arXiv:2603.24974v3 [math.OC] 19 Nov 2026

The Value of Information in Resource-Constrained Pricing

Ruicheng Ao† Jiashuo Jiang‡ David Simchi-Levi†,§

† Institute for Data, Systems, and Society, Massachusetts Institute of Technology
§ Department of Civil and Environmental Engineering and Operations Research Center, MIT
aorc@mit.edu, dslevi@mit.edu

‡ Department of Industrial Engineering and Decision Analytics,
Hong Kong University of Science and Technology
jsjiang@ust.hk

Abstract

Dynamic pricing with resource constraints requires balancing exploration of unknown demand patterns against exploitation of known information to maximize revenue. We develop algorithms for this problem across different information regimes. Products share perishable resources, and the firm posts prices and observes the realized requests, which may exceed the remaining capacity. In the full-information setting, a residual-correction policy that solves the fluid problem once attains $O(\log T)$ regret with bounded prices and arbitrary-rank resource matrices. In the no-information setting, deterministic exploration followed by a certainty-equivalent policy attains $O(\sqrt{T})$ regret when the limiting fluid solution is nondegenerate. In the informed-price setting, given an initial price-demand estimate, a feedback controller learns the implementing prices without estimating the slope matrix if the slope matrix is symmetric, a set containing it is known in advance, the informed price is aligned with the resource constraints, and an inverse-image condition holds. A selector chooses between the informed-price algorithm and the no-information algorithm from historical data. Its regret bound includes observation error, misalignment, inventory exits, and periods on the fixed price. We further show how offline surrogate models reduce the regret of the learning policy through variance reduction, even when misspecified. The bound includes calibration and screen failures, and the regret is reduced for all sufficiently large horizons. We finally demonstrate the empirical performances of our policies through complete numerical experiments.

Keywords: Dynamic pricing, resource constraints, revenue management, fluid approximation, value of information.

1 Introduction

Dynamic pricing of perishable resources has been studied in revenue management for more than three decades (Gallego and Van Ryzin 1994, Talluri and van Ryzin 2004). The firm must balance immediate revenue against long-term capacity. Firms typically start with prior knowledge of demand. Examples include an incumbent price-demand pair inferred from historical operations, a forecasting model trained offline, or a surrogate demand signal derived from customer features, context, or related markets. The firm must learn demand and use imperfect offline information at the same time.

For example, a seasonal retailer prices several products that share inventory and fulfillment capacity. It posts prices each period and then observes the realized requests, which may exceed the remaining capacity. Before the season it may have historical price–demand pairs, frozen forecasts, traffic indices, or scores from related categories. Errors in these inputs affect both revenue and resource use.

This problem is more difficult under hard resource constraints than in unconstrained pricing. Pricing errors can reduce one-period revenue, and they can also consume inventory or capacity needed later. Each decision changes the feasible set for all future periods. A pricing policy must manage the classical learning-earning tradeoff (Keskin and Zeevi 2014) and preserve feasibility under uncertainty.

A recurring obstacle in this constrained setting is *degeneracy*. When resources approach depletion or multiple products compete for the same capacity, the fluid optimizer is still unique under strict concavity, but its binding constraints can be redundant and its dual prices non-unique; the optimal basis then ceases to be isolated (Bumpensanti and Wang 2020, Jiang et al. 2025). A re-solving analysis can require the binding set to stay fixed as the remaining capacity changes; under degeneracy it need not, which is why a re-solving policy can oscillate. With imperfect forecasts or misspecified model outputs, the resulting pricing errors can trigger premature resource depletion.

This paper asks: *how should a resource-constrained pricing system use imperfect offline information while preserving worst-case robustness?* We separate the control problem from the information problem. We first introduce a *residual-correction heuristic* that solves the fluid problem once and then corrects the demand targets by the cumulative residuals. Because the fluid problem is solved only once, the policy retains a single fluid solution throughout. This heuristic underlies all subsequent algorithms. We then study how different forms of offline information interact with this heuristic.

We distinguish two kinds of offline information. The first is an *informed price*, represented by an anchor $(\boldsymbol{p}^0, \boldsymbol{d}^0)$: $\boldsymbol{p}^0$ is the price vector and $\boldsymbol{d}^0$ the approximate demand at that price. The firm uses the anchor to construct and assess a resource-feasible demand plan. The second is a *surrogate model* whose scores may be biased but can be correlated with the noise in the requests at a fixed price. We use its centered scores to form pseudo-observations for the exploration regression, which reduces estimation variance. An anchor can change the posted prices and therefore the regret rate, whereas a correlated surrogate can only reduce the variance of learning. We study the two separately. We ask when offline information should change the prices, when it should change only the estimator, and when it should not be used.

Throughout, demand is affine and offline information may be misspecified. Every policy posts bounded prices, serves requests by a fixed rule based on remaining capacity, and uses the same initial-fluid benchmark for comparison. This benchmark replaces random demand with its expectation and chooses one price vector subject to the initial capacity constraints; regret is this benchmark less the policy's expected revenue. Our main contributions are as follows.

1. **Control under degeneracy (Section 3).** Under the local fluid-price margin of Assumption 5, residual correction attains $O(\log T)$ regret with bounded prices and resource matrices of arbitrary rank. The bound applies regardless of active-set stability or dual-price positivity, including when dual prices are non-unique.

2. **Online learning (Section 4).** Without prior information, deterministic exploration followed by a certainty-equivalent policy attains $O(\sqrt{T})$ regret when the limiting fluid solution is nondegenerate, and this rate is minimax optimal. The constants in these bounds depend on fixed problem primitives, including the noise proxy and local margin, but not on T; the no-information constant also depends on the fixed exploration design and its tuning constants.

3. **Informed prices in a network (Section 5).** Our main result is a condition under which an exact informed price is sufficient for pricing in a network. Requests are bounded and the supplied parameter class contains the truth, with symmetric slopes satisfying known uniform spectral bounds. Under the inverse-image condition that can be checked before the season, the informed-price algorithm finds the corresponding prices without estimating the slope matrix (Theorem 3). This inverse-image condition ensures that targets near the plan have implementing prices inside the price set for every model in the supplied parameter class; it is distinct from the nondegeneracy condition, which keeps the binding set

stable for the online-learning result. The regret of the algorithm grows with the residual and slack of the plan, and both vanish when the informed price is exact and aligned with the resources. With zero slack, an $O(T^{-1/4})$ threshold on the residual bound marks the crossover between the informed and no-information regret bounds, and a selector chooses between the two algorithms using only historical data (Theorem 4). Hence the firm uses the informed price only when it can check, before the season, that the price is accurate enough.

4. **Variance reduction from misspecified surrogates (Section 6).** We use biased but correlated surrogate models as variance-reducing instruments rather than direct recommendations. Through control variates, surrogate information reduces learning variance and improves regret constants even under misspecification, because the paired signals are correlated with the noise in the requests. Our regret bound for this case includes the costs of calibration, centering, overlap and failure. Under its calibration and regularity conditions, and when variance reduction is positive in directions that affect revenue, Proposition 1 establishes that the surrogate-assisted algorithm has smaller expected regret than its least-squares counterpart for all sufficiently large T. A separate profit identity includes data-acquisition and computation costs, so the firm can compare the gain from a signal with its price.

Across all settings, regret decomposes into stochastic noise, parameter uncertainty, and initialization bias: the first is controlled by residual correction, anchors shift the bias-versus-learning tradeoff, and surrogates reduce estimation variance. Figure 1 summarizes this decomposition.

For a manager, the results translate into two decisions made before the season. First, use historical data to compute a bound η on the fluid optimality residual for the chosen plan and resource multiplier. If the plan check passes, the slack is $O(T^{-1/2})$ and η is small relative to $T^{-1/4}$, run the informed-price algorithm. Otherwise, use the no-information algorithm (Theorem 4 makes this choice using observed data only). Second, buy an auxiliary signal only if its expected gain in revenue from sales exceeds its acquisition and computation cost (Section 6.4).

1.1 Related Literature

Our work connects four strands of literature.

Pricing with resource constraints. Dynamic pricing with perishable resources dates to the fluid benchmark of Gallego and Van Ryzin (1994); see Jasin (2014), Bumpensanti and Wang (2020), Wang and Wang (2022), Jiang et al. (2025), Ao et al. (2024a,b, 2025) and references therein. In price-based revenue management, the firm sets continuous prices, as in this paper. With known demand and a nondegenerate fluid solution, self-adjusting price control (Jasin 2014) attains logarithmic regret against the fluid value. Re-solving (Wang and Wang 2022) attains constant regret against the optimal dynamic policy, whose value can lie order $\log T$ below the fluid value. Chen et al. (2021) extend self-adjusting control to unknown demand. Our residual-correction heuristic belongs to this stream. Its logarithmic bound is measured against the fluid value, so it complements the constant bound of Wang and Wang (2022), and it applies regardless of active-set stability or dual-price positivity. Quantity-based network revenue management (Reiman and Wang 2008, Jasin and Kumar 2012, Bumpensanti and Wang 2020) and online linear programming (Li and Ye 2022) instead decide which discrete requests to serve. Bumpensanti and Wang (2020) obtain uniformly bounded loss without a nondegenerate fluid solution. Here the firm controls only prices, and every request is served by a fixed rule.

Offline data, priors, and partial demand knowledge. For a single product, prior structure can lower regret from $O(\sqrt{T})$ to $O(\log T)$: when every price is informative (Broder and Rusmevichientong 2012), when part of the demand model is known (Keskin and Zeevi 2014), or when enough offline data are available (Bu et al. 2020). Bu et al. (2020) characterize phase transitions and an inverse-square law for optimal regret as a function of the size, location, and dispersion of offline data, showing that even imperfect historical prices carry

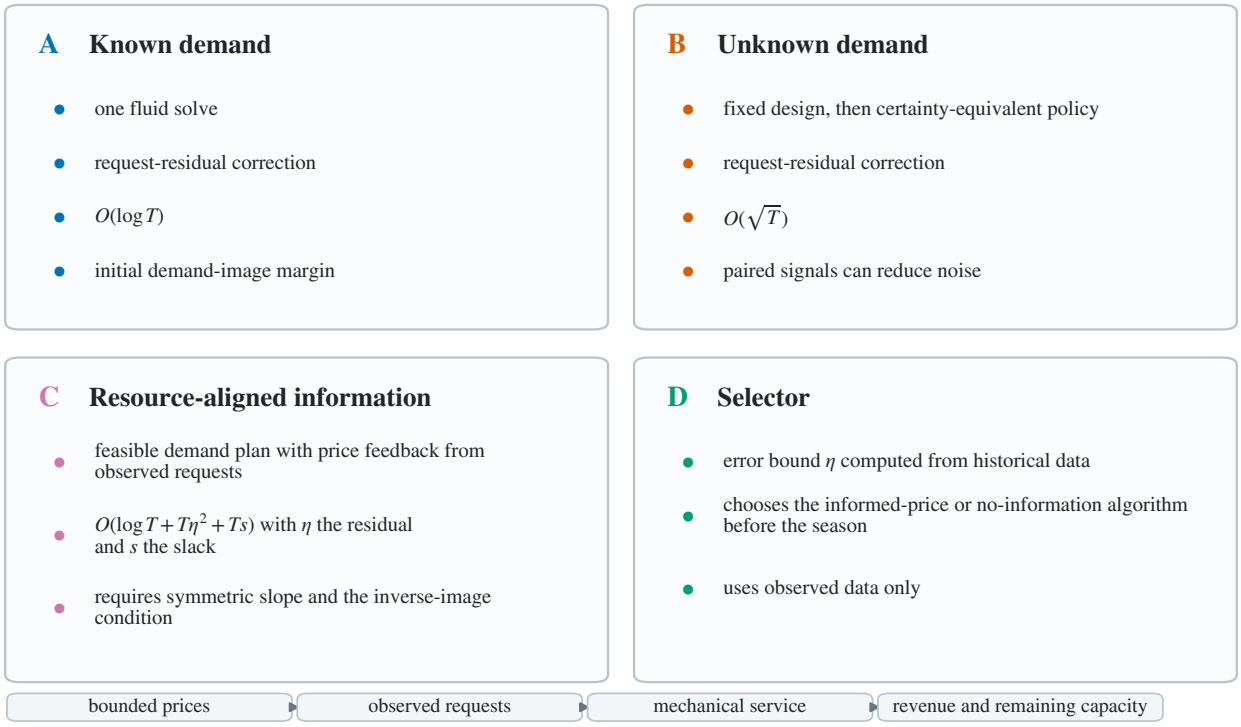


Figure 1: Information and pricing guarantees. With known demand, the firm can use residual correction. With no information, deterministic exploration followed by a certainty-equivalent policy gives the $O(\sqrt{T})$ baseline when the limiting fluid solution is nondegenerate. The informed-price algorithm attains $O(\log T)$ regret under an exactly aligned anchor and the inverse-image condition. Otherwise, its bound grows with the residual and slack. A selector chooses between the informed-price algorithm and the no-information algorithm using only historical data. Correlated offline surrogate models can reduce the variance of the exploration fit through control variates.

non-trivial information. Subsequent work extends these ideas to offline data integration, misspecification, and weaker forms of prior information (Ferreira et al. 2018, Wang et al. 2024, Li et al. 2021, Simchi-Levi and Xu 2022, Xu and Zeevi 2020). With shared resources, the information must also identify the response direction that moves the optimal product mix. Section 5 analyzes this direction.

Online learning under constraints and learning-augmented optimization. The bandits with knapsacks (BwK) literature addresses online decision-making under resource constraints, primarily for discrete actions and bandit feedback (Badanidiyuru et al. 2014, Agrawal and Devanur 2016, Agrawal et al. 2016, Liu et al. 2022, Ma et al. 2024, Sankararaman and Slivkins 2021, Sivakumar et al. 2022, Liu and Grigas 2022). A separate line on algorithms with predictions studies consistency and robustness when external advice may be imperfect (Lykouris and Vassilvitskii 2021, Mitzenmacher and Vassilvitskii 2022, Purohit et al. 2018, Wei and Zhang 2020, Bhaskara et al. 2020, 2021, 2023). Our informed-price result follows that consistency-robustness logic but operates in a different decision structure: continuous pricing with fluid optimization, irreversible resource consumption, and a selector that uses only historical data.

Auxiliary signals, prediction-powered methods, and AI-assisted operations. The surrogate framework builds on control variates and prediction-powered inference (Ao et al. 2026b,a) rather than direct action recommendation. We use prediction-powered and cross-prediction estimators (Angelopoulos et al. 2024, Zrnic and Candès 2024) in constrained pricing. The surrogate signal reduces estimation variance. This distinction matters under hard constraints: biased auxiliary models remain useful, whereas biased price recommendations can trigger premature resource depletion. Predictive systems, simulators, and LLM-based tools now provide decision support across operational settings—from optimization model formulation and solver interaction to inventory control and demand forecasting (Liang et al. 2026, Ao et al. 2026d,e, Duan et al. 2025, Baek et al. 2026b,a). We characterize when such imperfect signals should alter the posted prices (anchors) and when they should only reduce estimation variance (correlated auxiliaries).

Relation to the conference version. The conference version (Anonymous 2025) established three nominal rate regimes. This paper keeps its price set and resource constraints, and the firm observes the realized demand and serves requests mechanically. It adds the network results and the analysis of surrogate calibration and acquisition cost. The new results are Theorems 3 and 4, which give the condition under which the anchor is sufficient and the selector, and Proposition 3, which gives the minimax rate with an exact anchor. Appendix A relates each result to the conference version and states the scope of each rate and each numerical result. Our experiments separate gains from improved fitting and from additional information. Section 7.3 compares a revenue screen with 58 controls on 240 cases.

2 Model and Preliminaries

We now formalize the model, the fluid benchmark against which all policies are compared, and the assumptions used throughout. Over T periods, a firm sells n products that share m resources, with a known consumption matrix $A \in \mathbb{R}_+^{m\times n}$. In period t the firm observes the remaining capacity c^t, posts prices $p^t \in \mathcal{P} = [L, U]^n$ with $0 \le L < U < \infty$, and then observes the potential requests $d^t \in \mathbb{R}_+^n$. The firm controls only prices. A fixed service rule serves $\widehat{d}^t = \Phi(d^t, c^t)$ (Definition 1). The firm learns from requests d^t and earns revenue $(p^t)^\top \widehat{d}^t$. Capacity evolves as

$$c^{t+1} = c^t - A\widehat{d}^t, \qquad c^1 = C_T. \tag{1}$$

In seasonal retail, products are SKUs and resource rows represent inventory pools or shared fulfillment budgets. Unmet requests use no resources and earn no revenue (Talluri and van Ryzin 2004, Jasin and Kumar 2013, Chs. 2–3). Observing full requests follows the learning convention of Ferreira et al. (2018, §§2.1,4.1). Applications include airlines pricing perishable seats, retailers selling seasonal goods, and cloud providers allocating compute capacity—all settings where unsold inventory has no salvage value. We denote by $f(p)$

the vector of expected demand at price p; its affine specification is stated in the demand assumption below. We first describe the scaling of capacity with the horizon.

Assumption 1 (Fluid scaling). *(a) For fixed $\bar{c} \in \mathbb{R}^m_{++}$ and $K_C < \infty$, $\|C_T - T\bar{c}\|_\infty \le K_C$ at every horizon. Dimensions, prices, demand, A, noise proxy and fulfillment rule are fixed along the sequence.*

(b) $C_T \ge 0$ and $\{p \in \mathcal{P} : Af(p) \le C_T/T\} \ne \emptyset$ at every horizon.

