uniformly bounded. In particular, there is a fixed M with $|v_t - \bar{p}_{t-1}| \le M$, independent of T and τ. Write $K = c_+ M$ and $W = \overline{p} - \underline{p}$.

An exact update for group variation. At arrival t, abbreviate $m = \bar{p}_{t-1}$, $m_g = \bar{p}_{g_t,t-1}$, $d = m_g - m$, and $e = p_t - m$. As in the algorithm, set $m_g = m$ for an unobserved group. Updating $\sum_g N_g \bar{p}_g^2 - t\bar{p}^2$ gives

$$B_t - B_{t-1} = -d^2 + 2ed + \frac{(p_t - m_g)^2}{N_{g_t}(t)} - \frac{e^2}{t}. \tag{84}$$

The identity includes a group's first arrival and $t = 1$: a group with prior count zero contributes p_t^2 after the update. The penultimate term is at most $W^2/N_{g_t}(t)$, and these terms sum to at most

$$W^2 \sum_{g \in G} \sum_{j=1}^{N_g(T)} j^{-1} \le |G| W^2 (1 + \log T). \tag{85}$$

Accumulated gaps and correction strength. For $h_{g,t-1} = |\bar{p}_{g,t-1} - \bar{p}_{t-1}|$ and $t \ge 2$, the reflected update implies $h_{g,t-1} \le Q_{g,t} - Q_{g,t-1} + \tau$. Using $(x - y)/x \le \log(x/y)$ for $x, y > 0$ yields

$$\frac{h_{g,t-1}}{c_- + \lambda_f Q_{g,t}} \le \frac{1}{\lambda_f} \log \frac{c_- + \lambda_f Q_{g,t}}{c_- + \lambda_f Q_{g,t-1}} + \frac{\tau}{c_-}. \tag{86}$$

This step includes decreasing queues and reflection at zero. Sum over all groups and arrivals; the arriving-group terms are nonnegative and form a subset of this sum. Since $Q_{g,T} \le W(T-1)$, and the first-arrival gap is zero,

$$\sum_{t=1}^{T} \frac{h_{g_t,t-1}}{c_- + \lambda_f Q_{g_t,t}} \le \frac{|G|}{\lambda_f} \log \left(1 + \frac{\lambda_f W T}{c_-}\right) + \frac{|G| \tau T}{c_-}. \tag{87}$$

This is where accumulating past gaps enters the fairness bound: a larger queue reduces the contribution of a persistent gap to later departures from the market mean.

Projection of the market reference. Let $J_t = [\underline{p} + \delta_t, \overline{p} - \delta_t]$ and $\beta_t = \operatorname{dist}(\bar{p}_{t-1}, J_t)$, with $\bar{p}_0 \in [\underline{p}, \overline{p}]$. Proposition 2 and nonexpansiveness imply

$$|\widetilde{p}_t - \bar{p}_{t-1}| \le \frac{K}{c_- + \lambda_f Q_{g_t,t}} + \beta_t, \qquad 0 \le \beta_t \le \delta_t. \tag{88}$$

For fixed $0 < \alpha \le 1/4$, the expected sum of β_t^2 is finite. Set

$$Z_n = \sum_{s=1}^{n} \Delta_s, \qquad \ell_n = \sum_{s=1}^{n} \delta_s - n\delta_{n+1} > 0.$$

All past centers lie in their interior intervals, so

$$\underline{p} + \frac{1}{n} \sum_{s=1}^{n} \delta_s + \frac{Z_n}{n} \le \bar{p}_n \le \overline{p} - \frac{1}{n} \sum_{s=1}^{n} \delta_s + \frac{Z_n}{n}.$$

Hence $\beta_{n+1} > 0$ requires $|Z_n| > \ell_n$. Iterating the conditional moment-generating functions of the equiprobable signs gives $\mathbb{E}e^{uZ_n} \le e^{u^2 S_n/2}$, and thus $\Pr(|Z_n| > \ell_n) \le 2e^{-\ell_n^2/(2S_n)}$. Now $\ell_n \sim \alpha n^{1-\alpha}/(1-\alpha)$ and $S_n \sim n^{1-2\alpha}/(1-2\alpha)$; the tail bound decreases exponentially for sufficiently large n. Therefore

$$\sum_{t \ge 1} \mathbb{E}\beta_t^2 \le C_\alpha := 1 + 2 \sum_{n \ge 1} \delta_{n+1}^2 e^{-\ell_n^2/(2S_n)} < \infty. \tag{89}$$

The first term bounds β_1^2. This constant may depend on α; the argument does not include $\alpha = 0$.