We measure any policy by its *regret* relative to an idealized *fluid benchmark* that replaces random demand with its expectation and solves the resulting deterministic problem. The benchmark is the initial diagonal fluid problem: it posts one price and serves mean requests without rationing:

$$V^{\mathrm{Fluid}}(C_T) = T \max_{p \in \mathcal{P}} \{p^\top f(p) : Af(p) \le C_T/T\}. \tag{2}$$

Proposition 2 relates it to the time-varying untruncated diagonal relaxation of Gallego and Van Ryzin (1994, Theorem 2). Fluid regret is signed: it can be negative, because the benchmark serves mean requests without rationing. All comparisons use this same signed benchmark. Following the revenue management literature (Talluri and van Ryzin 2004, Bu et al. 2020), we impose two structural assumptions: affine demand, which captures price sensitivity with a finite-dimensional parameter space, and negative-definite slope matrices, which ensure that expected revenue is strictly concave in prices.

Assumption 2 (Affine demand). *(a) $f(p) = \alpha + Bp$ for $\alpha \in \mathbb{R}^n$, $B \in \mathbb{R}^{n \times n}$ and every $p \in \mathcal{P}$.*

(b) $\lambda_{\max}(B + B^\top) < 0$, and $f(p) \ge 0$ throughout $\mathcal{P}$.

Negative definiteness ensures strictly decreasing demand ($B_{jj} < 0$) and invertibility of B. Affine demand is standard in learning-based pricing (Keskin and Zeevi 2014, Javanmard and Nazerzadeh 2019, Bu et al. 2020). It gives affine inverse demand and concave quadratic revenue. In this setting, Bu et al. (2020) show that the interplay between offline data and online learning produces phase transitions and an inverse-square law. With no information, the analysis also requires the limiting fluid solution to be nondegenerate, while the known-demand result allows degenerate limiting fluid solutions.

Assumption 3 (Nondegenerate limiting fluid solution). *Let $\bar{c}$ be the limiting capacity in Assumption 1(a), $\mathcal{D} := \alpha + B\mathcal{P}$, $\ell(d) := d^\top B^{-1}(d - \alpha)$, and $\bar{x}^*$ the unique maximizer of ℓ over $\{d \in \mathcal{D} : Ad \le \bar{c}\}$, with price $\bar{p}^* := B^{-1}(\bar{x}^* - \alpha)$. For constants $\delta_P, \delta_D, a_{\min}, \lambda_{\min}, s_{\min} > 0$: $L + \delta_P \le \bar{p}^*_j \le U - \delta_P$ for every j; $\mathrm{dist}(\bar{x}^*, \partial\mathcal{D}) \ge \delta_D$; the binding rows $\bar{J} := \{i : A_i\bar{x}^* = \bar{c}_i\}$, if any, have full row rank with $\sigma_{\min}(A_{\bar{J}}) \ge a_{\min}$ and $\nabla\ell(\bar{x}^*) = A_{\bar{J}}^\top \bar{\lambda}_{\bar{J}}$ with $\bar{\lambda}_{\bar{J}} \ge \lambda_{\min}\mathbf{1}$ ($\bar{\lambda} = 0$ when $\bar{J} = \emptyset$); and every other row has slack $\bar{c}_i - A_i\bar{x}^* \ge s_{\min}$. For each horizon, d^*_T denotes the unique maximizer of ℓ over $\{d \in \mathcal{D} : Ad \le C_T/T\}$.*

Here, $\mathcal{D}$ is the set of feasible demand vectors, and $\ell(d)$ is the mean revenue in one period at the price $B^{-1}(d - \alpha)$. The limiting fluid solution is unique, its price is bounded away from the boundary of the price set, and every binding resource has a multiplier of at least $\lambda_{\min}$. We use this assumption to keep the binding rows of the fitted fluid problem unchanged for large T when the fitted demand parameters are sufficiently close to the true parameters and C_T/T is sufficiently close to $\bar{c}$. The firm learns from the realized requests. We make the following assumption on the information available before each pricing decision and on the noise in the requests.

Assumption 4 (Observed potential requests). *(a) On $(\Omega, \mathcal{F}, \{\mathcal{F}^t\}, \mathbb{P})$, prices are $\mathcal{F}^{t-1}$-measurable, or independently randomized before the current shock. Full requests $d^t = f(p^t) + \xi_t$ are observed even when service truncates. For every admissible predictable price, $\xi_t \ge -f(p^t)$ pathwise, $\mathbb{E}[\xi_t \mid \mathcal{F}^{t-1}] = 0$, and, for every unit v and $a \in \mathbb{R}$, $\mathbb{E}[e^{av^\top \xi_t} \mid \mathcal{F}^{t-1}] \le e^{a^2\sigma^2/2}$.*

(b) For a known $D_{\max} < \infty$, every predictable price sequence, constant sequences included, gives $0 \le d_j^t \le D_{\max}$ pathwise for all t and j; hence $0 \le f \le D_{\max}$ on $\mathcal{P}$.

Requests may exceed the remaining capacity, so we define the service rule.

Definition 1 (Mechanical service)**.** The fixed rule satisfies $0 \le \widehat{d} \le d$, $A\widehat{d} \le c$, and serves all requests whenever $Ad \le c$, pathwise. Experiments use proportional truncation $\widehat{d} = \gamma d$, where $\gamma = \min\{1, \min_{j:A_j d>0} c_j/(A_j d)\}$.

The fluid regret of a policy π is $\mathrm{Reg}_T^{\mathrm{fl}}(\pi) = V^{\mathrm{Fluid}}(C_T) - \mathbb{E}\sum_t (p^t)^\top \widehat{d}^t$, also written $\mathrm{Reg}_T^{\mathrm{fl}}$. Proofs write $p_t, c_t, D_t, \widehat{D}_t, \mathcal{F}_t$ for $p^t, c^t, d^t, \widehat{d}^t, \mathcal{F}^t$. The quantities $\widehat{d}^t$ and $\widehat{D}_t$ denote served requests in period t; other hatted symbols are parameter estimates, defined where they arise. A tilde denotes a pseudo-observation used only for estimation. Let $\bar{R} := \max_{p\in\mathcal{P}} p^\top f(p)$ be the largest one-period expected revenue. Anchor error is $e = f(p^0) - d^0$ with $\epsilon_0 = \|e\|_2$. The anchor error is unknown, unlike the error bound η computed from data and used by the selector. We write $x_+ = \max\{x, 0\}$ and $|S|$ for set size. All policies solve the same constrained control problem and differ only in the demand information available.

3 Algorithm and Logarithmic Regret with Full Information

When the demand function $f(\boldsymbol{p}) = \boldsymbol{\alpha} + B\boldsymbol{p}$ is known, resource constraints force a dynamic tradeoff. At each period, the firm must balance immediate revenue against future opportunity: selling too aggressively depletes resources needed later, while withholding inventory forgoes potential revenue.

Prior algorithms achieve $O(\log T)$ regret by repeatedly resolving the fluid problem (2), but require strong non-degeneracy assumptions. For instance, Jasin (2014) assumes the dual LP has a unique optimal vertex $\boldsymbol{\lambda}^*$ and that perturbing the constraint right-hand side preserves which constraints bind at optimality. Degeneracy is pervasive (Bumpensanti and Wang 2020): many constraints bind simultaneously, and the optimal dual solution need not be unique.

With known demand, the firm can solve the fluid problem once and then correct its demand target by the cumulative request residuals. Re-solving the fluid can change the binding constraints as the remaining capacities evolve. The initial fluid solution is fixed, while the residual corrections continue to adjust prices. The residuals record how far realized requests depart from the fixed target. The construction is related to the re-solving heuristic of Jasin (2014). Our analysis uses the diagonal benchmark, observed requests, and mechanical service, and it allows any active set and zero dual prices.

Write $\mathcal{D} = f([L, U]^n)$, $R : \mathbb{R}^n \to \mathbb{R}$, $R(\boldsymbol{d}) = \ell(\boldsymbol{d}) = \boldsymbol{d}^\top B^{-1}(\boldsymbol{d} - \boldsymbol{\alpha})$, and we also let $\boldsymbol{d}_T^*$ be the unique solution of the following problem

$$\max_{\boldsymbol{d}\in\mathcal{D}} \{R(\boldsymbol{d}) : A\boldsymbol{d} \le \boldsymbol{C}_T/T\}. \tag{3}$$

This is (2) written in demand coordinates. The known-demand result needs one more condition on this initial solution.

Assumption 5 (Local fluid-price margin)**.** *There is a constant $\delta_D > 0$, independent of T, such that $\mathrm{dist}(\boldsymbol{d}_T^*, \partial\mathcal{D}) \ge \delta_D$ for every instance in the fluid-scaled sequence.*

The margin keeps small residual corrections implementable inside the price set.

3.1 Algorithm Design

We now turn the target into a pricing policy. Fix an admissible fixed price $\boldsymbol{p}^{\mathrm{F}} \in [L, U]^n$. For $\sigma > 0$, define $a_0 = \frac{\delta_D^2}{8n\sigma^2}$, $K_0 = \frac{2n}{1-e^{-a_0}}$, and choose a constant $c_h \ge \max\left\{2, \frac{6}{a_0}, 1 + \frac{\log K_0}{a_0}\right\}$. When $\sigma = 0$, take $c_h = 2$. Set $h_T = \min\{T, \lceil c_h(1+\log T)\rceil\}$, $K_T = T - h_T$. The policy posts the fixed price in the final h_T periods.

Algorithm 1 Residual-Corrected Fluid Pricing

Input: C_T, A, f, c_h and a fixed price p^{F}.
Solve (3) once and fix K_T.
for $t = 1, \dots, T$ **do**
  if $t \le K_T$ and $q_s \in \mathcal{D}$ for every $s \le t$, with q_s from (4) **then**
    Post $B^{-1}(q_t - \alpha)$.
  else
    Post p^{F}.
  end if
  Observe requests, serve them mechanically, update capacity and record ξ_t.
end for

At the start of period $t \le K_T$, let

$$M_t = \sum_{s<t} \frac{\xi_s}{T-s}, \qquad q_t = d_T^* - M_t. \tag{4}$$

Algorithm 1 posts the price that implements the corrected target until the target leaves $\mathcal{D}$. From then on, and in the last periods, it posts the fixed price. In (4), the residual of period s enters each of the remaining $T-s$ targets with weight $1/(T-s)$, so the deviation of period s is absorbed over the rest of the horizon. The correction at period t has variance $O(\sigma^2/(T-t))$ in each direction. With high probability, the target is in $\mathcal{D}$ until the last $h_T = O(\log T)$ periods.

The residual bridge (30) gives the capacity identity (31). For $t \le K_T$, this identity shows by induction that every request through t is served provided the residual recursion stays in $\mathcal{D}$ through $t+1$. The argument uses only $A \ge 0$, so duplicate, slack and dependent rows are allowed.

3.2 Regret Analysis

The following result shows that, with known demand, residual correction attains logarithmic regret while allowing zero dual prices and changing active sets.

Theorem 1. *Suppose Assumptions 1, 2, 4(a) and 5 hold. Then Algorithm 1 satisfies* $\mathrm{Reg}_T^{\mathrm{fl}}(\pi) \le C_{\mathrm{FI}}(1 + \log T)$, *where* C_{FI} *depends on the fixed problem primitives, including the noise proxy and local margin, but not on* T.

The bound holds for any fixed price and any service rule in Definition 1, and for resource matrices of arbitrary rank, zero or nonunique capacity multipliers and changing binding sets. Thus, by solving the fluid problem once, the policy attains logarithmic regret against the fluid benchmark.

We now explain the proof in Appendix C.2. First, consider a hypothetical setting in which the residual recursion continues after the policy switches to the fixed price. This defines the virtual process, whose prices may lie outside the admissible price set. We bound three losses separately: the quadratic loss of the virtual process, the loss after the target leaves $\mathcal{D}$, and the loss during the fixed-price window. The first loss is the main one. Revenue is quadratic in the target, and the residuals are a martingale difference sequence. For every unit vector v and $t \le K_T$, $\mathrm{Var}(v^\top M_t) \le \sigma^2/(T-t)$, so the expected virtual loss in period t is at most $\frac{n\|B^{-1}+B^{-\top}\|_2}{2}\frac{\sigma^2}{T-t}$, and the total is at most $\frac{n\|B^{-1}+B^{-\top}\|_2\sigma^2}{2}(1+\log T)$. The exit probability is at most T^{-6}; a Cauchy–Schwarz argument bounds the expected exit contribution by $O(T^{-2})$, while the fixed-price window contributes at most $h_T \max_{d\in\mathcal{D}} R(d) = O(\log T)$.

4 Algorithm and Regret with No Information

With known demand, residual correction attains $O(\log T)$ regret (Section 3). When the demand function $f(\boldsymbol{p}) = \boldsymbol{\alpha} + B\boldsymbol{p}$ is unknown, the firm must simultaneously learn the parameters and set prices under resource constraints, balancing exploration (estimating $\boldsymbol{\alpha}, B$ from price experiments) against exploitation (using current estimates to maximize revenue)—the classical learning-earning tradeoff. Resource constraints compound this tradeoff: the continuous, high-dimensional state-action space renders dynamic programming intractable, and capacity coupling across periods creates path-dependent dynamics.

Sufficient exploration is the central challenge. As Keskin and Zeevi (2014) show, a myopic policy that always prices at the current estimated optimum leads to incomplete learning and poor revenue. For the fixed design, mean squared parameter error is bounded by $C/\lambda_{\min}(V_{\ell_T})$, where C depends only on n and σ and V_{ℓ_T} is the design matrix (7). The fixed-design error bound decreases as $\lambda_{\min}(V_{\ell_T})$ grows, so if prices cluster too tightly—always near the current estimate, for example—then insufficient cumulative price variation in some direction keeps $\lambda_{\min}(V_{\ell_T})$ small and estimation stagnates.

Therefore, the firm first posts a fixed design of prices that identifies demand. Algorithm 2 posts this design for $O(\sqrt{T})$ periods, fits projected least squares once, solves one fluid problem with the fitted model, and then corrects the demand targets by the cumulative residuals. The algorithm adapts the parametric self-adjusting heuristic of Chen et al. (2021) to affine demand with observed requests. The unknown parameters lie in a known compact class.

Assumption 6 (Learning parameter class)**.** *(a) Write $\Theta = [\boldsymbol{\alpha}\ B] \in \mathbb{R}^{n\times(n+1)}$. The firm knows a convex compact set $\mathfrak{T}$ containing the true Θ such that, for every $[\boldsymbol{\alpha}'\ B'] \in \mathfrak{T}$, $B' + B'^{\top} \preceq -\kappa_B I_n$, $\boldsymbol{\alpha}' + B'\boldsymbol{p} \geq \mathbf{0}$ for all $\boldsymbol{p} \in [L, U]^n$, for a fixed $\kappa_B > 0$.*

(b) For $z(p) = (1, p^{\top})^{\top}$, a fixed compact $\mathcal{K}_{\text{fp}} \subseteq \mathcal{P}$ supports a measure μ with $\lambda_{\min}(\mathbb{E}_{\mu}[z(p)z(p)^{\top}]) \geq c_{\text{id}} > 0$.

Projection onto $\mathfrak{T}$ keeps the fitted demand nonnegative and its slope matrix invertible, while part (b) permits prices that identify every demand coefficient.

Compactness implies the slope bound $\overline{B} := \sup_{\Theta'=[\boldsymbol{\alpha}'\ B']\in\mathfrak{T}} \|B'\|_{\text{op}} < \infty$. Definition 2 gives an explicit design satisfying Assumption 6(b), and Appendix H.1 constructs the historical class for the selector.

4.1 Algorithm Design

In the exploration phase, the firm first posts the fixed design of Definition 2. In the exploitation phase, it solves one fluid problem with the fitted model and corrects the demand targets by the cumulative residuals, as in Section 3.

Definition 2 (Exploration design)**.** Let $z(\boldsymbol{p}) = (1, \boldsymbol{p}^{\top})^{\top}$ and $n \geq 1$. The exploration design is a finite set $\{\boldsymbol{p}^{(j)}\}_{j=0}^{J_0-1} \subseteq \mathcal{K}_{\text{fp}}$, $J_0 \geq 1$, with a known $\kappa_0 > 0$ satisfying

$$\frac{1}{J_0}\sum_{j=0}^{J_0-1} z(\boldsymbol{p}^{(j)})z(\boldsymbol{p}^{(j)})^{\top} \succeq \kappa_0 I_{n+1}. \tag{5}$$

If a box $Q = \prod_{j=1}^{n}[\ell_j, u_j] \subseteq \mathcal{K}_{\text{fp}}$ with positive widths is known, take $v = (\ell+u)/2$, $0 < \Delta < \frac{1}{2}\min_j(u_j - \ell_j)$, $\boldsymbol{p}^{(0)} = v$ and $\boldsymbol{p}^{(j)} = v + \Delta e_j$ for $1 \leq j \leq n$; Lemma 7 gives (5) with $J_0 = n+1$ and an explicit κ_0.

For fixed design constants $c_L, c_h > 0$, set

$$c_L, c_h > 0 \text{ fixed}, \qquad q_T = \left\lceil \frac{c_L\sqrt{T}}{J_0} \right\rceil \in \mathbb{N}_{\geq 1}, \qquad \ell_T := L_T = J_0 q_T, \qquad h_T = \lceil c_h L_T \rceil, \qquad K_T = T - h_T. \tag{6}$$

Algorithm 2 Explore-Then-Adjust Pricing

Input: $\boldsymbol{C}_T, A, \mathfrak{T}$, the design, c_L, c_h and a fixed price $\boldsymbol{p}^{\mathrm{F}}$.
if $L_T + h_T \geq T$ **then**
    Post $\boldsymbol{p}^{\mathrm{F}}$ throughout.
else
    Cycle through the design for L_T periods.
    Fit projected least squares, $\widehat{\Theta}_{\ell_T} = \Pi_{\mathfrak{T}}(\Theta^{\mathrm{OLS}}_{\ell_T})$, and solve (8); if that problem is infeasible, post $\boldsymbol{p}^{\mathrm{F}}$ from then on.
    for $t = L_T + 1, \ldots, T$ **do**
        if $t \leq K_T$ and $\boldsymbol{q}_s \in \widehat{\mathcal{D}}$ for every $L_T < s \leq t$, with $\boldsymbol{q}_s$ from (9) **then**
            Post $\widehat{B}^{-1}_{\ell_T}(\boldsymbol{q}_t - \widehat{\alpha}_{\ell_T})$.
        else
            Post $\boldsymbol{p}^{\mathrm{F}}$.
        end if
    end for
end if
In every period, observe requests, serve them mechanically and update capacity.

Here q_T is the number of complete design cycles and L_T is the exploration length; h_T is the length of the final fixed-price phase and K_T is the last period at which residual adjustment is allowed. If $L_T + h_T \geq T$, the policy posts the fixed price throughout. We use $c_L = c_h = 1$ by default. Any fixed positive values are allowed.

During the first L_T periods, the policy cycles through the prices in (5). Let

$$V_{\ell_T} = \sum_{t=1}^{L_T} z(\boldsymbol{p}^t) z(\boldsymbol{p}^t)^\top, \tag{7}$$

compute rowwise ordinary least squares $\Theta^{\mathrm{OLS}}_{\ell_T}$, and freeze the joint projection $\widehat{\Theta}_{\ell_T} = \Pi_{\mathfrak{T}}(\Theta^{\mathrm{OLS}}_{\ell_T}) = [\widehat{\alpha}_{\ell_T}\ \widehat{B}_{\ell_T}]$. The regression uses potential requests, not served quantities. Because the policy cycles through the design, $V_{\ell_T} \succeq \kappa_0 L_T I_{n+1}$ by (5), so the root mean squared parameter error is $O(L_T^{-1/2})$, with a constant depending only on n, σ and κ_0, not on T (Lemma 8). The choice $L_T \asymp \sqrt{T}$ balances exploration against exploitation, yielding the $O(\sqrt{T})$ regret rate: exploration has a cost of $O(L_T)$, and the parameter error has a cost of $O(T/L_T)$ over the horizon.

Let $\boldsymbol{C}^+ = \boldsymbol{c}^{L_T+1}$ and $N = T - L_T$. Consider the estimated fluid program

$$\max_{\boldsymbol{d} \in \widehat{\mathcal{D}}} \left\{ \widehat{R}(\boldsymbol{d}) : A\boldsymbol{d} \leq \boldsymbol{C}^+/N \right\}, \quad \widehat{\mathcal{D}} = \widehat{\alpha}_{\ell_T} + \widehat{B}_{\ell_T}[L, U]^n, \quad \widehat{R}(\boldsymbol{d}) = \boldsymbol{d}^\top \widehat{B}^{-1}_{\ell_T}(\boldsymbol{d} - \widehat{\alpha}_{\ell_T}), \tag{8}$$

which chooses a demand target to maximize fitted revenue subject to the remaining capacity per period, $\boldsymbol{C}^+/N$. Here $\boldsymbol{C}^+$ is the capacity left after exploration, N the number of remaining periods, and $\widehat{\mathcal{D}}$ the demand vectors predicted by the fitted model. If this problem is infeasible, the policy posts a fixed price $\boldsymbol{p}^{\mathrm{F}} \in [L, U]^n$ thereafter. Otherwise, let $\widehat{\boldsymbol{d}}^*$ be its unique optimizer. For $t = L_T + 1, \ldots, K_T$, before any switch, set

$$\widehat{\Delta}^s = \boldsymbol{d}^s - (\widehat{\alpha}_{\ell_T} + \widehat{B}_{\ell_T} \boldsymbol{p}^s), \qquad \boldsymbol{q}_t = \widehat{\boldsymbol{d}}^* - \sum_{s=L_T+1}^{t-1} \frac{\widehat{\Delta}^s}{T - s}. \tag{9}$$

The guarantee assumes exact projection, optimization and membership checks. A numerical failure that switches to the fixed price adds its expected loss to the bound. The fixed design gives the Gram lower bound.

The residual recursion keeps the fitted target consistent with observed requests while keeping the parameter estimate fixed.

4.2 Regret Analysis

The next theorem shows that the choice $L_T \asymp \sqrt{T}$ gives $O(\sqrt{T})$ regret at every horizon, with a constant that depends on the primitives and the design but not on T.

Theorem 2 (No-information regret)**.** *Suppose Assumptions 1, 2, 3, 5 and 6 hold, with requests as in Assumption 4(a), service as in Definition 1, and the exploration design of Definition 2. For any fixed $c_L, c_h > 0$, Algorithm 2 satisfies $\mathrm{Reg}_T^{\mathrm{fl}}(\pi) \le C_{\mathrm{NI}}\sqrt{T}$ for every integer $T \ge 1$, where C_{NI} depends on the primitives and on c_L, c_h but not on T.*

The bound is $O(\sqrt{T})$, compared with $O(\log T)$ when demand is known, and the difference is the cost of learning demand parameters. The selector of Section 5.3 runs this policy after reading historical data, so the next lemma gives a history-conditional no-information bound.

Lemma 1 (History-conditional no-information bound)**.** *Suppose Assumptions 1, 2, 3, 5 and 6 hold, service follows Definition 1, and fix the exploration design of Definition 2 and $c_L, c_h > 0$. Let $\mathcal{H}$ be a pre-campaign sigma-field, $\mathcal{G}_t = \mathcal{H} \vee \mathcal{F}^t$, and let Assumption 4(a) hold with $\mathcal{G}_t$ in place of $\mathcal{F}^t$ and the same proxy σ. The prices $\boldsymbol{p}^t$ are those posted by Algorithm 2 run from period one, and V^{Fluid} is written as in Definition 8. For every integer $T \ge 1$, almost surely,*

$$V^{\mathrm{Fluid}}(C_T) - \mathbb{E}\left[\sum_{t=1}^{T} (\boldsymbol{p}^t)^\top \hat{\boldsymbol{d}}^t \mid \mathcal{H}\right] \le C_{\mathrm{NI}}\sqrt{T}, \tag{10}$$

where C_{NI} depends on the primitives and on c_L, c_h, but not on T or on $\mathcal{H}$. The request assumptions must remain valid after conditioning on the historical data; this permits the estimation and regret bounds to hold conditionally, with the same constant for almost every realized history.

The proof is in Appendix D. The next lemma shows that the $O(\sqrt{T})$ rate is minimax optimal.

Lemma 2 (Pure-pricing no-information lower bound)**.** *Consider $f_\eta(p) = 1/2 - p/4 + \eta(p-1)$, $|\eta| \le 1/16$, a one-product subclass with price set $[1/2, 3/2]$, Bernoulli potential requests, unit resource coefficient, initial capacity $C_T = T$, potential requests observed before mechanical fulfillment, and regret measured against the diagonal pure-pricing fluid comparator (2). Then, for every integer horizon $T \ge 1$, every nonanticipating no-information policy π satisfies $\sup_{\eta \in \{-T^{-1/4}/16, T^{-1/4}/16\}} \mathrm{Reg}_T^{\mathrm{fl}}(\pi; f_\eta) = \Omega(\sqrt{T})$.*

The proof is in Appendix D. As in the unconstrained problem of Keskin and Zeevi (2014), for every η the demand at the price $p = 1$ equals $1/4$, and $p = 1$ is optimal when $\eta = 0$. At this price, the observed requests do not depend on η. With $C_T = T$ and Bernoulli potential requests, the resource constraint does not bind, so the bound reflects the cost of learning. No policy without prior information can improve on the $O(\sqrt{T})$ rate. Sections 5 and 6 study how offline information can attain a smaller regret.

5 Improvement with Informed Prices

Sections 3 and 4 give the $O(\log T)$ and $O(\sqrt{T})$ regret of the two extremes. Firms often possess an initial price-demand observation $(\boldsymbol{p}^0, \boldsymbol{d}^0)$ from historical data, pilot studies, or predictive models calibrated on related markets. This setting bridges the full- and no-information extremes. When the error of the observation is

small, the objective is to approach $O(\log T)$ regret. When it is large, the objective is to match or improve on the no-information $O(\sqrt{T})$ baseline.

An informed observation $(\boldsymbol{p}^0, \boldsymbol{d}^0)$, the anchor, tells the firm the approximate demand at one price. Its error is $\epsilon_0 = \|f(\boldsymbol{p}^0) - \boldsymbol{d}^0\|_2$. Theorem 3 gives a condition under which an anchor aligned with the resources is sufficient for the controller to find the corresponding prices using a class of slope matrices in place of an estimate. The controller is given a class that contains the matrix. Section 5.3 builds this class, the plan and the error bound from historical data before the season, with matrix estimation confined to the period before the season. Theorem 4 then chooses between the informed-price algorithm and the no-information algorithm before the season. This section uses the network model and the initial fluid benchmark of Section 2.

5.1 When an Anchor Identifies the Decision

We now show when an anchor, a feasible plan and a parameter set are sufficient for pricing. We denote by $e = f(p^0) - d^0$ the anchor error vector, so $\epsilon_0 = \|e\|_2$; each pair (B', e') specifies a possible slope matrix and anchor error vector. The controller starts from a feasible demand plan d, with remaining target $y_1 = Td$ and per-period target $q_t = y_t/(T - t + 1)$. We now motivate the feedback rule. If (B, e) were known, the firm could post the inverse price $p_{B,e}(q_t)$; that price depends on unknown parameters, so the controller instead adjusts prices using observed requests: when requests exceed the target, projected feedback raises the next price. The controller uses request feedback in place of an estimate of B. It stops updating when the target leaves the ball of radius r around the plan or a terminal condition is satisfied, then posts a fixed price. Appendix E gives the controller and its logarithmic tracking and fallback budgets. Before the season, we fix a class $\mathfrak{C}$ of pairs (B', e') that contains the truth, where each B' is symmetric with $\mu I \preceq -B' \preceq MI$. In order to control the tracking error, these spectral bounds make the deterministic feedback update contract toward the target price. The price for a target q under a pair is $p_{B',e'}(q) = p^0 + (B')^{-1}(q - d^0 - e')$. The firm also chooses a plan d with $Ad \le \rho$, a multiplier $\lambda \ge 0$ and a radius $0 < r < \min_j d_j$, and sets $u = A^\top \lambda$. The slack of the plan is $s = \lambda^\top(\rho - Ad)$. Under the inverse-image condition, $p_{B',e'}(q) \in \mathcal{P}$ for every pair in the class and every target within r of the plan. The only condition on A is entrywise nonnegativity. Definition 12 gives these inputs in full.

The next theorem gives the regret of this controller.

Theorem 3 (Informed pricing with residual and slack)**.** *Adopt Definitions 4, 10 and 12, with Assumption 4 and service by Definition 1 and (1), and run Definitions 13–14 with the budgets in (45). Every path is mechanically feasible and every active request is served fully. For $g_B = 2d - d^0 + B(p^0 - u) - e$ and the initial-fluid comparator (2), the conditional signed loss $L_T(\mathcal{G}_0; \pi^{\mathrm{gpf}})$ defined in Definition 11 satisfies, almost surely on the $\mathcal{G}_0$-event where the displayed hypotheses hold,*

$$L_T(\mathcal{G}_0; \pi^{\mathrm{gpf}}) \le (M + \mu)P^{\mathrm{tr}}_{T,d} + T\{s + \|g_B\|_2^2/(2\mu)\} + (R_{\max} + 2q_+\|u\|_1)J^{\mathrm{tr}}_{T,d}. \tag{11}$$

A known bound

$$R_{\mathrm{grad}}(d, \lambda; \mathfrak{C}) := \sup_{(B',e')\in\mathfrak{C}} \|2d - d^0 + B'(p^0 - u) - e'\|_2 \le \eta \tag{12}$$

replaces $\|g_B\|^2$ by η^2.

The bound (11) has three terms. The first and the last are the tracking and fallback budgets, which are $O(1 + \log T)$ for fixed inputs; their constant factors depend on $n, L, U, \mu, M, \sigma, r, h, D_{\max}, q_+$ and $\|u\|_1$, and are independent of T (see the explicit construction in (45)). The second term is the bias of the plan: s is the slack of the plan, and g_B is the residual of the optimality condition of the true fluid problem at the plan and multiplier. Like the bias $\epsilon_0^2 T$ in the conference version, this bias accumulates over all T periods. Corollary 1 shows that it vanishes when the anchor is exact and aligned with the resources. In that case the bound is logarithmic, below the $O(\sqrt{T})$ rate of the no-information algorithm.

Corollary 1 (Alignment and its residual relaxation)**.** *Apply Theorem 3 with $d = d^*$ and $\lambda = \lambda^0$, where $p^0 = A^\top \lambda^0$, $\lambda^0 \geq 0$, $d^* = d^0/2$, $\rho = Ad^*$, and (12) is verified at d^*. Then (11) has $s = 0$, $u = p^0$ and $\|g_B\|^2 = \epsilon_0^2$. At $e = 0$, $p_B(d^*)$ is the unique initial-fluid solution. More generally, a feasible d, λ and class bound $\|e'\| \leq \bar{\epsilon}$ give the bound $\eta_{\text{al}} = \bar{\epsilon} + \|2d - d^0\|_2 + M\|p^0 - A^\top \lambda\|_2$, retaining the same plan's budgets and slack.*

In Corollary 1, the displayed alignment conditions give zero slack, and an exact anchor also makes the residual zero; the final bound allows errors in the anchor and departures from alignment. When the anchor is not aligned, the plan and the multiplier can instead be computed from the KKT conditions of a nominal fluid problem, as in the next corollary.

Corollary 2 (Nominal-KKT plan correction)**.** *Let $f_C(p) = d^0 + C(p - p^0)$ with $C = C^\top \prec 0$, and let p_C in the interior of $\mathcal{P}$ and $\lambda_C \geq 0$ satisfy the KKT conditions of the nominal fluid problem $\max_{p \in \mathcal{P}} \{p^\top f_C(p) : Af_C(p) \leq \rho\}$: with $d_C = f_C(p_C)$ and $u_C = A^\top \lambda_C$, $Ad_C \leq \rho$, $d_C + C(p_C - u_C) = 0$, $\lambda_C^\top(\rho - Ad_C) = 0$. Suppose every hypothesis of Theorem 3 holds with plan $d = d_C$, multiplier $\lambda = \lambda_C$ and a radius $0 < r < \min_j d_{Cj}$, including (39), for the class $\mathfrak{C}_C = \{(C + E, e') : E = E^\top, \|E\|_{op} \leq \delta_B, \|e'\|_2 \leq \bar{\epsilon}\} \cap \mathfrak{C}$ containing the truth. Then $s = 0$ and, with $\eta = \bar{\epsilon} + \delta_B\|p^0 - u_C\|_2$, $R_{\text{grad}}(d_C, \lambda_C; \mathfrak{C}_C) \leq \eta$; in particular $\|g_B\|_2 \leq \eta$, so (11) holds with $\|g_B\|_2^2$ replaced by η^2.*

Thus the errors of the nominal fluid problem are included in the error bound η. The class, the plan, and the anchor are needed for the logarithmic rate. Appendix F shows that, for two products and an exact anchor, the minimax regret is of order $\sqrt{T}$, the rate of the no-information algorithm. With multiple products the prices stay close to p^*, so the design matrix grows in the direction of $p^* - p^0$ and grows slowly in the other directions, even when p^* is bounded away from p^0. The controller in Theorem 3 uses the class $\mathfrak{C}$ and the plan in place of an estimate of B.

5.2 Historical Accuracy and Policy Selection

Without knowing ϵ_0, an algorithm faces a dilemma: trust the observation (risking that bias accumulates over time) or learn from scratch (forgoing the benefit of a good prior). This is reminiscent of the consistency-robustness tradeoff in learning-augmented algorithms (Lykouris and Vassilvitskii 2021, Mitzenmacher and Vassilvitskii 2022): to improve upon the baseline using predictions, we must know the prediction quality. Here, the anchor $(\boldsymbol{p}^0, \boldsymbol{d}^0)$ plays the role of the prediction. ϵ_0 measures the anchor error, whereas η bounds the fluid optimality residual over the parameter class for the chosen plan and multiplier. A natural question is whether the anchor error could be learned adaptively during the selling horizon. In our setting, unlike in unconstrained settings, testing the informed price multiple times to estimate ϵ_0 consumes capacity that subsequent price corrections cannot restore. For this reason, the selector reads historical data before the season.

Theorem 4 chooses between the informed-price and no-information algorithms before the season, using only the data and the coverage guarantee. The anchor error ϵ_0 is an ℓ_2 error in request units: if the relative error of the anchor for each product j is at most r_{rel} of its mean request $\bar{d}_j$, then $\epsilon_0 \leq r_{\text{rel}}\|\bar{d}\|_2$, which equals $r_{\text{rel}}\bar{d}_1\sqrt{n}$ when all products have the same mean. We denote by μ_0 the fixed lower curvature bound and by c_{ref} the fixed scale used in the selector's scores; their formal definitions are given in the selector construction. The selector chooses the informed-price algorithm exactly when the eligibility event G_T holds and the score inequality $s + \eta^2/(2\mu_0) \leq c_{\text{ref}}\left[T^{-1/2} - (1 + \log T)/T\right]$ is satisfied. With zero slack this means η is below an $O(T^{-1/4})$ threshold (for example 0.1 at $T = 10{,}000$, up to constants that depend on μ_0 and c_{ref}), and the budget is logarithmic when $\eta = O(\sqrt{\log T/T})$. The threshold $T^{-1/4}$ is driven by two competing costs. If the algorithm trusts the informed price but the observation is inaccurate, the resulting bias accumulates as $O(\eta^2 T)$. If it learns from scratch, the learning cost is $O(\sqrt{T})$. The transition occurs where these two

costs balance, $\eta^2 T \approx \sqrt{T}$, that is, $\eta \approx T^{-1/4}$. With zero slack and successful eligibility checks, historical data improve the selector's score when the fluid residual bound η falls below the stated threshold. When the class is built from the historical sample of Definition 20, with N requests at each of $n+1$ prices, the radii of Lemma 11 are $O(\sqrt{\log(1/\delta)/N})$ for fixed n and price step Δ_p; the constant factor depends on n, Δ_p, and σ_{off}, the historical noise scale. With δ_T polynomially small and zero slack, the informed budget is then $O(\log T + T \log T/N)$: a sample of $O(\sqrt{T} \log T)$ requests per price matches the no-information rate, and a sample of order T gives the logarithmic rate. These scales are specific to this construction. A class supplied with a smaller η is an alternative to this sampling construction. Historical data are acquired before the season. Section 6 treats calibration during the season.