Proof of Theorem 3. Write $e = (\widetilde{p}_t - m) + \Delta_t$ in (84). The conditional mean of $2d\Delta_t$ is zero. With $h = |d|$, completion of the square in (88) gives

$$-d^2 + 2d(\widetilde{p}_t - m) \le \frac{2Kh}{c_- + \lambda_f Q_{g_t,t}} - h^2 + 2\beta_t h \le \frac{2Kh}{c_- + \lambda_f Q_{g_t,t}} + \beta_t^2.$$

Summing the update, discarding its nonpositive last term, and applying (85), (87), and (89), we obtain

$$\mathbb{E}B_T \le \frac{2K|G|}{\lambda_f} \log\left(1 + \frac{\lambda_f WT}{c_-}\right) + \frac{2K|G|\tau T}{c_-} + |G|W^2(1 + \log T) + C_\alpha. \tag{90}$$

For every observed pair, weighted Cauchy–Schwarz gives

$$|\bar{p}_{g,T} - \bar{p}_{h,T}|^2 \le \left(\frac{1}{N_g(T)} + \frac{1}{N_h(T)}\right) B_T.$$

Under coverage, the maximum is at most $2B_T/(\pi_{\min}T)$. Consequently,

$$\begin{aligned}\mathbb{E}\Delta(\mathbf{p})^2 \le{}& \frac{4K|G|\tau}{c_-\pi_{\min}} + \frac{4K|G|}{\lambda_f \pi_{\min} T} \log\left(1 + \frac{\lambda_f WT}{c_-}\right) \\ &+ \frac{2|G|W^2(1 + \log T) + 2C_\alpha}{\pi_{\min} T}.\end{aligned} \tag{91}$$

For fixed $\lambda_f > 0$, this is $O(\tau + \log(eT)/T)$, with constants independent of T and τ. At $\tau = 0$ it proves the asserted root-mean-square rate. For positive tolerance it quantifies the effect of τ on the bound; it does not identify the queue threshold with an identical terminal threshold.

Pathwise convergence at zero tolerance. Without averaging over shock signs, (88) implies $|e| \le K/(c_- + \lambda_f Q_{g_t,t}) + 2\delta_t$. Using $-h^2 + 4\delta_t h \le 4\delta_t^2$ in the same update gives

$$B_T \le \frac{2K|G|}{\lambda_f} \log\left(1 + \frac{\lambda_f WT}{c_-}\right) + \frac{2K|G|\tau T}{c_-} + 4S_T + |G|W^2(1 + \log T). \tag{92}$$

At $\tau = 0$, coverage and $\alpha > 0$ yield $\Delta(\mathbf{p}) = O(T^{-\alpha} + \sqrt{\log(eT)/T})$ on each path satisfying the boundedness and design conditions, proving almost-sure convergence under the probabilistic model. If coverage holds at all sufficiently large prefixes, $h_{g,t} \le \Delta(\mathbf{p}_{1:t})$ and the zero-tolerance recursion give

$$Q_{g,T+1} = O(T^{1-\alpha} + \sqrt{T\log(eT)}) = o(T).$$

Queue sublinearity is thus a consequence rather than a premise. Under the i.i.d. design alternative of Theorem 2, the strong law gives an eventual positive eigenvalue bound almost surely. Bounded responses then bound the greedy prices on each such path, so the pathwise zero-tolerance convergence also holds with eventual group coverage. The deterministic constants in (91) use the pathwise design condition stated in the theorem.

E.8 Why a current-gap penalty need not eliminate disparity

The logarithmic bound above uses accumulated gaps rather than only their current value. The distinction already matters in a two-group instance with known demand, so it is separate from estimation error.

Proposition 11 (A distinction from a current-gap penalty)**.** *Suppose two groups have distinct constant oracle prices v_A, v_B in the interior of the price interval, with a common known curvature $c = -b > 0$. Each group has infinitely many arrivals and the shock magnitude tends to zero. A correction with coefficient $k[|\bar{p}_{g_t,t-1} - \bar{p}_{t-1}| - \tau]^+$, for fixed finite $k \ge 0$ and $\tau \ge 0$, cannot have terminal group-price disparity converge to zero. In contrast, FRSD with $\lambda_f > 0$ and $\tau = 0$ has vanishing disparity when group coverage holds.*

Proof. Suppose the current-gap rule had $|\bar{p}_{A,t} - \bar{p}_{B,t}| \to 0$. Each group-to-market gap is a group-share multiple of this difference, so its correction coefficient q_t tends to zero. The unprojected center is

$$z_t = \frac{cv_{g_t} + q_t \bar{p}_{t-1}}{c + q_t}.$$

Bounded prices imply $z_t - v_{g_t} \to 0$. Since both oracle prices are interior and $\delta_t \to 0$, projection is eventually inactive and $p_t - v_{g_t} \to 0$. Averaging along the infinitely many arrivals of each group then gives $\bar{p}_{A,t} \to v_A$ and $\bar{p}_{B,t} \to v_B$, contradicting their distinct values. For FRSD, the greedy prices are bounded constants, so the pathwise proof (92) applies directly and gives vanishing disparity under coverage. □

The proposition concerns vanishing disparity, not uniform finite-horizon superiority. A current-gap penalty weakens to zero if its input vanishes, restoring distinct revenue-maximizing prices. An accumulated queue can retain a nonzero correction when the current gap is small. Positive-tolerance experiments examine the revenue and learning costs of this persistence at finite horizons.

F Structure of the Fair Pricing Benchmark

Section 3 defines the revenue cost of terminal fairness under known demand. We now characterize that benchmark for a fixed sequence of contexts and group arrivals. Every group considered below has $N_g(T) > 0$; write $N_g = N_g(T)$, let $o_t = p_t^\star \in I$, and put $c = -b > 0$. Contexts may vary within a group, so its oracle prices need not be constant.

F.1 Fair pricing as a projection problem

Let A be the group-averaging matrix: $A_{g,t} = 1/N_g$ when $g_t = g$ and zero otherwise. Then $(A\mathbf{p})_g$ is group g's average price. Define the feasible set

$$C_\tau = \{\mathbf{p} \in I^T : (A\mathbf{p})_g - (A\mathbf{p})_h \le \tau \text{ for every } g, h\}. \tag{93}$$

It is a nonempty compact convex set: the restrictions are linear and any constant feasible price is in C_τ for $\tau \ge 0$.

Proposition 12 (Geometry of the fair benchmark)**.** *For each $\tau \ge 0$, the benchmark has a unique price path and*

$$\mathbf{p}^F(\tau) = \arg\min_{\mathbf{p} \in C_\tau} \|\mathbf{p} - \mathbf{o}\|_2^2, \qquad \mathrm{PoF}_T(\tau) = c\|\mathbf{p}^F(\tau) - \mathbf{o}\|_2^2. \tag{94}$$

For every $\mathbf{p} \in C_\tau$,

$$\sum_t \{r_t(o_t) - r_t(p_t)\} \ge \mathrm{PoF}_T(\tau) + c\|\mathbf{p} - \mathbf{p}^F(\tau)\|_2^2. \tag{95}$$

Proof. The quadratic revenue identity gives $r_t(o_t) - r_t(p_t) = c(p_t - o_t)^2$. Summing proves the optimization equivalence and the value formula. Existence follows from compactness and continuity; strict convexity of squared distance gives uniqueness. Write $\mathbf{f} = \mathbf{p}^F(\tau)$. For any feasible $\mathbf{p}$, the segment $\mathbf{f} + s(\mathbf{p} - \mathbf{f})$ is feasible for $s \in [0, 1]$. Since its squared distance to $\mathbf{o}$ is minimized at $s = 0$, the right derivative there is nonnegative:

$$(\mathbf{f} - \mathbf{o})^\top(\mathbf{p} - \mathbf{f}) \ge 0. \tag{96}$$

Expanding the square yields

$$\|\mathbf{p} - \mathbf{o}\|_2^2 = \|\mathbf{f} - \mathbf{o}\|_2^2 + \|\mathbf{p} - \mathbf{f}\|_2^2 + 2(\mathbf{f} - \mathbf{o})^\top(\mathbf{p} - \mathbf{f}).$$

Discarding the final nonnegative term and multiplying by c proves (95). □

Equation (95) separates the unavoidable revenue cost of fairness from the additional loss associated with distance to the best fair price path. Both are measured through price deviations, as in the regret decomposition.

For fixed C_τ, the optimal path is stable with respect to changes in the oracle vector. Let $\mathbf{f}$ and $\mathbf{f}'$ denote the projections of $\mathbf{o}$ and $\mathbf{o}'$ onto C_τ. Applying (96) twice and adding gives

$$\|\mathbf{f} - \mathbf{f}'\|_2^2 \le (\mathbf{o} - \mathbf{o}')^\top(\mathbf{f} - \mathbf{f}') \le \|\mathbf{o} - \mathbf{o}'\|_2\|\mathbf{f} - \mathbf{f}'\|_2.$$

Consequently,

$$\|\mathbf{f} - \mathbf{f}'\|_2 \le \|\mathbf{o} - \mathbf{o}'\|_2. \tag{97}$$

The distance to the set is also 1-Lipschitz: choosing $\mathbf{f}'$ as a feasible candidate for $\mathbf{o}$ and reversing their roles proves

$$|\operatorname{dist}(\mathbf{o}, C_\tau) - \operatorname{dist}(\mathbf{o}', C_\tau)| \le \|\mathbf{o} - \mathbf{o}'\|_2. \tag{98}$$

With the same curvature c, this controls the square root of the fairness cost divided by c.

F.2 Separating group means from within-group variation

Let $\nu_g = N_g^{-1} \sum_{t:g_t=g} o_t$ denote the oracle group mean, and let z_g be a proposed group-average price. For any price path with those means,

$$\sum_t (p_t - o_t)^2 = \sum_g N_g (z_g - \nu_g)^2 + \sum_g \sum_{t:g_t=g} \{(p_t - z_g) - (o_t - \nu_g)\}^2. \tag{99}$$

To verify this, write $p_t - o_t = (z_g - \nu_g) + [(p_t - z_g) - (o_t - \nu_g)]$ within each group. The residual in brackets sums to zero, so the cross term vanishes.