5.3 A Pre-Campaign Informed-Price Selector

An exact anchor aligned with the resources makes the bias term vanish in the informed-price regret bound. The firm is initially uncertain whether these conditions hold. A selector chooses between the informed-price algorithm and the no-information algorithm using only observed data. Before period 1, the selector reads the historical data H and computes a parameter set, a demand plan, a resource multiplier and an error bound. An eligibility screen checks spectral bounds, resource feasibility and inverse-price margins. If the screen fails, the firm runs the no-information algorithm. Definitions 15–16 in Appendix G give the full construction and screen.

Definition 3 (Symmetric bounded-request class)**.** The true affine response in Assumption 2(a) satisfies $B = B^\top$, and Assumption 4(b) holds with a fixed known finite $D_{\max}$. Symmetry requires reciprocal cross-price effects, while bounded requests cap revenue in each period and control the loss when historical coverage fails.

Here $s = \lambda^\top(\rho - Ad)$ is unused resource capacity valued at λ, and η bounds the direct residual; the symbols μ_0 and c_{ref} are the curvature bound and score scale, respectively. On the eligibility event G_T, the score of the informed-price algorithm is $a_T = c_{\text{ref}}(1 + \log T) + T\{s + \eta^2/(2\mu_0)\}$, and $a_T = \infty$ otherwise. The score of the no-information algorithm is $b_T = c_{\text{ref}}\sqrt{T}$. Thus a_T adds the bias of the plan to the $O(\log T)$ bound of the informed-price algorithm, and b_T reflects the $O(\sqrt{T})$ regret rate of the no-information algorithm, with the constant accounted for through K. The firm runs the informed-price algorithm only when the eligibility screen passes and $a_T \le b_T$. Otherwise, it runs the no-information algorithm. Definition 17 gives this rule in full.

The next theorem bounds the regret of the selected policy by the smaller of the two scores, up to a constant and the term for coverage failure.

Theorem 4 (Pre-campaign informed-price selector)**.** *Suppose the hypotheses of Theorem 2 hold, the truth lies in the class of Definition 3, and $\rho = C_T/T$. Let π^{sel} be the selected policy of Definition 17, built from the historical sample of Definition 20, the maps and gates of Definitions 15–16, the coverage event (64) from (68), Algorithm 2 and the constants in Definition 22, with Assumption 4(a) for $\mathcal{G}_t = \mathcal{H} \vee \mathcal{F}^t$. On the eligibility event G_T, the maps satisfy $s \ge 0$ and the same-class direct residual bound $R_{\text{grad}}(d, \lambda; \mathfrak{C}) \le \eta$. With $R_{\max} = nUD_{\max}$,*

$$\text{Reg}_T^{\text{fl}}(\pi^{\text{sel}}) \le K\mathbb{E}\min\{a_T, b_T\} + TR_{\max}\delta_T, \quad K = \max\{1, C_{\log}/c_{\text{ref}}, C_{\text{NI}}/c_{\text{ref}}\}. \tag{13}$$

On a subfamily satisfying

$$\Pr(E_T \cap G_T^c) = 0, \tag{14}$$

suppose $s = 0$ and fixed c_η, c_{samp} satisfy

$$\eta \le c_\eta \epsilon_0 + c_{\text{samp}}\sqrt{(1 + \log T)/T} \quad \text{on } E_T, \qquad \epsilon_0 = \|d^0 - f(p^0)\|_2. \tag{15}$$

Algorithm 3 Pre-Campaign Selection

1: Before period one, compute the historical estimates and radii, set $b_T = c_{\text{ref}}\sqrt{T}$, and initialize $a_T = \infty$.
2: Check the spectral conditions before computing inverses. If they pass, compute the plan and error bound and perform the remaining eligibility checks; if any eligibility check fails, keep $a_T = \infty$.
3: If all eligibility checks pass, set $a_T = c_{\text{ref}}(1 + \log T) + T\{s + \eta^2/(2\mu_0)\}$.
4: **if** $a_T \le b_T$ **then**
5: Run the informed-price algorithm of Proposition 4.
6: **else**
7: Run Algorithm 2 for the entire horizon.
8: **end if**

Then

$$\text{Reg}_T^{\text{fl}}(\pi^{\text{sel}}) \le C_{\text{inf}}\mathbb{E}\min\{\sqrt{T}, \epsilon_0^2 T + 1 + \log T\} + TR_{\max}\delta_T, \tag{16}$$

where $C_{\text{inf}} = K\max\{1, c_{\text{ref}} + 2c_{\text{samp}}^2/\mu_0, 2c_\eta^2/\mu_0\}$.

The proof is in Appendix H.2. Bound (16) interpolates between the two regimes. Up to the term for coverage failure, the regret is $O(\log T)$ when ϵ_0 is of order $\sqrt{\log T/T}$, $O(\epsilon_0^2 T)$ when the error is larger, and at most $O(\sqrt{T})$, the rate of the no-information algorithm. The condition $\Pr(E_T \cap G_T^c) = 0$ ensures that the eligibility screen passes almost surely whenever historical coverage holds, allowing the informed score to yield the interpolation bound; without it, a covered history can still select the no-information algorithm, and the first bound remains valid. Algorithm 3 summarizes the selection. Selection uses the observed G_T, s and η together with the curvature bound μ_0 and score scale c_{ref}, both fixed before the historical data is read. The regret constants $C_{\log}$ and C_{NI} enter only the bound. When the screen fails, the bound is that of the no-information algorithm. Both bounds concern regret measured against the fluid benchmark, before accounting for the historical sampling fee; profit after that fee subtracts the fee of Definition 21 once.

The following inputs are fixed before the historical data is read: the identifying design with κ_0, the constants $c_L, c_h, p^{\text{F}}, \mu_0, M_0, r_0, q_0, q_1, u_0$ and the score scale. Thus the screen and the scores use tuning parameters fixed before that record is read. The constants t_{00}, h_0 follow from Definition 22. The historical sample size N and level δ determine the radii of the mean estimates, and the selected d, λ determine η.

Section EC.4.3 reports an empirical comparison of a tuned feedback variant with a matched historical inverse and a no-history baseline.

6 Variance Reduction via Offline Surrogate Models

In the no-information setting (Section 4), the firm must learn the demand parameters (α, B) purely from online observations, incurring $O(\sqrt{T})$ regret. Online platforms routinely maintain pre-trained predictive models based on historical data, customer features, and contextual signals. Such *offline auxiliary data* can provide correlated predictions through surrogate models for unobserved demand at deployment prices. For instance, an e-commerce platform may have a predictive model trained on customer demographics and browsing patterns, even though actual sales under the deployment pricing strategy have not been observed. Ride-hailing services may similarly possess traffic and weather data that correlate with demand.

We show that biased or misspecified *offline surrogate models* reduce estimation variance and improve the constants in the $O(\sqrt{T})$ regret bound. The analysis allows surrogate bias and errors in the ranking of demands across prices. The requirement is *informative correlation*: at a fixed price, the surrogate must be correlated with the noise in the requests. A uniform bound on the remaining noise reduces the statistical part of the regret. Appendix J.1 adds the errors from estimating the coefficient and the center.

6.1 Problem Setup and Offline Surrogate Model

Before the season, the firm may hold historical observations and a demand surrogate trained under another pricing policy. Historical request–surrogate pairs, or pairs observed earlier in the season, calibrate the coefficient and the center. Centering residuals accommodates bias in the surrogate mean. Section EC.2 describes who provides each input, when, at what cost, and what must stay stable between history and the season.

In retail, a frozen model maps current traffic or search measurements and the price to a score for each SKU. The model is fixed, but its inputs vary with current demand, and logs pair the scores with uncensored requests. Only correlation at a fixed price, conditional on the history, helps: a forecast that depends only on price and history has no conditional variance and gives no gain. Scores arrive with the requests, after prices are posted. We make the following assumption on the paired observations.

Assumption 7 (Paired surrogate observations and calibration geometry)**.** *(a) Fix a deterministic calibration region $\mathcal{S}_{\text{sur}} \subseteq [L,U]^n$ and a constant $\lambda_S > 0$ independent of T. Whenever a price $\boldsymbol{p} \in \mathcal{S}_{\text{sur}}$ is posted, the observation consists of the paired sample $(\boldsymbol{D}, \boldsymbol{S})$. With $\boldsymbol{\xi} = \boldsymbol{D} - f(\boldsymbol{p})$ and $\boldsymbol{Z} = \boldsymbol{S} - \mu_S(\boldsymbol{p})$, $\mathbb{E}[\boldsymbol{Z} \mid \boldsymbol{p}, \mathcal{F}^{t-1}] = 0$. Uniformly on $\mathcal{S}_{\text{sur}}$, these conditional moments are deterministic functions of the posted price: $\Sigma_S(\boldsymbol{p}) := \mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^\top \mid \boldsymbol{p}^t = \boldsymbol{p}, \mathcal{F}^{t-1}] \succeq \lambda_S I$ and $\Sigma_{dS}(\boldsymbol{p}) := \mathbb{E}[\boldsymbol{\xi}\boldsymbol{Z}^\top \mid \boldsymbol{p}^t = \boldsymbol{p}, \mathcal{F}^{t-1}]$ is smooth. The optimal coefficient is $\Gamma^\star(\boldsymbol{p}) = \Sigma_{dS}(\boldsymbol{p})\Sigma_S(\boldsymbol{p})^{-1}$.*

(b) Let $\Sigma_d(\boldsymbol{p}) := \mathbb{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top \mid \boldsymbol{p}^t = \boldsymbol{p}, \mathcal{F}^{t-1}]$, $\Sigma_{\text{eff}}(\boldsymbol{p}) := \Sigma_d(\boldsymbol{p}) - \Sigma_{dS}(\boldsymbol{p})\Sigma_S(\boldsymbol{p})^{-1}\Sigma_{Sd}(\boldsymbol{p})$, with $\Sigma_{Sd} := \Sigma_{dS}^\top$. There is a constant $\rho_\star^2 \in [0,1]$ such that, on the calibration region $\mathcal{S}_{\text{sur}}$, $\Sigma_{\text{eff}}(\boldsymbol{p}) \preceq (1-\rho_\star^2)\Sigma_d(\boldsymbol{p})$ for all $\boldsymbol{p} \in \mathcal{S}_{\text{sur}}$, and the adjusted noise $\boldsymbol{\xi}_t^\star := \boldsymbol{\xi}_t - \Gamma^\star(\boldsymbol{p}^t)\boldsymbol{Z}^t$ satisfies $\mathbb{E}[\boldsymbol{\xi}_t^\star \mid \boldsymbol{p}^t, \mathcal{F}^{t-1}] = 0$ and is conditionally sub-Gaussian with variance proxy $(1-\rho_\star^2)\sigma^2$ on $\mathcal{S}_{\text{sur}}$.

(c) The finite prices in (5) belong to $\mathcal{S}_{\text{sur}}$. At each such price, the paired calibration sample identifies $\mu_S(\boldsymbol{p})$ and $\Gamma^\star(\boldsymbol{p})$, either through positive design-point mass or through a prespecified local estimator. The pair $(\boldsymbol{\xi}, \boldsymbol{Z})$ is conditionally sub-Gaussian uniformly over these prices. Nuisance estimates used in period t are computed from independent offline data or from observations available before period t.

The policy computes $\widehat{\Gamma}_t$ and the center from independent offline data in $\mathcal{F}^0$ or from earlier pairs, so both are predictable. Clipping to a known compact set that contains $\Gamma^\star(\boldsymbol{p})$ on $\mathcal{S}_{\text{sur}}$ gives $\|\widehat{\Gamma}_t\|_{\text{op}} \le M_\Gamma$. Appendix J.1 bounds the resulting plug-in and calibration terms.

The correlation floor $\rho_\star^2$ holds uniformly over prices and over demand directions with positive variance, so it cannot exceed $\inf_{p, v: v^\top \Sigma_d(p) v > 0} v^\top \Sigma_{dS}(p)\Sigma_S(p)^{-1}\Sigma_{Sd}(p)v / (v^\top \Sigma_d(p) v)$, with $p \in \mathcal{S}_{\text{sur}}$. The uniform sub-Gaussian proxy in Assumption 7(b) gives the statistical factor in Appendix J.1.

Before period 1, the policy checks that the exploration prices are covered and runs Algorithm 2 if they are not. After the estimate is frozen, the policy uses actual requests only.

6.2 Variance Reduction via Control Variates

We adapt classical control variates: given a variable X with known mean, the estimator $\tilde{Y} = Y - \beta(X - \mathbb{E}[X])$ is unbiased for $\mathbb{E}[Y]$ and, with $\beta = \text{Cov}(Y,X)/\text{Var}(X)$, has reduced variance whenever $\text{Cov}(Y,X) \neq 0$. Applying this to the paired observations $(\boldsymbol{D}^t, \boldsymbol{S}^t(\boldsymbol{p}^t))$, the *pseudo-observation* at an exploration price is

$$\widetilde{\boldsymbol{D}}^t := \boldsymbol{D}^t - \widehat{\Gamma}_t(\boldsymbol{p}^t)\{\boldsymbol{S}^t(\boldsymbol{p}^t) - \widehat{m}_t(\boldsymbol{p}^t)\}. \tag{17}$$

Here $\widehat{m}_t(\boldsymbol{p})$ is the center, an estimate of $\mu_S(\boldsymbol{p})$ that is predictable and computed from independent offline data or earlier pairs. Pseudo-observations enter only the exploration fit, while capacity, service, revenue and the residual updates use requests. Appendix J.1 derives the covariance projection, the Schur-complement variance and the fixed-design estimator. With the optimal coefficient $\Gamma^\star$ and with $\widehat{m}_t = \mu_S$, the conditional

variance of the pseudo-observation is the Schur complement $\Sigma_{\text{eff}}(\boldsymbol{p})$ of Assumption 7(b). This is the residual variance after removing the component of $\boldsymbol{D}^t$ linearly predictable from the centered surrogate. For scalar targets, it equals $\text{Var}(\xi)(1-\rho^2)$, where ρ is the correlation: $\rho = 0.5$ yields 25% reduction, $\rho = 0.7$ yields 49%, and $\rho = 0.9$ yields 81%. By the Schur complement formula, correlation matters more than accuracy: because the surrogate is centered, a high-correlation, moderately biased surrogate outperforms a low-correlation, low-bias predictor.

Finite centers. The center is estimated separately at each price, so its variance and possible bias persist across online repeats. Appendix J.1 keeps these terms and the dependence across folds, and Section EC.2 gives the clipped empirical correction.

6.3 Theoretical Guarantees

The surrogate is used only for exploration regression. The estimation error has two components: an oracle variance floor, the Schur complement, which is irreducible even with the true $\Gamma^\star$, and an empirical error from the plug-in estimates of the coefficient and the center. Let $\eta_{m,L}$ and $\eta_{\Gamma,L}$ be bounds on the errors of the center and the coefficient at the exploration prices, obtained before the season on an event that fails with probability at most $\delta_{\text{cal}} \le T^{-3}$. With $D_{\mathfrak{T}}$ the diameter of the parameter class, the error of the fit is measured by $\bar{e}_L^2 = \{(1-\rho_\star^2)\sigma^2 + \eta_{\Gamma,L}^2\}/L_T + \eta_{m,L}^2 + D_{\mathfrak{T}}^2\delta_{\text{cal}}$. Definition 23 gives these conditions in full.

Theorem 5 (Regret with surrogate-assisted learning)**.** *Suppose the hypotheses of Theorem 2 and Assumption 7 hold. Let π^{surr}, the envelopes and $\bar{e}_L$ be as in Definition 23. For $L_T + h_T < T$,*

$$\text{Reg}_T^{\text{fl}}\left(\pi^{\text{surr}}\right) \le C\left\{L_T + h_T + \log\frac{T}{h_T} + T\bar{e}_L^2 + h_T\bar{e}_L\log\frac{T}{h_T} + T\delta_{\text{cal}}\right\} + \Delta_{\text{surr}}(T), \tag{18}$$

where $\delta_{\text{surr}} := \Pr(\mathcal{B}_{L_T}^c)$ is bounded by the explicit load, estimator and deterministic geometry-gate failures, and $\Delta_{\text{surr}}(T) := T\bar{R}\delta_{\text{surr}} + C_{\text{exit}}(T-h_T)\overline{\pi}_{\text{ex},L_T}$ with the exit constant and integrated exit term of Appendix J. For $L_T + h_T \ge T$, complete fallback has regret at most $T\bar{R}$. In particular, if $L_T \asymp h_T \asymp \sqrt{T}$, $\eta_{\Gamma,L} = O(1)$, and $\eta_{m,L} = O(L_T^{-1/2})$, the regret is $O(\sqrt{T})$. Relative to the raw exploration fit, the displayed statistical envelope is strictly smaller whenever $\eta_{\Gamma,L}^2 + L_T\eta_{m,L}^2 + L_T D_{\mathfrak{T}}^2\delta_{\text{cal}} < \rho_\star^2\sigma^2$.

The bound (18) separates the costs. The terms $L_T + h_T$ count the exploration periods and the last periods. The term $T\bar{e}_L^2$ is the cost of the estimation error after the fit is frozen. The term $(1-\rho_\star^2)\sigma^2/L_T$ is the contribution of the oracle noise proxy to the mean squared error bound. The coefficient error $\eta_{\Gamma,L}$ enters in the same way. The error of the center, $\eta_{m,L}$, enters without the factor $1/L_T$, so the center must be estimated with error $O(L_T^{-1/2})$. The remaining terms account for the last periods, calibration failure, the failure of the gates and exits. The theorem's displayed calibration inequality ensures that the surrogate reduces the statistical error envelope.

Appendix J.1 proves the theorem and constructs the envelopes and radius gates from primitive sampling conditions for large horizons, with every acquisition charged separately. The envelopes are known in advance and determine $r_{L,T}$. The constant c_{sur} is used only in the analysis. Total regret also depends on how the fitted parameters change inventory and fallback losses. Appendix J.1 also analyzes a linear estimator that combines corrected historical observations with online requests; that extension requires positive online repeats, a specified error radius, and exact parameter projection (Theorem 6).

The theorem bounds the regret of each policy separately. Proposition 1 compares the actual expected regret of the two policies.

Proposition 1 (Actual regret improvement from the surrogate estimator)**.** *Let π_T^{raw} and π_T^{cv} use the same design, service rule, certainty-equivalent residual policy and last periods $h_T = \lceil c_h(1+\log T)\rceil$, and differ*

only in the exploration estimator (raw least squares or the pseudo-observations (17)). Under the hypotheses of Theorem 7 in Appendix J.2, including $\gamma_Q \geq \underline{\gamma}_Q > 0$,

$$\mathrm{Reg}_T^{\mathrm{fl}}\left(\pi_T^{\mathrm{raw}}\right) - \mathrm{Reg}_T^{\mathrm{fl}}\left(\pi_T^{\mathrm{cv}}\right) = \frac{N_T}{L_T}\gamma_Q + o(\sqrt{T}) > 0$$

for all sufficiently large T*, where* γ_Q *is defined in Definition 24.*

Exits, the last periods and calibration failure are included in the regret of each policy. The gap γ_Q is positive when the covariance matrix of the noise in the requests is positive definite, $\rho_\star > 0$, and $Q_{\mathrm{dyn}} \neq 0$. The comparison is in gross revenue from sales. The gain is $O(N_T/L_T)$, which is $O(\sqrt{T})$ under the exploration schedule. The calibration samples that the proposition uses, $O(\log^2 T)$ paired observations and $O(L_T(1+\log T)^2)$ center observations at each design price, are charged in Section 6.4 when they are purchased. These sample sizes bound the proposition's sampling requirements from above. Whether purchasing these observations is worthwhile depends on their cost, as in (19). The theorem is in Appendix J.2, and its complete proof is in Section EC.1 of the e-companion.

The informed-price and surrogate-assisted approaches serve different roles: informed pricing requires an accurate informed price with a known error bound to achieve $O(\log T)$, while surrogate assistance uses abundant but potentially misspecified offline data to improve the constant in $O(\sqrt{T})$.

6.4 Acquisition Cost and Deployment Value

We now determine how much a firm should pay for the data.

Let Y_T^{S} and Y_T^{B} be the revenue from sales of the surrogate policy and a comparator. If the fee is known in advance and C_{data} is the expected incremental cost of data and computation, the expected incremental profit relative to the comparator is

$$G_T(C_{\mathrm{data}}) := \mathbb{E}(Y_T^{\mathrm{S}} - Y_T^{\mathrm{B}}) - C_{\mathrm{data}} = \mathrm{Reg}_T^{\mathrm{fl}}(\pi^{\mathrm{B}}) - \mathrm{Reg}_T^{\mathrm{fl}}(\pi^{\mathrm{S}}) - C_{\mathrm{data}}. \tag{19}$$

For additional scores, centers and newly purchased request vectors, $C_{\mathrm{data}} = \mathbb{E}[c_{\mathrm{pair}}N_{\mathrm{pair}} + c_{\mathrm{center}}N_{\mathrm{center}} + c_{\mathrm{lab}}N_{\mathrm{lab}}^{\mathrm{new}} + C_{\mathrm{comp}}]$. Existing labels shared by both policies are sunk. Each incremental item is charged once.

Under the same fee, the most a firm should pay for the data is the expected gain in revenue from sales over the comparator. Under the conditions of Proposition 1, π_T^{cv} earns more expected revenue than π_T^{raw} for all sufficiently large horizons. Whether the purchase pays depends on the fee through (19). Section EC.4 estimates this gain on independent paired paths, with simultaneous intervals across scenarios and controls. A fixed cost shifts both ends of each interval equally.

7 Numerical Experiments

Sections 3–6 establish regret bounds for residual correction, informed pricing and surrogate-assisted learning. The experiments ask two questions: does the resource plan identify the pricing decision, and what is the profit of a purchased signal? The benchmark is calibrated to retail transactions (Chen 2015) (Section EC.4). All network policies use the same parameter set, except the oracle, which knows the slope matrix. Revenue includes exploration, failed gates, stockouts and periods at the fixed price. A revenue gain and a reduction in signed fluid regret are the same contrast expressed over different denominators. Sections EC.2–EC.6 of the e-companion give the complete designs, evaluation seeds and test families.

7.1 Full Information and No Information

This study examines how the regret of Algorithms 1 and 2 grows with the horizon. Figure 2 uses five horizons and 200 replications per policy. With unknown demand, regret has log–log slope 0.495 (bootstrap interval $[0.488, 0.501]$), consistent with the $O(\sqrt{T})$ rate. With known demand, regret grows more slowly. In a separate experiment with 2,048 fresh replications per horizon, a mean-zero adjustment chosen on an independent pilot gives a coefficient on $\log T$ of 1.521 (interval $[0.924, 2.119]$); the adjustment improves precision and preserves expected regret. Sections EC.3 and EC.3.2 give the designs and the adjustment. The observed rates match the theory. Full-information regret grows slowly, which is the cost of stochastic constraint management alone, and no-information regret grows as $\sqrt{T}$, the cost of learning demand.

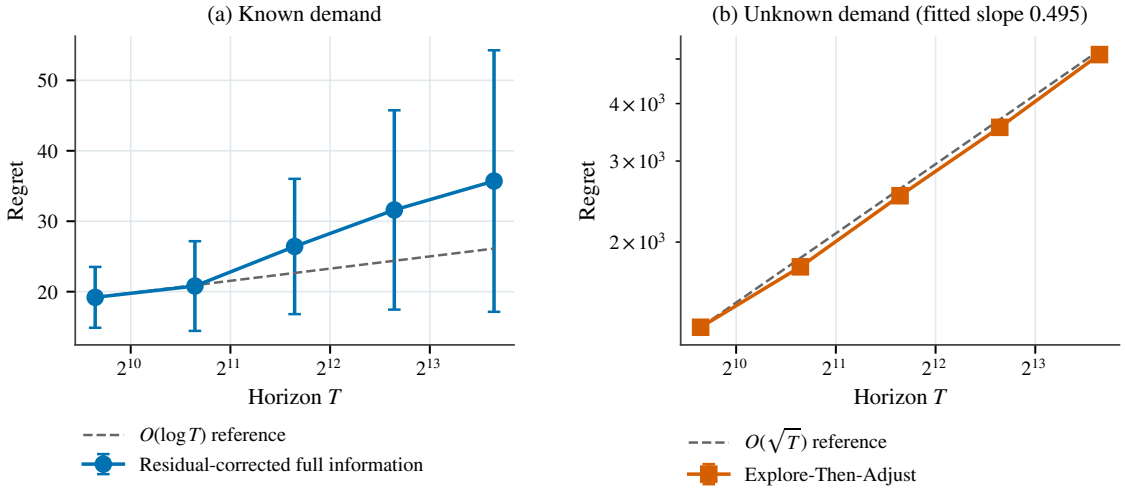


Figure 2: Residual-correction regret: Algorithm 1 with terminal multiplier $c_h = 40$ (a), and Algorithm 2 (b). Means use 200 replications, and bars represent 1.96 standard errors. Dashed references are logarithmic and square-root, respectively.

7.2 Informed Prices and Plan Correction

We study the informed-price algorithm with exact aligned anchors and shifted exact anchors. With exact aligned anchors, regret stays below 2.6 for two products and 4.7 for four products up to $T = 102{,}400$ (Figure 3). Regret varies little over the tested horizons, consistent with the $O(1 + \log T)$ upper bound in Theorem 3 and Corollary 1. With the same parameter set for all policies, plan correction has higher mean revenue than the original target in all 12 shifted cells. It also has higher mean revenue than both nominal-inverse and fixed pricing in all 24 cells (Figure 4). Section EC.5 describes the design and the family of 96 simultaneous revenue contrasts for low and moderate noise.

7.3 Comparison with Controls

We study the revenue screen when the firm does not know in advance which case it faces. The screen compares the revenue of the candidate, the surrogate-assisted algorithm under test, with that of the incumbent, the strongest benchmark policy. We compare the screened policy with 58 benchmark policies on 240 cases and four signal regimes. The policies, the targets and the analysis were fixed before the comparison seeds were run. Every policy is run on four seeds per case and regime. These seeds are disjoint from the development seeds and from the seeds of an earlier comparison (Section EC.6). All policies use the same random noise.

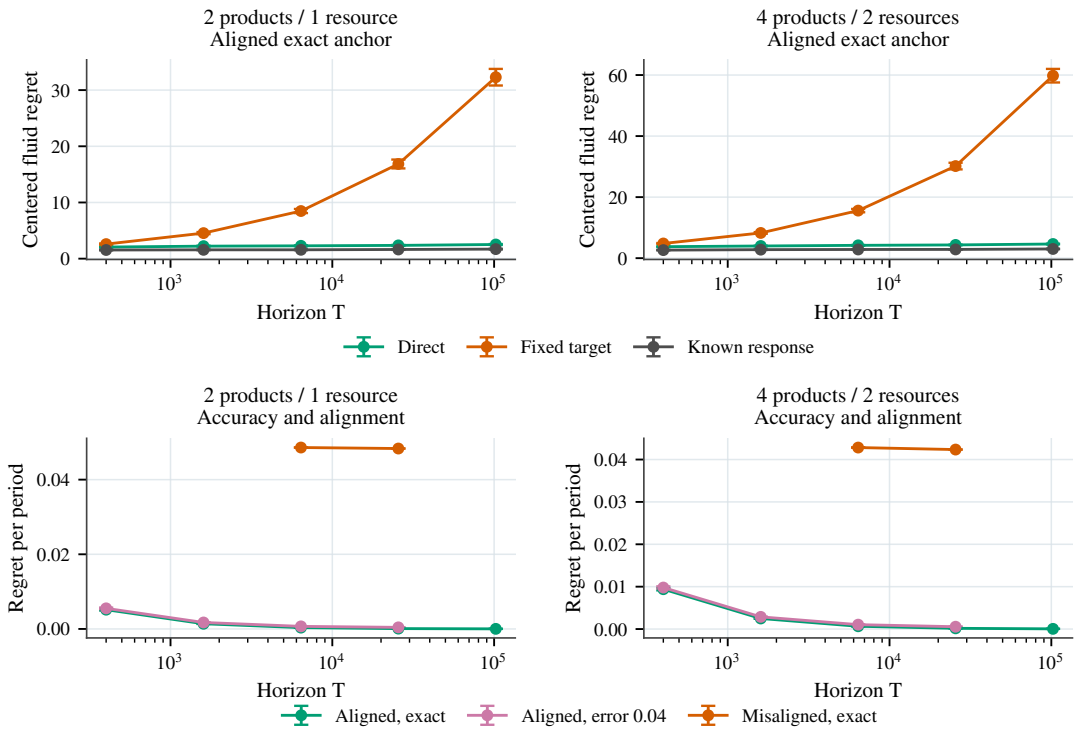


Figure 3: Informed pricing: exact aligned anchors (top), and the regret per period of the informed-price algorithm under forecast error and misalignment (bottom), labeled "Direct" in the top panels. Means and pointwise 95% intervals of centered signed fluid regret, including the fixed-price periods, from 1,000 paired replications. The known-response controller is an oracle. Section EC.5 gives the design and the revenue contrasts.

The requests of each policy are generated at its own posted prices. The four targets are a positive revenue difference, a positive profit-after-fees difference, a revenue gain above 1% of control revenue, and a revenue gain above 5% of benchmark regret. Tests use case-clustered normal approximations with equal case weights and a Bonferroni correction over the 232 components.

The screened policy keeps the fitted candidate and pays its full data fee. It posts the candidate's prices for 32 periods and then posts the maximum price of the incumbent for the next 32 periods. Each block gives a request-based revenue score: the mean revenue of the posted prices at the observed requests, scaled by a resource-load factor (Section EC.6). The factor reduces the score when the mean observed requests would exhaust a resource before the horizon ends. The policy then switches permanently to the incumbent prices if their score exceeds the candidate's score by more than the combined standard error of the two blocks. Otherwise it continues with the candidate. Every probe period counts toward revenue and remaining capacity, and the policy uses only its own prices and observed requests. Continuing with the candidate unless the incumbent's score is higher applies the idea of the selector in Theorem 4 to fitted policies.

Pooled over the 240 cases, all 232 components pass, that is, each of the four targets is met for each of the 58 benchmark policies. Across the 58 benchmark policies, some of which have identical revenue paths, pooled revenue gains range from 3.5% to 5.9% (median 5.1%), and regret reductions range from 47% to 60% (Figure 5). Against the strongest benchmark policy, the incumbent, the gain is 3.5% (standardized statistic 8.65). On the seeds of the earlier comparison, the gains range from 3.6% to 6.3%.

The gain could come from the fit or from the screen. To separate the two, a second policy uses the same

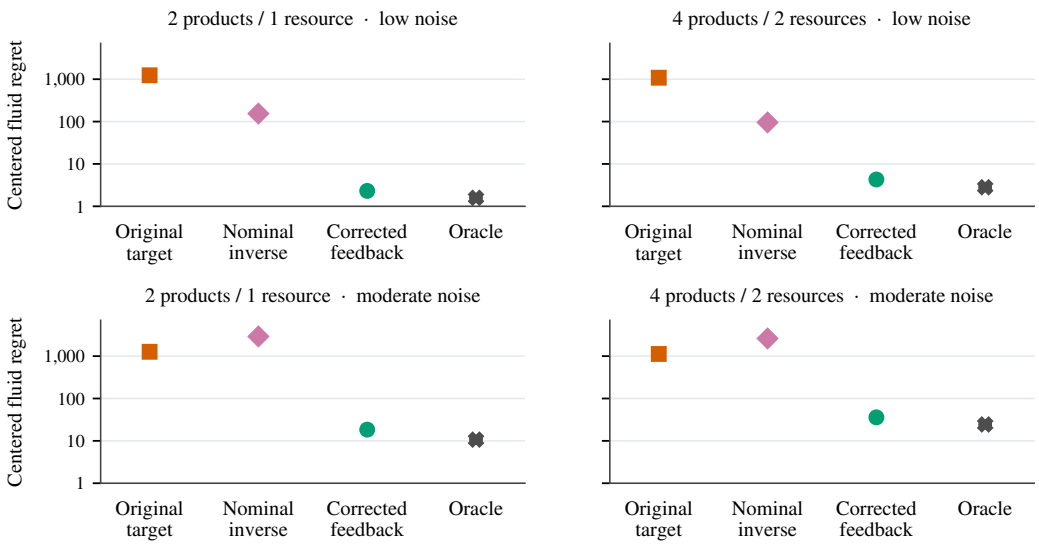


Figure 4: Shifted exact anchors, $T = 25{,}600$: centered fluid regret and pointwise 95% intervals from 1,000 paired replications (log scale). All policies except the oracle use the same parameter set; the oracle knows the slope matrix.

screen, design, controller and decision rule, but estimates the demand parameters by least squares from the same historical data. It pays only for the requests it uses. In a test that was fixed in advance, the screened policy earns 2.8% more revenue than this policy (standardized statistic 11.6), and its profit after fees is also higher. Against the incumbent, the least-squares policy with the same screen has a gain of 0.7%, so about four fifths of the revenue gain over the incumbent is due to the fit. The revenue difference between the screened policy and the least-squares policy is positive at every cross-price coupling and in each group of cases, including the cases in which the signal has no information about the requests, so part of the gain is due to the estimation procedure and not to the signal. Section EC.6 reports these results, the candidate without a screen and the earlier comparison.

7.4 Acquisition Costs

This study examines the profit of a purchased signal. Profit after fees includes the candidate's full data fee even when the screen switches to the incumbent prices. Revenue and profit are measured before computation cost. The empirical fits use history-only data (Section EC.2) and an observed-support projection, defined in the same section. By contrast, Theorems 5 and 6 analyze the policy with positive online repeats and exact Frobenius projection.

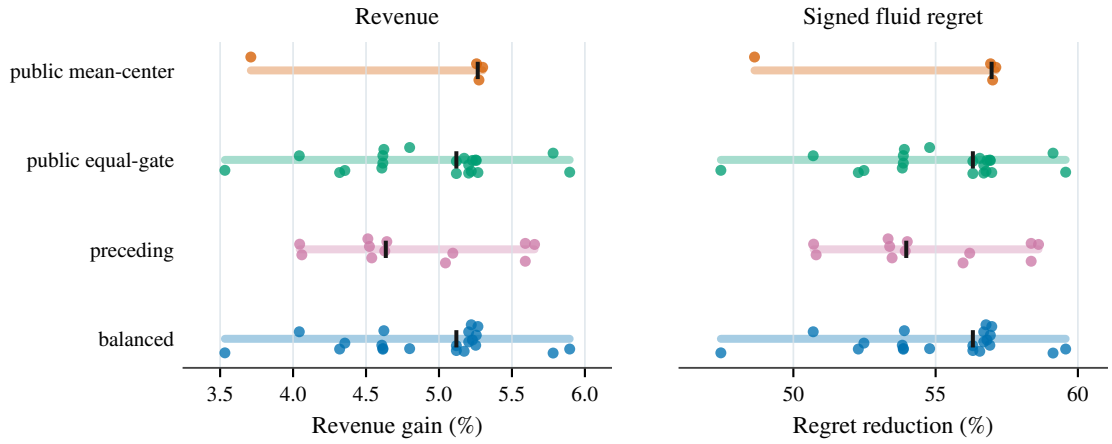


Figure 5: Pooled effects in the comparison with 58 benchmark policies. Each dot is one benchmark policy, grouped by configuration family. Horizontal bars span each family, and black ticks mark family medians. Revenue gain and regret reduction use their own pooled denominators.