The first term is the cost of moving group means. The second is the additional cost of changing within-group price variation. Without active price bounds, it is minimized to zero by a common shift within each group:

$$p_t = o_t + z_{g_t} - \nu_{g_t}. \tag{100}$$

Thus terminal group fairness can preserve all within-group oracle differences while changing only the group levels.

Proposition 13 (Reduction to group-average prices)**.** *Consider the weighted problem*

$$\min_{\mathbf{z}:\max_g z_g - \min_g z_g \le \tau} \sum_g N_g (z_g - \nu_g)^2. \tag{101}$$

It has a unique solution $\mathbf{z}^\star$*. If the shifted prices in* (100)*, using* $\mathbf{z}^\star$*, belong to* I*, they are the fair benchmark and*

$$\mathrm{PoF}_T(\tau) = c \sum_g N_g (z_g^\star - \nu_g)^2. \tag{102}$$

For arbitrary price bounds, the minimum in (101) *remains a lower bound on* $\mathrm{PoF}_T(\tau)/c$*.*

Proof. The feasible set in (101) is closed, convex, and nonempty. Since all $N_g > 0$, the objective is strictly convex and coercive, giving a unique minimum. Every feasible full price path has feasible group means, and (99) bounds its cost below by the weighted group-mean objective. The shifted path attains this lower bound whenever it is price-feasible, proving both assertions. □

This reduction applies to the contextual model: only the oracle means enter the group-level problem, while the context-driven deviations $o_t - \nu_g$ are retained by the shifts. The next subsection gives the corresponding exact construction when some shifted prices reach the bounds.

F.3 An exact construction with price bounds

For a group g and target mean $z \in I$, define

$$\phi_g(z) = \min \left\{ \sum_{t:g_t=g} (p_t - o_t)^2 : p_t \in I,\ \frac{1}{N_g} \sum_{t:g_t=g} p_t = z \right\}. \tag{103}$$

The feasible set is nonempty because all prices can equal z, and strict convexity makes its minimizing vector unique.

Proposition 14 (Clipped shifts within each group)**.** *The minimizer in* (103) *has the form*

$$p_t(z) = \mathrm{Proj}(o_t + s_g, I), \qquad \frac{1}{N_g} \sum_{t:g_t=g} \mathrm{Proj}(o_t + s_g, I) = z, \tag{104}$$

for a scalar s_g*. Such a scalar exists for every* $z \in I$*. Although it need not be unique at an endpoint or a clipping plateau, the resulting prices are unique. The full benchmark is obtained by minimizing*

$$\sum_g \phi_g(z_g) \quad \textit{over } z_g \in I, \quad \max_g z_g - \min_g z_g \le \tau, \tag{105}$$

and then applying (104) *within each group.*

Proof. The average of the clipped shifts is continuous and nondecreasing in s_g. It equals $\underline{p}$ for all sufficiently negative s_g and $\overline{p}$ for all sufficiently positive s_g. The intermediate value theorem supplies a shift for each target mean. For a fixed such shift, each $p_t(z)$ minimizes $(p - o_t)^2 - 2s_g p$ over I, since this expression differs from $(p - o_t - s_g)^2$ by a constant. For any other vector $\mathbf{q}$ with the same group mean, summing these coordinatewise inequalities gives

$$\sum_{t:g_t=g} \{(p_t(z) - o_t)^2 - 2s_g p_t(z)\} \le \sum_{t:g_t=g} \{(q_t - o_t)^2 - 2s_g q_t\}.$$

The linear sums cancel because the means agree, proving optimality. Strict convexity gives uniqueness of the prices. Finally, every full price path induces a feasible vector of group means and incurs at least $\sum_g \phi_g(z_g)$. The groupwise minimizers attain that value for any feasible mean vector, proving the exact reduction. □

The scalar equations can be solved by monotone root search. The outer problem is convex: if two target means have optimizing vectors, a convex combination of the vectors is feasible for the corresponding combined mean, and convexity of the squared objective bounds its cost by the same combination of costs. Hence every ϕ_g is convex.

The shift also provides a useful supporting slope. For a shift s_g representing z and any other mean z', the same coordinatewise minimization argument gives

$$\phi_g(z') \ge \phi_g(z) + 2N_g s_g(z' - z). \tag{106}$$

Thus $2N_g s_g$ is a subgradient relative to the mean interval. This gives a direct optimality check for computed benchmarks. Finally, clipping a common shift is monotone in o_t, so the benchmark preserves the ordering of oracle prices within each group.

F.4 A COMMON BAND FOR THE GROUP MEANS

The unconstrained group-mean problem (101) admits a one-dimensional representation. A vector has range at most τ exactly when its entries lie in some interval $[h, h + \tau]$. For fixed h, the minimizing means are $z_g(h) = \mathrm{Proj}(\nu_g, [h, h + \tau])$. Therefore the outer objective is

$$F(h) = \sum_g N_g\{(h - \nu_g)_+^2 + (\nu_g - h - \tau)_+^2\}. \tag{107}$$

It is continuously differentiable and convex, with

$$F'(h) = 2\sum_{\nu_g < h} N_g(h - \nu_g) - 2\sum_{\nu_g > h+\tau} N_g(\nu_g - h - \tau). \tag{108}$$

At an optimum, the total weighted upward adjustment of the low-mean groups equals the total weighted downward adjustment of the high-mean groups. Groups already inside the band keep their oracle means.

If $\max_g \nu_g - \min_g \nu_g \le \tau$, any band containing all means attains zero and $\mathbf{z}^\star = \boldsymbol{\nu}$. Otherwise a root of (108) lies between $\min_g \nu_g$ and $\max_g \nu_g - \tau$: the derivative at the first point is nonpositive and at the second nonnegative. It can be located by monotone bisection or by identifying the active lower and upper groups. If these groups are L and U, respectively, the balance equation gives

$$h = \frac{\sum_{g \in L \cup U} N_g \nu_g - \tau \sum_{g \in U} N_g}{\sum_{g \in L \cup U} N_g}, \tag{109}$$

with membership checked against $\nu_g < h$ and $\nu_g > h + \tau$. At a band endpoint the corresponding adjustment is zero, so either adjacent active-set convention gives the same means.

For exact equality of group means, $\tau = 0$, all groups are assigned the common weighted mean

$$z_g^\star = \frac{1}{T}\sum_h N_h \nu_h = \frac{1}{T}\sum_t o_t. \tag{110}$$

When the associated shifts stay inside I, equation (102) becomes the weighted between-group variance of the oracle means multiplied by c. The within-group oracle variation is still preserved. This identifies which part of oracle price variation creates the cost of group-average fairness.

F.5 Two groups and an explicit fairness cost

For two groups, suppose $\nu_A \geq \nu_B$, and let $d = (\nu_A - \nu_B - \tau)^+$. The solution of (101) is

$$z_A^\star = \nu_A - \frac{N_B}{T} d, \qquad z_B^\star = \nu_B + \frac{N_A}{T} d. \tag{111}$$

To see this when $d > 0$, write $x = z_A - \nu_A$ and $y = z_B - \nu_B$. The binding gap constraint is $y - x = d$. Substitution into $N_A x^2 + N_B y^2$ gives a strictly convex quadratic in x, whose derivative is $2Tx + 2N_B d$. Its zero yields the two shifts. When $d = 0$, the oracle means are already feasible and both shifts vanish.

The resulting mean-level cost is

$$N_A(z_A^\star - \nu_A)^2 + N_B(z_B^\star - \nu_B)^2 = \frac{N_A N_B}{T} d^2. \tag{112}$$

Whenever the shifted oracle prices remain feasible,

$$\mathrm{PoF}_T(\tau) = (-b)\frac{N_A N_B}{T}(\nu_A - \nu_B - \tau)_+^2. \tag{113}$$

This extends the constant-within-group construction in Appendix A.4 to contextual oracle prices with feasible common shifts. For the constant-oracle class itself, feasibility follows automatically because the adjusted prices lie between the two oracle prices.

The group sizes determine how the adjustment is shared: the larger group moves less in mean price, since moving each of its prices incurs a larger total cost. Writing $\pi = N_A/T$, the cost equals $(-b)T\pi(1-\pi)d^2$. This displays the dependence on horizon, group balance, and the excess oracle gap separately.

F.6 Verifiable bounds for a computed fair benchmark

The benchmark can also be checked through a primal price path and nonnegative multipliers. Let B contain one row $A_g - A_h$ for every ordered pair $g \neq h$, so feasibility is $B\mathbf{p} \leq \tau\mathbf{1}$ together with $\mathbf{p} \in I^T$. For any vector $\lambda \geq 0$, write $d = B^\top\lambda$ and define

$$p_t(\lambda) = \mathrm{Proj}\left(o_t - \frac{d_t}{2c}, I\right), \quad L(\lambda;\tau) = \sum_t \{c(p_t(\lambda) - o_t)^2 + d_t p_t(\lambda)\} - \tau\mathbf{1}^\top\lambda. \tag{114}$$

These expressions require only scalar clipping. They give a lower bound on the fairness cost without first solving the full constrained problem.