8 Conclusions

This paper studies how a resource-constrained pricing system should use imperfect offline information while preserving worst-case robustness. When the anchor is aligned with the resources and the inverse-image condition holds, within a supplied, truth-containing symmetric slope class satisfying the prescribed spectral bounds, the informed-price algorithm finds the corresponding prices without estimating the slope matrix and attains a regret bound that includes loss due to the residual, slack, exit and the fixed price. A selector chooses between the informed-price algorithm and the no-information algorithm using only observed data. At the two extremes, residual correction attains $O(\log T)$ regret with known demand under the local fluid-price margin condition. With no information, deterministic exploration followed by a certainty-equivalent policy attains $O(\sqrt{T})$ regret when the limiting fluid solution is nondegenerate; the constant depends on the fixed problem primitives, the exploration design and its tuning constants, but not on T. Under all hypotheses of Proposition 1, including a strictly positive γ_Q, surrogate-assisted learning reduces regret relative to the matched raw least-squares policy for all sufficiently large horizons. Together, these algorithms decompose regret into three sources: stochastic noise ($O(\log T)$), parameter uncertainty ($O(\sqrt{T})$ or $O(\log T)$ depending on information), and initialization bias ($O(\eta^2 T)$ from informed prices). Together with the no-information lower bound, this decomposition establishes $O(\sqrt{T})$ as the minimax-optimal no-information regret rate.

Three findings translate directly to practice: (i) under the eligibility conditions and with zero slack, an error bound of order $T^{-1/4}$ marks the crossover between the informed and no-information regret bounds, and a logarithmic guarantee follows when $\eta = O(\sqrt{\log T/T})$ and the coverage-failure contribution is $O(\log T)$; (ii) offline surrogate models reduce variance only when they are correlated with the noise in the requests at a fixed price; and (iii) the combined incremental cost of data and computation should not exceed the expected gain in revenue from sales over the comparator.

In our experiments, the value of a signal depends on the comparator, on how well the policy recovers from errors, and on acquisition cost. In a paired study with 240 cases, the revenue screen raises pooled revenue and profit after fees, before computation cost, against all 58 controls, with revenue gains from 3.5% to 5.9% (median 5.1%), including 3.5% against the incumbent. Under the same screen, the surrogate-assisted

algorithm earns 2.8% more revenue than a policy based on a least-squares estimate.

These results point to three directions for future work. First, our analysis assumes that demand is affine. Extending the regret guarantees to nonlinear parametric models requires new proof techniques. Second, the selector reads historical data before the season. An interesting direction is to adaptively infer the reliability of offline data. Third, regret guarantees for extensions of the surrogate framework to nonparametric demand models with general kernels require concentration techniques beyond the analysis in Section 6.

Data and Code Accessibility

Replication materials supply code, frozen configurations, records, instructions for comparisons and figures, software versions, checksums, seeds and inference families, distinguishing saved-data analysis from reruns (complete supported-precision and deployment replays: Section EC.4). Public UCI Online Retail calibration (Chen 2015) includes processed summaries and retrieval and preprocessing instructions.

References

Agrawal S, Devanur N (2016) Linear contextual bandits with knapsacks. *Advances in neural information processing systems* 29.

Agrawal S, Devanur NR, Li L (2016) An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. *Conference on Learning Theory*, 4–18 (PMLR).

Angelopoulos AN, Duchi JC, Zrnic T (2024) PPI++: Efficient prediction-powered inference. *arXiv preprint arXiv:2311.01453* URL https://arxiv.org/abs/2311.01453.

Anonymous (2025) Learning to price with resource constraints: From full information to machine-learned prices. *Advances in Neural Information Processing Systems*, URL http://dx.doi.org/10.52202/085713-3053.

Ao R, Chen H, Gao S, Li H, Simchi-Levi D (2026a) Best arm identification with llm judges and limited human. *arXiv preprint arXiv:2601.21471* .

Ao R, Chen H, Liu H, Simchi-Levi D, Sun WW (2026b) Ppi-svrg: Unifying prediction-powered inference and variance reduction for semi-supervised optimization. *arXiv preprint arXiv:2601.21470* .

Ao R, Chen H, Simchi-Levi D, Zhu F (2024a) Online local false discovery rate control: A resource allocation approach. *arXiv preprint arXiv:2402.11425* .

Ao R, Dong J, Luo G, Simchi-Levi D (2026c) Service-induced congestion in memory-constrained LLM serving. *arXiv preprint arXiv:2606.15555* .

Ao R, Fu H, Simchi-Levi D (2024b) Two-stage online reusable resource allocation: Reservation, overbooking and confirmation call. *arXiv preprint arXiv:2410.15245* .

Ao R, Luo G, Simchi-Levi D, Wang X (2025) Optimizing llm inference: Fluid-guided online scheduling with memory constraints. URL https://arxiv.org/abs/2504.11320.

Ao R, Simchi-Levi D, Wang X (2026d) Optirepair: Closed-loop diagnosis and repair of supply chain optimization models with llm agents. *arXiv preprint arXiv:2602.19439* .

Ao R, Simchi-Levi D, Wang X (2026e) Solver-in-the-loop: Mdp-based benchmarks for self-correction and behavioral rationality in operations research. *arXiv preprint arXiv:2601.21008* .

Badanidiyuru A, Langford J, Slivkins A (2014) Resourceful contextual bandits. *Conference on Learning Theory*, 1109–1134 (PMLR).

Baek J, Chen Y, Chi Z, Ma W (2026a) Evaluating llm-persona generated distributions for decision-making. *arXiv preprint arXiv:2602.06357* .

Baek J, Fu Y, Ma W, Peng T (2026b) Ai agents for inventory control: Human-llm-or complementarity. *arXiv preprint arXiv:2602.12631* .

Bhaskara A, Cutkosky A, Kumar R, Purohit M (2020) Online linear optimization with many hints. URL https://arxiv.org/abs/2010.03082.

Bhaskara A, Cutkosky A, Kumar R, Purohit M (2021) Logarithmic regret from sublinear hints. URL https://arxiv.org/abs/2111.05257.

Bhaskara A, Cutkosky A, Kumar R, Purohit M (2023) Bandit online linear optimization with hints and queries. Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 2313–2336 (PMLR), URL https://proceedings.mlr.press/v202/bhaskara23a.html.

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Operations Research* 60(4):965–980.

Bu J, Simchi-Levi D, Xu Y (2020) Online pricing with offline data: Phase transition and inverse square law. *International Conference on Machine Learning*, 1202–1210 (PMLR).

Bumpensanti P, Wang H (2020) A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science* 66(7):2993–3009.

Chen D (2015) Online Retail. UCI Machine Learning Repository, URL http://dx.doi.org/10.24432/C5BW33, dataset, CC BY 4.0.

Chen Q, Jasin S, Duenyas I (2021) Joint learning and optimization of multi-product pricing with finite resource capacity and unknown demand parameters. *Operations Research* 69(2):560–573.

Duan Y, Hu Y, Jiang J (2025) Ask, clarify, optimize: Human-llm agent collaboration for smarter inventory control. *arXiv preprint arXiv:2601.00121* .

Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using thompson sampling. *Operations research* 66(6):1586–1602.

Gallego G, Van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science* 40(8):999–1020.

Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (New York: Springer).

Jasin S (2014) Reoptimization and self-adjusting price control for network revenue management. *Operations Research* 62(5):1168–1178.

Jasin S, Kumar S (2012) A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* 37(2):313–345.

Jasin S, Kumar S (2013) Analysis of deterministic lp-based booking limit and bid price controls for revenue management. *Operations Research* 61(6):1312–1320.

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

Jiang J, Ma W, Zhang J (2025) Degeneracy is ok: Logarithmic regret for network revenue management with indiscrete distributions. *Operations Research* .

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations research* 62(5):1142–1167.

Li X, Ye Y (2022) Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research* 70(5):2948–2966.

Li Y, Xie H, Lin Y, Lui JC (2021) Unifying offline causal inference and online bandit learning for data driven decision. *Proceedings of the Web Conference 2021*, 2291–2303.

Liang K, Lu Y, Mao J, Sun S, Yang C, Zeng C, Jin X, Qin H, Zhu R, Teo CP (2026) Large-scale optimization model auto-formulation: Harnessing llm flexibility via structured workflow. *arXiv preprint arXiv:2601.09635* .

Liu H, Grigas P (2022) Online contextual decision-making with a smart predict-then-optimize method. *arXiv preprint arXiv:2206.07316* .

Liu S, Jiang J, Li X (2022) Non-stationary bandits with knapsacks. *Advances in Neural Information Processing Systems* 35:16522–16532.

Lykouris T, Vassilvitskii S (2021) Competitive caching with machine learned advice. *Journal of the ACM (JACM)* 68(4):1–25.

Ma W, Xia D, Jiang J (2024) High-dimensional linear bandits with knapsacks. *International Conference on Machine Learning*, 34008–34037 (PMLR).

Mitzenmacher M, Vassilvitskii S (2022) Algorithms with predictions. *Communications of the ACM* 65(7):33–35.

Purohit M, Svitkina Z, Kumar R (2018) Improving online algorithms via ml predictions. *Advances in Neural Information Processing Systems* 31.

Reiman MI, Wang Q (2008) An asymptotically optimal policy for a quantity-based network revenue management problem. *Mathematics of Operations Research* 33(2):257–282.

Sankararaman KA, Slivkins A (2021) Bandits with knapsacks beyond the worst case. *Advances in Neural Information Processing Systems* 34:23191–23204.

Simchi-Levi D, Xu Y (2022) Bypassing the monster: A faster and simpler optimal algorithm for contextual bandits under realizability. *Mathematics of Operations Research* 47(3):1904–1931.

Sivakumar V, Zuo S, Banerjee A (2022) Smoothed adversarial linear contextual bandits with knapsacks. *International Conference on Machine Learning*, 20253–20277 (PMLR).

Talluri KT, van Ryzin GJ (2004) *The Theory and Practice of Revenue Management* (New York: Springer).

Wang Y, Wang H (2022) Constant regret resolving heuristics for price-based revenue management. *Operations Research* 70(6):3538–3557.

Wang Y, Zheng Z, Shen ZJM (2024) Online pricing with polluted offline data. *Available at SSRN 4320324* .

Wei A, Zhang F (2020) Optimal robustness-consistency trade-offs for learning-augmented online algorithms. *Advances in Neural Information Processing Systems* 33:8042–8053.

Xu Y, Zeevi A (2020) Upper counterfactual confidence bounds: a new optimism principle for contextual bandits. *arXiv preprint arXiv:2007.07876* .

Zrnic T, Candès EJ (2024) Cross-prediction-powered inference. *arXiv preprint arXiv:2309.16598* URL https://arxiv.org/abs/2309.16598.

A Relation to the Conference Version and Scope of the Results

The NeurIPS 2025 version (Anonymous 2025) proves logarithmic regret with full information (its Theorem 2.1), square-root regret with unknown demand (Theorem 3.1) and an interpolation for informed pricing (Theorem 4.2). This paper keeps those rate exponents, the price set and the resource constraints. The firm observes the realized demand and serves requests mechanically, and regret is the signed difference between the initial fluid value and revenue from sales. Table 1 relates each result to its counterpart, and Table 2 lists the conditions of each result beyond the model of Section 2 and what it guarantees. Rate claims refer to these theorems, and revenue and net-value claims to the experiments in Section 7.

Table 1: Results of the conference version and of this paper.

Topic	Conference version	This paper
Actions and service	Bounded price set; demand targets with service accounting inside the algorithms	Every policy posts prices in the box and serves requests by the fixed rule of Definition 1; revenue and capacity follow the served requests.
Benchmark	Initial fluid value	The same program (2), with signed regret against revenue from sales (Section 2).
Full information	Theorem 2.1: logarithmic regret with boundary attraction	Theorem 1: residual correction after one fluid solve attains logarithmic regret under the local margin of Assumption 5, for any rank of A and zero or nonunique multipliers.
No information	Theorem 3.1: square-root regret with perturbed prices	Theorem 2: deterministic exploration, a frozen projected fit and residual adjustment attain square-root regret under Assumption 3.
Informed pricing	Theorem 4.2: interpolation in the anchor error	Theorem 3: a feedback controller learns the implementing prices in a network, with regret in the residual and slack of the plan; Theorem 4: an observed-data selector with the interpolation (16) under alignment, the symmetric class and the inverse-image condition.
Rate with an exact anchor	—	Proposition 3: for two products and an exact anchor, the minimax regret is of order $\sqrt{T}$, the rate of the no-information algorithm (Appendix F).
Surrogate information	—	Theorem 5: statistical envelope with calibration, centering, overlap and failure terms; Proposition 1 and Theorem 7: actual regret comparison; identity (19): acquisition and computation costs.

The regret bound of the conference version for the informed price (its Theorem 4.2) rests on two steps of its proof. Its Lemma E.2 bounds the minimum eigenvalue of the design matrix $\sum_s (p^s - p^0)(p^s - p^0)^\top$ below by $\mathrm{tr}(\cdot)/n$, and its constant depends on $\|p^* - p^0\|_2^{-2}$. The first step is exact for a single product. Proposition 5 shows that the second step is needed: when $\|p^* - p^0\|_2$ can be of order $T^{-1/4}$, the minimax regret is of order $\sqrt{T}$ even when the anchor is exact. With multiple products the prices stay close to p^*, so the design matrix grows in the direction of $p^* - p^0$ and grows slowly in the other directions, even when p^* is bounded away from p^0; the proof of the conference version did not discuss this case in detail. Proposition 3 gives the minimax regret of order $\sqrt{T}$ for a fixed family with two products. Theorem 3 gives the condition under which the anchor identifies the direction that moves the optimal solution, and under this condition, Theorem 4 attains the regret bound of the conference version. The lower bounds of the conference version

Table 2: Conditions and guarantees of the main results.

Result	Conditions beyond the model of Section 2	Guarantee
Theorem 1 (full information)	Local fluid-price margin (Assumption 5).	Logarithmic regret of Algorithm 1 against the initial fluid value; any rank of A, zero or nonunique multipliers and changing binding sets.
Theorem 2 (no information)	Assumptions 3, 5 and 6; the exploration design of Definition 2.	Regret of Algorithm 2 at most $C_{\mathrm{NI}}\sqrt{T}$ for every $T \geq 1$.
Theorem 3 (informed pricing)	The standing model of Definition 4; a symmetric class with spectral bounds $\mu \leq M$ containing the truth; bounded requests (Assumption 4(b)); a feasible plan, multiplier and radius with the inverse-image condition (39).	Conditional signed loss at most the tracking and fallback budgets plus $T\{s + \|g_B\|_2^2/(2\mu)\}$; residual and slack vanish for an exact aligned anchor.
Theorem 4 (selector)	The hypotheses of Theorem 2; the truth in the class of Definition 3; the historical sample of Definition 20 at level δ_T; the eligibility screen of Definition 16.	Regret at most $K\mathbb{E}\min\{a_T, b_T\} + TR_{\max}\delta_T$ with observable a_T, b_T; on the aligned subfamily, the interpolation (16).
Theorem 5 (surrogate envelope)	The hypotheses of Theorem 2; Assumption 7; the calibration envelopes and radius gate.	The bound (18) with calibration, centering, overlap and failure terms; order $\sqrt{T}$ when $L_T \asymp h_T \asymp \sqrt{T}$.
Proposition 1 and Theorem 7 (actual regret)	The hypotheses of Theorem 7 (Appendix J.2), including Assumption 3, calibration samples of order $\log^2 T$ pairs and $L_T(1 + \log T)^2$ centers at each design price, and $\gamma_Q \geq \underline{\gamma}_Q > 0$.	$\mathrm{Reg}_T^{\mathrm{fl}}(\pi_T^{\mathrm{raw}}) - \mathrm{Reg}_T^{\mathrm{fl}}(\pi_T^{\mathrm{cv}}) = (N_T/L_T)\gamma_Q + o(\sqrt{T})$, positive for all sufficiently large T.
Proposition 3 (rate with an exact anchor)	A fixed two-product family with a shared resource; the policy knows the family and that the anchor is exact.	Minimax regret of order $\sqrt{T}$, the rate of Theorem 2.

(its Propositions 4.1 and 4.3) hold without knowledge of ϵ_0. Propositions 3 and 5 hold even with knowledge of the exact anchor and the family.

The tuning parameters of the conference version map to the current policies as follows. The perturbation scale σ_0 is replaced by the fixed identifying design of Definition 2 and its schedule constants c_L, c_h. $c_L = c_h = 1$ is the default, and any fixed positive values keep the $\sqrt{T}$ rate. The boundary-attraction threshold ζ is replaced by residual correction, whose only tuning constant is the terminal-window multiplier c_h of Section 3, given explicitly in terms of the margin and noise. The informed tolerance τ is replaced by the score scale $c_{\mathrm{ref}} > 0$ of Theorem 4, which is fixed before historical data is read. It shifts the selection threshold $a_T \le b_T$ and enters the constant K in (13).

B Standing Model and Notation

The following definitions fix the model and the notation shared by the proofs.

Definition 4 (Standing model and notation for the three analyses)**.** Throughout the three analyses, $n, m \ge 1$, $\mathcal{P} = [L, U]^n$ with $0 \le L < U < \infty$, $A \in \mathbb{R}_+^{m\times n}$, $\rho \in \mathbb{R}_+^m$, and $C_T := T\rho$. A posted price p^t is predictable and the observed potential request is

$$d^t = f(p^t) + \xi_t, \qquad \xi_t \in \mathcal{F}^t, \qquad \mathbb{E}[\xi_t \mid \mathcal{F}^{t-1}] = 0, \quad \mathbb{E}[e^{\theta v^\top \xi_t} \mid \mathcal{F}^{t-1}] \le e^{\theta^2\sigma^2/2} \quad (\|v\|_2 = 1,\ \theta \in \mathbb{R}). \tag{20}$$

Requests satisfy $0 \le d_j^t \le D_{\max}$ pathwise, and constant prices are admissible, so $0 \le f_j(p) \le D_{\max}$ on the whole box. The fixed mechanical rule satisfies $0 \le \widehat{d}^t \le d^t$, $A\widehat{d}^t \le c^t$, and $\widehat{d}^t = d^t$ whenever $Ad^t \le c^t$, with $c^{t+1} = c^t - A\widehat{d}^t$ and $c^1 = C_T$. By Assumption 2(a), $f(p) = \alpha + Bp$ on the entire box, so $d^s = \Theta z_s + \xi_s$ for the design observations. For the fixed-design leaf, $z_s := (1, (p^s)^\top)^\top$ is deterministic before ξ_s, $V_{\ell_T} := \sum_{s=1}^{\ell_T} z_s z_s^\top \succeq \kappa_0 \ell_T I_d$, $J_0 \ge 1$, $q := q_T \ge 1$, $\ell_T := L_T := J_0 q > 0$, and $d := n + 1$. The unprojected and projected estimators are, respectively,

$$\widetilde{\Theta}_{\ell_T} := \Big(\sum_{s=1}^{\ell_T} d^s z_s^\top\Big) V_{\ell_T}^{-1}, \qquad \widehat{\Theta}_{\ell_T} := \Pi_{\mathfrak{T}}^{\|\cdot\|_F}(\widetilde{\Theta}_{\ell_T}), \qquad \Theta = [\alpha\ B]. \tag{21}$$

Here $\kappa_0 > 0$ and $\mathfrak{T}$ is compact, convex, and contains Θ. For the fluid and frozen-fit leaves, put $\mathcal{D}_\vartheta := \alpha_\vartheta + B_\vartheta \mathcal{P}$, $K(c) := \{v \in \mathcal{D}_\vartheta : Av \le c\}$, $R_\vartheta(v) := v^\top B_\vartheta^{-1}(v - \alpha_\vartheta)$, $W_\vartheta(c) := \max_{v\in K(c)} R_\vartheta(v)$, $c_0 := C_T/T$, and let $x^* = d_0 = d_T^*$ maximize $W_\vartheta(c_0)$. The change of variables in (2) gives $V^{\mathrm{Fluid}}(C_T) = TW_\vartheta(c_0) = T\ell(x^*)$, where $\ell(v) := R_\vartheta(v)$. Let $\bar{R} := \max_{p\in\mathcal{P}} p^\top f(p) \le nUD_{\max}$. Since the fluid problem maximizes $p^\top f(p)$ over a subset of $\mathcal{P}$, $V^{\mathrm{Fluid}}(C_T) \le T\bar{R}$. The image-margin gate supplies a common ball $B(d_0, r_0) \subseteq \mathcal{D}_\vartheta \cap \mathcal{D}_{\widehat{\vartheta}}$ whose inverse prices lie in $\mathcal{P}$. Both images are nonnegative. For the feedback leaf, $r^{\mathrm{fl}} := V^{\mathrm{Fluid}}(C_T)/T$ and $V := n\sigma^2$, $H_0 := 0$, $H_k := \sum_{j=1}^k j^{-1}$ for $k \ge 1$, and $R_{\max} := \bar{R}$, so $r^{\mathrm{fl}} \le R_{\max}$. The quantity $\mathrm{Reg}_T^{\mathrm{fl}} := V^{\mathrm{Fluid}}(C_T) - \mathbb{E}\sum_{t=1}^T (p^t)^\top \widehat{d}^t$ is the expected signed loss.

Definition 5 (Frozen-fit geometry and the virtual controller before stopping)**.** For a freeze time $s < T$, put $N := T - s$, $b := c^{s+1}/N$, $c_0 := C_T/T$, $\Delta := \|b - c_0\|_\infty$ and $e := \|\widehat{\Theta}_s - \Theta\|_F$. Write $\vartheta := (\alpha, B)$ and, for every class member $\eta = (\alpha_\eta, B_\eta)$, define $\mathcal{D}_\eta := \alpha_\eta + B_\eta\mathcal{P}$ and $K_\eta(c) := \{v \in \mathcal{D}_\eta : Av \le c\}$. $\widehat{\vartheta} = (\widehat{\alpha}, \widehat{B})$, $\widehat{p}(v) := \widehat{B}^{-1}(v - \widehat{\alpha})$ and $g(v) := \alpha + B\widehat{p}(v)$ for every $v \in \mathbb{R}^n$. The inverse is defined on all of $\mathbb{R}^n$ because every member of the learning class has a bounded inverse. Let $x := X(\widehat{\vartheta}, b)$ and $v := X(\vartheta, b)$ be the fitted and true maximizers in $K_{\widehat{\vartheta}}(b)$ and $K_\vartheta(b)$, where X is a fixed Borel map returning the maximizer (unique, by strict concavity of each class member's revenue) when the set is nonempty and the fixed point d_0 otherwise. When $K_{\widehat{\vartheta}}(b) = \varnothing$ the controller uses complete fallback from $s+1$, and this occurs only off the good event, since the gate's feasibility clause $c^{s+1} - NAx \ge 0$ with $x \in \mathcal{D}_{\widehat{\vartheta}}$ makes $K_{\widehat{\vartheta}}(b)$ nonempty. Thus x, v are $\mathcal{G}_s$-measurable.

Set $J := \{i : A_i x^* = c_{0,i}\}$, and use a multiplier $\lambda_J \geq 0$ with $\nabla \ell(x^*) = A_J^\top \lambda_J$ when the KKT gate below holds. The local-margin part of that gate (where $\mathcal{D}$ in Assumption 5 denotes $\mathcal{D}_\vartheta$) is the true-image inclusion $\mathbb{B}_2(x, r_0) \subseteq \mathcal{D}_\vartheta$. When $J \neq \varnothing$ it also requires A_J to have full row rank with $\sigma_{\min}(A_J) \geq s_J > 0$ (for $J = \varnothing$ set $\lambda_J = 0$ and $C_\lambda = 0$ below). Set $C_f := 1 + \sup_{p \in \mathcal{P}} \|p\|_2$, so $\sup_{p \in \mathcal{P}} \|\widehat{\alpha} + \widehat{B}p - (\alpha + Bp)\|_2 \leq C_f e$. Fix positive constants $c_\Delta, r_{\mathrm{loc}}, C_\Delta, C_g, C_a, C_x$ and s_J depending only on the named primitive bounds and this geometry, never on T, s, e, Δ. All displayed norms are Euclidean. Also write $\|\nabla \ell\|_{2,\max} := \sup_{v \in \mathcal{D}_\vartheta \cap \mathbb{R}^n} \|\nabla \ell(v)\|_2 < \infty$. For $A \geq 0$, a positive row has $A_i \mathbf{1} > 0$ and $a_A := \min_{i : A_i \mathbf{1} > 0} A_i \mathbf{1}$. A zero row has $A_i \mathbf{1} = 0$ and is identically zero. When there is a positive row set $\tau_b := \Delta / a_A$, and set $\tau_b := 0$ if all rows are zero. The deterministic geometry gate is

$$
\begin{gathered}
\|v - d_0\|_2 \leq C_\Delta \sqrt{\Delta}, \quad \|x - v\|_2 \leq C_g(e + \sqrt{\Delta}), \quad \|g(x) - x\|_2 \leq C_a(e + \sqrt{\Delta}), \\
\|x - x^*\|_2 \leq C_x(e + \sqrt{\Delta}), \\
\|x - x^*\|_2 \leq r_0/4, \quad c^{s+1} - NAx \geq 0, \\
A_J x = b_J, \quad A_{J^c} x < b_{J^c}, \quad \nabla \ell(x^*) = A_J^\top \lambda_J, \quad \text{and, if } J \neq \varnothing,\ \lambda_J > 0,\ \sigma_{\min}(A_J) \geq s_J > 0, \\
4 \log(N/h) \max\{C_H r_s, C_a(e + \sqrt{\Delta})\} \leq \min\{\log 2, r_0/8\}, \\
\Delta \leq c_\Delta, \quad e \leq r_{\mathrm{loc}}, \quad C_g(e + \sqrt{\Delta}) + \sqrt{n}\, \tau_b \leq r_0/2.
\end{gathered}
\tag{22}
$$

The last line is the local-margin gate. The complete shift argument is in the lemma proof. Put $a := g(x) - x$ and $H := B\widehat{B}^{-1} - I$. For every class member B', the condition $B' + B'^\top \preceq -\kappa_B I$ implies, for unit w, $\kappa_B/2 \leq -w^\top B' w \leq \|B' w\|$, hence $\|B'^{-1}\| \leq 2/\kappa_B$. Consequently, with $C_H := 2/\kappa_B$, $H = (B - \widehat{B})\widehat{B}^{-1}$ and $\|H\| \leq C_H e$.

Let $\mathcal{H}$ contain pre-campaign data and randomization and tie the freeze filtration to the standing one by $\mathcal{G}_t := \mathcal{H} \vee \mathcal{F}^t$. Prices are $\mathcal{G}_{t-1}$-measurable. The hypothesis is that, for every $t > s$, the pre-campaign history is uninformative about the future shock:

$$
\mathbb{E}[\xi_t \mid \mathcal{G}_{t-1}] = 0, \qquad \mathbb{E}[e^{\theta w^\top \xi_t} \mid \mathcal{G}_{t-1}] \leq e^{\theta^2 \sigma^2/2} \quad (\|w\|_2 = 1,\ \theta \in \mathbb{R}). \tag{23}
$$

The actual residual controller starts with $y_{s+1} := Nx$ and $q_t := y_t/(T - t + 1)$. Its predictable active test is

$$
I_t^{\mathrm{sur}} := \mathbf{1}\{s < t \leq T - h,\ q_t \in \mathcal{D}_{\widehat{\vartheta}},\ \widehat{p}(q_t) \in \mathcal{P},\ \|q_t - d_0\|_2 \leq r_0/2\}. \tag{24}
$$

It posts $\widehat{p}(q_t)$ when $I_t^{\mathrm{sur}} = 1$, observes $D_t = f(\widehat{p}(q_t)) + \xi_t$, and updates $y_{t+1} := y_t - D_t$. Otherwise it posts p_s^{F} and freezes the auxiliary state. Equivalently,

$$
q_{t+1} = q_t - \frac{D_t - q_t}{T - t} \quad \text{when } I_t^{\mathrm{sur}} = 1, \qquad q_{s+1} = x. \tag{25}
$$

It posts a fixed $\mathcal{G}_s$-measurable box price p_s^{F} after the first failed active test and throughout the last periods $t > T - h$. Mechanical service continues under Definition 1.

For comparison, define on every sample path the unstopped virtual process

$$
\widetilde{z}_{s+1} = 0, \quad \widetilde{q}_t = x + \widetilde{z}_t, \quad \widetilde{p}_t = \widehat{p}(\widetilde{q}_t), \quad \widetilde{D}_t = g(\widetilde{q}_t) + \xi_t, \quad \widetilde{X}_t = \widetilde{p}_t^\top \widetilde{D}_t, \tag{26}
$$

$$
\widetilde{z}_{t+1} = \left(I - \frac{H}{T - t}\right)\widetilde{z}_t - \frac{a + \xi_t}{T - t}, \qquad s + 1 \leq t \leq T - h. \tag{27}
$$

Define the virtual exit event $\mathcal{E}_s := \{\max_{s+1 \leq t \leq T-h+1} \|\widetilde{q}_t - d_0\|_2 > r_0/2\}$ and $\pi_{\mathrm{ex},s} := \Pr(\mathcal{E}_s \mid \mathcal{G}_s)$. The actual/virtual induction is pathwise: on $\mathcal{E}_s^c$, the two prices, requests, service decisions and residuals coincide through $T - h$. For predictable weights w_t, Assumption 4(a) gives

$$
\Pr\left(\left|\sum_t w_t^\top \xi_t\right| > \sigma \sqrt{2vx} \mid \mathcal{G}_s\right) \leq 2e^{-x} \quad \text{when } \sum_t \|w_t\|_2^2 \leq v. \tag{28}
$$

C Proofs for the Full-Information Setting

We prove the linear-quadratic identities and full-information guarantee for observed requests and mechanical fulfillment. The self-adjusting architecture follows Jasin (2014). For comparison, Wang and Wang (2022, Theorem 2) gives logarithmic re-solving regret with positive duals. Its Theorems 3 and 5 give constant regret, respectively for a single product away from the boundary inventory point and multiple products under active-set stability (C1). Li and Ye (2022, Assumption 3 and Proposition 2) concerns online-linear-programming regularity only.

Proposition 2 (Stationary reduction for diagonal pure-pricing relaxations)**.** *Let* $\mathcal{D}_\theta = f_\theta([L,U]^n)$ *and suppose that* R_θ *is concave on* $\mathcal{D}_\theta$. *Then* $\max_{\boldsymbol{d}_1,\dots,\boldsymbol{d}_T\in\mathcal{D}_\theta}\left\{\sum_{t=1}^T R_\theta(\boldsymbol{d}_t) : \sum_{t=1}^T A\boldsymbol{d}_t \le \boldsymbol{C}\right\} \le T\max_{\boldsymbol{d}\in\mathcal{D}_\theta}\{R_\theta(\boldsymbol{d}) : A\boldsymbol{d}\le \boldsymbol{C}/T\}$.

Proof. For any feasible sequence let $\bar{\boldsymbol{d}} = T^{-1}\sum_{t=1}^T \boldsymbol{d}_t$. The affine image of the convex price set is convex, so $\bar{\boldsymbol{d}}\in\mathcal{D}_\theta$, and feasibility gives $A\bar{\boldsymbol{d}}\le \boldsymbol{C}/T$. Jensen's inequality yields $T^{-1}\sum_t R_\theta(\boldsymbol{d}_t)\le R_\theta(\bar{\boldsymbol{d}})$. Taking the supremum proves the claim. □

C.1 Regularity Lemmas for the Linear-Quadratic Model

Definition 6 (Linear-quadratic re-solve primitives)**.** Let $R_\theta(\boldsymbol{d}) = p_\theta(\boldsymbol{d})^\top\boldsymbol{d} = \boldsymbol{d}^\top B^{-1}(\boldsymbol{d}-\boldsymbol{\alpha})$, $r_\theta(\boldsymbol{p}) = \boldsymbol{p}^\top(\boldsymbol{\alpha}+B\boldsymbol{p})$. For a compact capacity-ratio set $\mathcal{Q}$ with nonempty re-solve feasible sets, write $\mathcal{P}(\boldsymbol{q}) = \{\boldsymbol{p}\in[L,U]^n : A(\boldsymbol{\alpha}+B\boldsymbol{p})\le\boldsymbol{q}\}$.

Lemma 3 (Value-map regularity lemma)**.** *Under Definition 6 and Assumption 2, use* $\theta=(\boldsymbol{\alpha},B)$ *and the maps just defined.* r_θ *is a quadratic function on* $[L,U]^n$ *with constant Hessian* $B+B^\top\prec 0$. *The demand-coordinate value map* R_θ *is a quadratic function on* $f_\theta([L,U]^n)$ *with constant Hessian* $M_d := B^{-1}+B^{-\top} = B^{-\top}(B+B^\top)B^{-1}\prec 0$. *Consequently* R_θ *has* $\|M_d\|_2$*-Lipschitz gradient and second-order growth modulus* $\kappa_d = -\lambda_{\max}(M_d)/2 > 0$ *on the whole demand image: for all* $\boldsymbol{d},\boldsymbol{d}'$ *in that image,* $R_\theta(\boldsymbol{d}')\le R_\theta(\boldsymbol{d})+\nabla R_\theta(\boldsymbol{d})^\top(\boldsymbol{d}'-\boldsymbol{d})-\kappa_d\|\boldsymbol{d}'-\boldsymbol{d}\|_2^2$. *The quadratic extension satisfies the same inequality on* $\mathbb{R}^n$. *If* d^* *maximizes it on a convex set* K, *its variational inequality therefore gives* $R_\theta(d)-R_\theta(d^*)\le-\kappa_d\|d-d^*\|^2$ *for every* $d\in K$.

Proof. Differentiation gives $\nabla^2 r_\theta = B+B^\top\prec 0$, which also makes B nonsingular. Substituting $p = B^{-1}(d-\alpha)$ gives $R_\theta = d^\top B^{-1}(d-\alpha)$ and $M_d = B^{-1}+B^{-\top} = B^{-\top}(B+B^\top)B^{-1}\prec 0$. Its constant Hessian gives both Lipschitz continuity and the exact second-order inequality. The maximizing variational inequality makes the linear term nonpositive. □

C.2 Proof of Theorem 1

We first record the accounting identity behind the residual correction. For analysis, continue the recursion in (4) after a switch and define

$$\boldsymbol{p}_t^{\mathrm{v}} = B^{-1}(\boldsymbol{q}_t-\boldsymbol{\alpha}),\qquad \boldsymbol{D}_t^{\mathrm{v}} = \boldsymbol{q}_t+\boldsymbol{\xi}_t. \tag{29}$$

After a switch these are virtual quantities, possibly outside the admissible price and request sets.

Lemma 4 (Residual bridge and pre-cutoff feasibility)**.** *Use Assumption 4(a),* $A\ge 0$, *Definition 1 and Algorithm 1, with nonsingular* B, $\mathcal{D}=\alpha+B[L,U]^n\subseteq\mathbb{R}^n_+$ *and fluid-feasible* d_T^*. *For the virtual definitions (29), every integer* $1\le t<T$ *satisfies*

$$\sum_{v=1}^{t}(D_v^{\mathrm{v}}-d_T^*) = (T-t)\sum_{s=1}^{t}\boldsymbol{\xi}_s/(T-s). \tag{30}$$

If $1 \le t \le K_T$, $t+1 \le T$, and $q_1, \dots, q_{t+1} \in \mathcal{D}$, the actual policy has not switched, every request through t is served, and

$$C_T - \sum_{v=1}^{t} A\boldsymbol{d}^v = (C_T - TAd_T^*) + (T-t)Aq_{t+1} \ge 0. \tag{31}$$

Terminal fallback after K_T is prescribed and is distinct from an image-exit switch.