Proposition 15 (Certificate for the fair benchmark)**.** *For every $\lambda \geq 0$ and feasible price path $\mathbf{q} \in C_\tau$,*

$$L(\lambda;\tau) \leq \mathrm{PoF}_T(\tau) \leq c\|\mathbf{q} - \mathbf{o}\|_2^2. \tag{115}$$

If $\mathbf{p}(\lambda)$ is feasible and

$$\lambda^\top(\tau\mathbf{1} - B\mathbf{p}(\lambda)) = 0, \tag{116}$$

then it is the unique benchmark and both bounds coincide. For an arbitrary feasible $\mathbf{q}$, their gap is at least $c\|\mathbf{q} - \mathbf{p}(\lambda)\|_2^2$.

Proof. For any feasible $\mathbf{q}$,

$$c\|\mathbf{q} - \mathbf{o}\|_2^2 \geq c\|\mathbf{q} - \mathbf{o}\|_2^2 + \lambda^\top(B\mathbf{q} - \tau\mathbf{1}).$$

The expression on the right separates across coordinates after writing $d = B^\top\lambda$. Its minimum over I^T is attained by (114), by completing each square. Hence it is at least $L(\lambda;\tau)$, proving the lower bound; using $\mathbf{q}$ as a feasible candidate proves the upper bound. More explicitly, with $\mathbf{p} = \mathbf{p}(\lambda)$,

$$c\|\mathbf{q} - \mathbf{o}\|_2^2 - L(\lambda;\tau) = \sum_t \{c(q_t - o_t)^2 + d_t q_t - c(p_t - o_t)^2 - d_t p_t\} \tag{117}$$

$$+ \lambda^\top(\tau\mathbf{1} - B\mathbf{q}). \tag{118}$$

The final term is nonnegative. Each coordinate objective is a quadratic with curvature $2c$ minimized at p_t over I. The projection variational inequality gives its increase at q_t at least $c(q_t - p_t)^2$. Summing proves the gap bound. If $\mathbf{p}$ is itself feasible and (116) holds, substituting $\mathbf{q} = \mathbf{p}$ makes both terms in (118) zero. Uniqueness follows from Proposition 12. □

For the two-group constant-oracle case with excess gap $d_0 = \nu_A - \nu_B - \tau > 0$, put all multiplier mass on the constraint $z_A - z_B \le \tau$ and choose

$$\lambda_{A,B} = 2c\frac{N_A N_B}{T} d_0.$$

Its coordinates d_t are $\lambda_{A,B}/N_A$ in group A and $-\lambda_{A,B}/N_B$ in group B. The prices in (114) are therefore exactly the shifts in (111). They are feasible, their group gap equals τ, and complementarity holds. This provides an independent optimality certificate for the closed-form fairness cost.

If a multiplier vector certifies the optimum at tolerance τ, it also gives a supporting bound at every other tolerance $\tau' \ge 0$:

$$\mathrm{PoF}_T(\tau') \ge \mathrm{PoF}_T(\tau) - (\tau' - \tau)\mathbf{1}^\top \lambda. \tag{119}$$

Indeed, $L(\lambda; \tau') = L(\lambda; \tau) - (\tau' - \tau)\mathbf{1}^\top \lambda$, and the lower bound applies at τ'. Thus the same certificate quantifies how relaxation of the fairness requirement can change the optimal revenue cost.

F.7 How the fairness cost changes with tolerance

Proposition 16 (Monotonicity and convexity of the price of fairness)**.** *For a fixed context and group sequence, $\tau \mapsto \mathrm{PoF}_T(\tau)$ is nonnegative, nonincreasing, and convex on $[0, \infty)$. It is zero whenever the oracle path is terminally τ-price-fair. In addition,*

$$0 \le \mathrm{PoF}_T(\tau) \le c\sum_t (o_t - \bar{o})^2, \qquad \bar{o} = T^{-1}\sum_t o_t. \tag{120}$$

Proof. If $\tau \le \tau'$, then $C_\tau \subseteq C_{\tau'}$, so the minimum squared distance is nonincreasing. It is nonnegative and equals zero when $\mathbf{o}$ is feasible. For convexity, choose tolerances τ_1, τ_2 and their optimizers $\mathbf{f}_1, \mathbf{f}_2$. For $a \in [0, 1]$, the vector $a\mathbf{f}_1 + (1-a)\mathbf{f}_2$ stays in I^T. Every ordered group difference is at most $a\tau_1 + (1-a)\tau_2$, so this vector is feasible at the combined tolerance. Convexity of squared distance gives

$$\begin{aligned}\mathrm{PoF}_T(a\tau_1 + (1-a)\tau_2) &\le c\|a(\mathbf{f}_1 - \mathbf{o}) + (1-a)(\mathbf{f}_2 - \mathbf{o})\|_2^2 \\ &\le a\,\mathrm{PoF}_T(\tau_1) + (1-a)\,\mathrm{PoF}_T(\tau_2).\end{aligned}$$

Finally, the constant path $p_t = \bar{o}$ belongs to I^T and is fair for every $\tau \ge 0$. Its cost proves (120). □

On the two-group constant-oracle class, the marginal cost is explicit:

$$\frac{d}{d\tau}\mathrm{PoF}_T(\tau) = -2(-b)\frac{N_A N_B}{T}(\nu_A - \nu_B - \tau)^+. \tag{121}$$

The derivative is also zero at the point where the tolerance reaches the oracle gap. The largest benefit from relaxing the tolerance occurs when the remaining excess gap is large, and the benefit declines continuously to zero as the oracle path becomes feasible. This is the benchmark trade-off underlying the fairness-cost term in Theorem 4.

G Information and the Learning Lower Bound

The learning construction in Appendix A.4 relates estimation accuracy to price deviations that incur regret. This section makes the adaptive likelihood calculation and the constants explicit. It uses the same Gaussian learning family and the same fixed policy–estimator pair as Theorem 4; the bounded-noise concentration results above concern the upper-bound model instead.

G.1 A fixed neighborhood of demand instances

Recall $b(h) = b_0 + h$ and $\eta_g(h) = -(2b_0 + h)p_0$. The oracle is

$$p^\star(h) = p_0 - \frac{hp_0}{2(b_0 + h)}.$$

The family can be chosen on an interval independent of the horizon and policy. Let $c_0 = -b_0 > 0$ and let $d_p = \min\{p_0 - \underline{p}, \overline{p} - p_0\} > 0$. Choose $h_0 > 0$ such that

$$h_0 \le c_0/2, \qquad h_0 < \overline{b} - b_0, \qquad h_0|p_0|/c_0 < d_p. \tag{122}$$

For $0 \le h \le h_0$, the slope remains in its negative interval and

$$|p^\star(h) - p_0| = \frac{h|p_0|}{2|b_0 + h|} \le \frac{h|p_0|}{c_0} < d_p.$$

Thus the oracle stays interior. The coefficients $\eta_g(h)$ are bounded on this fixed interval, as required by the class containing the construction. Because $p_0 \neq 0$ and $h_0 > 0$, the endpoint oracle separation $d_0 = |p^\star(h_0) - p_0|$ is strictly positive.

The mean demand has the useful form

$$m_h(p) = m_0(p) + h(p - p_0). \tag{123}$$

At $p = p_0$, all instances therefore produce the same conditional demand distribution. Moving away from p_0 creates information, and the squared size of that movement is exactly the revenue loss under instance zero, up to the curvature $-b_0$.

G.2 Likelihood ratios for adaptive prices

An adaptive policy uses a probability kernel $K_t(dp_t \mid H_{t-1})$ given the observed history, where any parameter-independent randomization can be included in H_{t-1} or sampled through the kernel. The kernel is the same function under every instance. The conditional demand density given the history and current price is

$$f_h(y \mid p_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - m_h(p_t))^2}{2\sigma^2}\right].$$

The joint law is built by alternating the common action kernels and these demand kernels. In the likelihood ratio between instances zero and h, the common action kernels cancel. Equivalently, applying the chain rule one arrival at a time yields

$$\log \frac{d\mathbb{P}_0}{d\mathbb{P}_h} = \sum_{t=1}^{T} \log \frac{f_0(D_t \mid p_t)}{f_h(D_t \mid p_t)}. \tag{124}$$

Gaussian densities are positive everywhere, so these conditional likelihood ratios are well defined. A deterministic policy is included as a point-mass action kernel.

Under instance zero, write $D_t = m_0(p_t) + \varepsilon_t$. With $d_t = m_h(p_t) - m_0(p_t)$, the conditional log-density ratio is

$$\frac{(D_t - m_h(p_t))^2 - (D_t - m_0(p_t))^2}{2\sigma^2} = \frac{d_t^2 - 2d_t\varepsilon_t}{2\sigma^2}.$$

Its conditional expectation is $d_t^2/(2\sigma^2)$ because the next Gaussian noise has mean zero given the history and price. Taking expectations in (124) and using (123) gives

$$\mathrm{KL}(\mathbb{P}_0 \| \mathbb{P}_h) = \frac{h^2}{2\sigma^2}\mathbb{E}_0 \sum_{t=1}^{T} (p_t - p_0)^2 = \frac{h^2 \bar{R}_0(T)}{2\sigma^2(-b_0)}. \tag{125}$$

The expectation is under instance zero throughout. Adaptivity changes the distribution of the observed prices, but it does not add a likelihood term for a policy kernel shared by the two instances.

G.3 From indistinguishability to estimation error

Let $\widehat{b}$ be the final estimator, including any parameter-independent randomization. Define the event $A = \{\widehat{b} \ge b_0 + h/2\}$. Under instance zero, A entails squared error at least $h^2/4$; under instance h, its complement entails the same lower bound. Consequently,

$$\mathbb{E}_0[(\widehat{b} - b_0)^2] + \mathbb{E}_h[(\widehat{b} - b_0 - h)^2] \ge \frac{h^2}{4}\{\mathbb{P}_0(A) + \mathbb{P}_h(A^c)\}. \tag{126}$$

By the definition of total variation,

$$\mathbb{P}_0(A) + \mathbb{P}_h(A^c) = 1 + \mathbb{P}_0(A) - \mathbb{P}_h(A) \geq 1 - \|\mathbb{P}_0 - \mathbb{P}_h\|_{\mathrm{TV}}.$$

The supremum risk L_T is at least the average of these two risks. Thus

$$L_T \geq \frac{h^2}{8}\left(1 - \|\mathbb{P}_0 - \mathbb{P}_h\|_{\mathrm{TV}}\right). \tag{127}$$

This step is valid for any final estimator, not just the shock-product estimator used by FRSD.

For completeness, the total-variation bound used here can be reduced to a binary calculation. For any event B, coarse-graining the likelihood ratio onto B and B^c, using convexity of $x \log x$ on each set, gives

$$\mathrm{KL}(P\|Q) \geq p\log(p/q) + (1-p)\log((1-p)/(1-q)), \quad p = P(B),\ q = Q(B).$$

For fixed $q \in (0,1)$, the second derivative of the right-hand side as a function of p is $1/[p(1-p)] \geq 4$. Its value and first derivative vanish at $p = q$, so it is at least $2(p-q)^2$. Endpoint cases follow by limits. Taking the supremum over B gives

$$\|P - Q\|_{\mathrm{TV}} \leq \sqrt{\mathrm{KL}(P\|Q)/2}. \tag{128}$$

In particular, a relative-entropy budget of $1/8$ leaves a total-variation distance of at most $1/4$, which makes the right-hand side of (127) at least $3h^2/32$.

G.4 Constants independent of the policy and horizon

The endpoint instances give a strictly positive regret floor before any observations. Put $c_b = -(b_0 + h_0)$ and $d_0 = |p^\star(h_0) - p_0|$. For any first price p,

$$(p - p_0)^2 + (p - p^\star(h_0))^2 \geq d_0^2/2.$$

The first-price distribution is the same under both instances. Since their revenue curvatures are at least c_b, and all subsequent oracle losses are nonnegative,

$$R_T \geq \tfrac{1}{2}\{\bar{R}_0(T) + \bar{R}_{h_0}(T)\} \geq c_b d_0^2/4 =: r_0 > 0. \tag{129}$$

Choose

$$\kappa = \min\{r_0 h_0^2, \sigma^2(-b_0)/4\}, \qquad h = \sqrt{\kappa/R_T}. \tag{130}$$

The fixed family contains this instance because $R_T \geq r_0$ implies $h \leq \sqrt{\kappa/r_0} \leq h_0$. Equation (125) gives

$$\mathrm{KL}(\mathbb{P}_0\|\mathbb{P}_h) = \frac{\kappa \bar{R}_0(T)}{2\sigma^2(-b_0)R_T} \leq \frac{\kappa}{2\sigma^2(-b_0)} \leq \frac{1}{8}. \tag{131}$$

The testing inequality therefore yields

$$L_T \geq \frac{3\kappa}{32R_T}, \qquad R_T \geq \frac{c_L}{L_T}, \qquad c_L = 3\kappa/32. \tag{132}$$

All constants are fixed by the demand family and price interval. The choice of the test point may depend on the policy's worst-case regret, but the family and c_L do not. For finite L_T the displayed testing bound also shows that L_T is positive; if $L_T = \infty$, the reciprocal bound is read as zero.

G.5 The fairness term and its attainable cost

The other subclass in Theorem 4 has two separated constant oracle prices. Equation (113) specializes to its exact known-demand benchmark,

$$\mathrm{PoF}_T(\tau) = (-b_0)\frac{N_A N_B}{T}(\Delta^\star - \tau)^2, \qquad \Delta^\star > \tau.$$

Every terminally fair price path has at least this oracle loss. The shifts in (111) attain the bound and remain feasible, so the cost does not depend on a loose inequality in the construction.

The coverage condition can sharpen its coefficient. Since $N_A + N_B = T$ and both counts are at least $\pi_{\min} T$, one necessarily has $\pi_{\min} \leq 1/2$. With $x = N_A/T \in [\pi_{\min}, 1 - \pi_{\min}]$, concavity and symmetry of $x(1-x)$ give

$$\frac{N_A N_B}{T} = Tx(1-x) \geq T\pi_{\min}(1-\pi_{\min}) \geq T\pi_{\min}^2. \tag{133}$$

The main theorem uses the last, simpler coefficient. The exact finite-sample factor above preserves the dependence on observed group counts.

Finally, the supremum over a class containing both constructions is at least the supremum over each subclass. Combining the information cost (132) with the attained fairness cost yields the maximum in Theorem 4. The constructions connect two distinct reasons for departing from an oracle price: acquiring demand information and meeting a group-average price requirement.