Proof. In the sum each ξ_s has coefficient $1 - (t-s)/(T-s) = (T-t)/(T-s)$. This gives the bridge and the capacity identity first for virtual demands. Image membership makes posted prices admissible and actual requests equal $q_s + \xi_s$. Inductively, the identity at s gives $c^s - A\boldsymbol{d}^s \ge 0$, so work conservation gives full service. Expectations are unconditional with respect to this future event. □

Lemma 5 (Expected virtual revenue loss)**.** *Let $R(\boldsymbol{d}) = \boldsymbol{d}^\top B^{-1}(\boldsymbol{d} - \boldsymbol{\alpha})$, let $\boldsymbol{d}_T^*$ solve (3), and let $\boldsymbol{q}_t$ follow (4). Under Assumption 4(a), set $L_R = \|B^{-1} + B^{-\top}\|_2$. For every deterministic $K \le T$, $\sum_{t=1}^{K} \mathbb{E}\big[R(\boldsymbol{d}_T^*) - R(\boldsymbol{q}_t)\big] \le \frac{L_R n \sigma^2}{2} H_{T-1}$, $\quad H_{T-1} = \sum_{k=1}^{T-1} \frac{1}{k}$, $\quad H_0 := 0$.*

Proof. The exact quadratic expansion has zero expected linear term because $q_t - d_T^*$ is a martingale sum. Its expected quadratic term is at most $(L_R n \sigma^2/2) \sum_{s<t} (T-s)^{-2}$. Sum over $t \le K$ and use $\sum_{t=1}^{T} \sum_{s<t} (T-s)^{-2} = H_{T-1}$. □

Lemma 6 (Exit and fallback charge)**.** *Under Assumption 5, the probability that the policy switches before the last periods is at most T^{-6}. The combined expected revenue loss from that event and the last periods is $O(\log T)$.*

Proof. The noiseless case is immediate. Otherwise let $\mathcal{G}_T = \{\max_{t \le K_T+1} \|M_t\| \le \delta_D/2\}$. Coordinate sub-Gaussian tails and $\sum_{s<t}(T-s)^{-2} \le 1/(T-t)$ give $\Pr(\|M_t\| > \delta_D/2) \le 2ne^{-a_0(T-t)}$. Summation and the prescribed c_h give $\Pr(\mathcal{G}_T^c) \le K_0 e^{-a_0(h_T-1)} \le T^{-6}$. When $K_T = 0$ there is no preterminal switch. The local margin and bridge give full service on $\mathcal{G}_T$. For virtual earned revenue $X_t = R(q_t) + (p_t^{\mathrm{v}})^\top \xi_t$, predictability gives the unconditional identity $\mathbb{E}X_t = \mathbb{E}R(q_t)$. Sub-Gaussian fourth moments give $\sup_{T, t \le K_T} \mathbb{E}X_t^2 \le C_X$. Actual revenue Y_t equals X_t on $\mathcal{G}_T$ and is always nonnegative. We now apply the Cauchy–Schwarz inequality to bound the expected revenue loss on the failure event. Thus $\mathbb{E}\sum_{t \le K_T} Y_t \ge \mathbb{E}\sum_{t \le K_T} X_t - \sqrt{C_X}\, T \Pr(\mathcal{G}_T^c)^{1/2}$, with failure cost $O(T^{-2})$. The final window costs at most $h_T \max_{d \in \mathcal{D}} R(d) = O(\log T)$. □

Proof of Theorem 1. Because B is invertible, (2) is exactly (3) under the change of variables $\boldsymbol{d} = f(\boldsymbol{p})$. Hence $V^{\mathrm{Fluid}}(\boldsymbol{C}_T) = TR(\boldsymbol{d}_T^*)$. We now apply Lemma 5 with $K = K_T$, and combine it with Lemma 6. Since $H_{T-1} \le 1 + \log T$, all three contributions, virtual mean-revenue loss, the failure event, and the last periods, are bounded by $C_{\mathrm{FI}}(1 + \log T)$. □

D Proofs for the No-Information Setting

The proof uses the self-adjusting argument of Chen et al. (2021) with fixed-design estimation, observed requests, and explicit resource capacities.

Definition 7 (Probability convention)**.** Use the filtration, predictable prices, uncensored requests, conditional vector sub-Gaussian proxy $\sigma > 0$, and mechanical fulfillment of Section 2. Thus $\boldsymbol{\xi}_t \in \mathcal{F}^t$, $\mathbb{E}[\boldsymbol{\xi}_t \mid \mathcal{F}^{t-1}] = 0$, and every unit-direction conditional mgf is bounded by $\exp(\lambda^2\sigma^2/2)$ for every real λ. Estimation uses $\boldsymbol{d}^t$. Capacity and revenue use $\hat{\boldsymbol{d}}^t$. The resource matrix is entrywise nonnegative. Complete design cycles have $q_T \ge 1$ and $L_T = J_0 q_T \ge J_0$.

D.1 Fixed-design estimation

Lemma 7 (Finite identifying design)**.** *For a box $Q = \prod_j [\ell_j, u_j] \subseteq \mathcal{K}_{\text{fp}}$, choose $v = (\ell + u)/2$, $0 < \Delta < \min_j(u_j - \ell_j)/2$, and prices $v, v + \Delta e_j$. This gives $J_0 = n + 1$ and the valid constant $\kappa_0^{-1} = (n+1)\{(1 + \mathbf{1}^\top v/\Delta)^2 + \|v\|_2^2/\Delta^2 + 2n/\Delta^2\}$.*

Proof. If the regressor matrix Z maps (a, b) to y, then $b_j = (y_j - y_0)/\Delta$ and $a = (1 + \mathbf{1}^\top v/\Delta)y_0 - \sum_j v_j y_j/\Delta$. Hence the braces above equal $\|Z^{-1}\|_F^2$, and $Z^\top Z/(n+1) \succeq I/\{(n+1)\|Z^{-1}\|_F^2\}$. □

Lemma 8 (Fixed-design OLS)**.** *Adopt Definitions 4 and 7, the affine demand of Assumption 2(a), and the compact convex class in Assumption 6(a). Cycle the deterministic fixed design for $\ell_T = J_0 q$ periods and Frobenius-project OLS onto that class. With $d = n + 1$ and $V_{\ell_T} = \sum_{s \le \ell_T} z_s z_s^\top \succeq \kappa_0 \ell_T I_d$, one may take $C_{\text{ols}} = 5$ so that $\mathbb{E}\|\widehat{\Theta}_{\ell_T} - \Theta\|_F^2 \le \frac{C_{\text{ols}} n d \sigma^2}{\kappa_0 \ell_T}$, $\Pr\left\{\|\widehat{\Theta}_{\ell_T} - \Theta\|_F > C_{\text{ols}}\sigma\sqrt{\frac{n\{d+\log(n/\delta)\}}{\kappa_0 \ell_T}}\right\} \le \delta$ for $0 < \delta < 1$.*

Proof. Write $\ell := \ell_T$, $V := V_\ell$, and let $e_j := \widetilde{\theta}_{\ell,j} - \theta_j \in \mathbb{R}^d$ denote the transpose of the jth row error. By (20) and (21), $e_j = V^{-1} \sum_{s=1}^{\ell} z_s \xi_{s,j}$. The deterministic design makes $z_s \in \mathcal{F}^{s-1}$. For $u \in \mathbb{S}^{d-1}$ set $b_s = u^\top V^{-1/2} z_s$ and $S_r = \sum_{s \le r} b_s \xi_{s,j}$. Then $\sum_s b_s^2 = 1$ and the tower property gives

$$\mathbb{E}e^{\lambda S_r} \le e^{\lambda^2 b_r^2 \sigma^2/2} \mathbb{E}e^{\lambda S_{r-1}}, \qquad \mathbb{E}e^{\lambda u^\top V^{1/2} e_j} \le e^{\lambda^2 \sigma^2/2}.$$

These are the uncensored observations $d^s = \Theta z_s + \xi_s$. OLS uses only uncensored requests. Since $\cosh(ax) \ge 1 + a^2x^2/2$ and the same mgf bounds $\mathbb{E}[\cosh(a\xi_{s,j}) \mid \mathcal{F}^{s-1}]$, letting $a \downarrow 0$ gives $\mathbb{E}[\xi_{s,j}^2 \mid \mathcal{F}^{s-1}] \le \sigma^2$.

We now prove the tail bound. A maximal $1/2$-separated subset $\mathcal{N}$ of $\mathbb{S}^{d-1}$ has $|\mathcal{N}| \le 5^d$ because its radius-$1/4$ balls are disjoint in the radius-$5/4$ ball, and $\|w\|_2 \le 2\max_{u \in \mathcal{N}} |u^\top w|$. Markov's inequality applied to $e^{\theta Y}$ and $e^{-\theta Y}$ with $\theta = \sqrt{2x}/\sigma$ (the Chernoff bound) gives $\Pr\{|Y| > \sigma\sqrt{2x}\} \le 2e^{-x}$. With $x = d\log 5 + \log(2n/\delta)$, the net and row unions give $\Pr\{\|\widetilde{\Theta}_\ell - \Theta\|_F > 2\sigma\sqrt{2nx/(\kappa_0 \ell)}\} \le \delta$. For $r = \log(n/\delta) > 0$, $d = n + 1 \ge 2$ and $d\log 5 + \log 2 + r \le 2(d + r)$ (as $e^4 > 50$), this is the stated tail with $C_{\text{ols}} = 5$.

We now prove the moment bound. If $s < t$, then $\mathbb{E}[\xi_{s,j}\xi_{t,j}] = 0$. Thus, for $S_j = \sum_s z_s \xi_{s,j}$, $\mathbb{E}[S_j S_j^\top] \preceq \sigma^2 V$ and $\mathbb{E}\|e_j\|_2^2 \le \sigma^2 \operatorname{tr}(V^{-1}) \le d\sigma^2/(\kappa_0 \ell)$; summing rows gives the moment bound. Metric projection onto the closed convex class is well defined and nonexpansive: the variational inequality $\langle X - \Pi X, Y - \Pi X\rangle_F \le 0$ for Y in the class, added for X and X', gives $\|\Pi X - \Pi X'\|_F \le \|X - X'\|_F$. Being continuous, it is Borel. Hence $\Pi_{\mathfrak{T}}(\Theta) = \Theta$ transfers both bounds to $\widehat{\Theta}_\ell$. □

Definition 8 (True fluid comparator)**.** Write $\mathcal{D}_\theta = \alpha + B[L, U]^n$, $R_\theta(d) = d^\top B^{-1}(d - \alpha)$ and $W_\theta(q) = \max\{R_\theta(d) : d \in \mathcal{D}_\theta, Ad \le q\}$. Set $c_0 = C_T/T$, denote its optimizer by $d_0 = d_T^*$ and its price by $p_T^* = B^{-1}(d_0 - \alpha)$, and use $V^{\text{Fluid}}(C_T) = TW_\theta(c_0)$.

Definition 9 (Common proof setting)**.** Use the primitive hypotheses of Theorem 2, Definition 1, and Definitions 7–8. Fix the number of resources, products, design points and design constants, and $q_T = \lceil c_L \sqrt{T}/J_0 \rceil$, $\ell_T = L_T = J_0 q_T$. When $\ell_T < T$, put $N = T - \ell_T$, $C^+ = c^{\ell_T + 1}$, $q^+ = C^+/N$ and $e_{\ell_T} = \|\widehat{\Theta}_{\ell_T} - \Theta\|_F$, $\Delta_{\ell_T} := \|q^+ - c_0\|_\infty$. All constants below depend only on the fixed compact class, price set, the margin condition, design/noise and resource primitives. A deterministic T_0 may be enlarged to cover the explicitly used localization and terminal thresholds.

Lemma 9 (Stability of the limiting fluid solution)**.** *Adopt Definition 4 and Assumptions 1, 2, 3, 5 and 6(a). Thus $\mathcal{P} = [L, U]^n$ is a nonempty compact convex closed box, $A \ge 0$, and $f(p) = \alpha + Bp$. For $\vartheta = (\alpha, B)$ and $\eta = (\alpha_\eta, B_\eta)$ write*

$$\ell_\eta(v) := v^\top B_\eta^{-1}(v - \alpha_\eta), \qquad \mathcal{D}_\eta := \alpha_\eta + B_\eta \mathcal{P}, \qquad K_\eta(c) := \{v \in \mathcal{D}_\eta : Av \le c\}.$$

Every $\mathcal{D}_\eta$ is compact and convex and every nonempty $K_\eta(c)$ is compact and convex. There are primitive constants $r_*, c_*, r_{\mathrm{img}}, L_*, T_0 > 0$ and

$$s_* := \tfrac{1}{2}\min\{s_{\min}, \lambda_{\min}, a_{\min}\} \quad \text{if } \bar{J} \neq \varnothing, \qquad s_* := s_{\min}/2 \quad \text{if } \bar{J} = \varnothing,$$

such that the following holds for every $T \geq T_0$. Put $c_0 = C_T/T$ and let $x^* = x_{\vartheta,c_0} = d_T^*$ be the true maximizer in $K_\vartheta(c_0)$. It exists by compactness and fluid feasibility, and λ^* denotes its multiplier. Multipliers are vectors in $\mathbb{R}_+^m$ whose components outside the binding set are zero. If η belongs to the learning class of Assumption 6(a), $\|\eta - \vartheta\|_F \leq r_*$ and $\|b - c_0\|_\infty \leq c_*$, then the fitted problem $\max\{\ell_\eta(v) : v \in K_\eta(b)\}$ has a unique maximizer $x_{\eta,b}$ and multiplier $\lambda_{\eta,b}$, its binding set is $J = \bar{J}$, and

$$A_J x_{\eta,b} = b_J, \qquad A_{J^c} x_{\eta,b} < b_{J^c}, \qquad \lambda_{\eta,b,J} \geq s_* \mathbf{1} \quad (J \neq \varnothing), \tag{32}$$

$$\|x_{\eta,b} - x^*\|_2 + \|\lambda_{\eta,b} - \lambda^*\|_2 \leq L_*\{\|\eta - \vartheta\|_F + \|b - c_0\|_\infty\}. \tag{33}$$

The true and fitted images contain the same ball $\mathbb{B}_2(x^*, r_{\mathrm{img}})$, and both inverse-price maps send that ball into $\mathcal{P}$. If $J = \varnothing$, the multiplier is zero and the active-rank and multiplier clauses are vacuous.

Proof. Put $Q_\eta := B_\eta^{-1} + B_\eta^{-\top}$ and $r_\eta := B_\eta^{-1}\alpha_\eta$. On a fixed neighborhood of ϑ, compactness of the learning class and its inverse bound give primitive $\mu, M, L_H, L_r > 0$ with $-Q_\eta \succeq \mu I$, $\|Q_\eta\| \leq M$, $\|Q_\eta - Q_\vartheta\| \leq L_H\|\eta - \vartheta\|_F$ and $\|r_\eta - r_\vartheta\| \leq L_r\|\eta - \vartheta\|_F$. At $(\vartheta, \bar{c})$, fluid feasibility and compactness of $\mathcal{P}$ give the maximizer $\bar{x}^*$ from Assumption 3.

Suppose first that $\bar{J} \neq \varnothing$. Write $A_J = A_{\bar{J}}$, $a_J = \|A_J\|_2$, $\sigma_J = \sigma_{\min}(A_J)$, and use $\nu = -\lambda_J$. In particular $\bar{\nu} := -\bar{\lambda}_{\bar{J}}$ and $Q := Q_\vartheta$. The restricted KKT system is $K\binom{x}{\nu} = \binom{r_\vartheta}{\bar{c}_J}$ with $K = [Q_\vartheta \; A_J^\top; A_J \; 0]$. For $S = A_J Q_\vartheta^{-1} A_J^\top$, S is negative definite and $\|S^{-1}\| \leq M/\sigma_J^2$. The explicit block inverse is

$$K^{-1} = \begin{pmatrix} Q^{-1} - Q^{-1}A_J^\top S^{-1} A_J Q^{-1} & Q^{-1}A_J^\top S^{-1} \\ S^{-1}A_J Q^{-1} & -S^{-1} \end{pmatrix}, \qquad \|K^{-1}\| \leq C_K := \mu^{-1} + \frac{M}{\sigma_J^2}\left(1 + \frac{a_J}{\mu}\right)^2.$$

(The bound is the sum of the four block norms.) If $C_K L_H \|\eta - \vartheta\|_F \leq 1/2$, the Neumann resolvent bound gives $\|K_\eta^{-1}\| \leq 2C_K$, where $K_\eta = [Q_\eta \; A_J^\top; A_J \; 0]$. With $R_0 := \|(\bar{x}^*, \bar{\nu})\|_2$ and $\|b_J - \bar{c}_J\|_2 \leq \sqrt{|J|}\|b - \bar{c}\|_\infty$, we obtain

$$\|(x', \nu') - (\bar{x}^*, \bar{\nu})\|_2 \leq L_{\mathrm{sol}}(\|\eta - \vartheta\|_F + \|b - \bar{c}\|_\infty), \quad L_{\mathrm{sol}} := 2C_K(L_r + L_H R_0 + \sqrt{|J|}).$$

Set $\lambda' = -\nu'$ and $\bar{\lambda} = -\bar{\nu}$. These are the capacity multipliers with the maximizing-KKT sign convention.

Let $M_A := \max_i \|A_i\|_2$ and $C_p := \frac{2}{\kappa_B}(L_{\mathrm{sol}} + 1 + \|\bar{p}^*\|_2)$. Choose a radius $r_{\mathrm{bar}} > 0$ below $\{(2C_K L_H)^{-1}, \delta_P/(2C_p), s_{\min}/[2(1 + M_A L_{\mathrm{sol}})], \lambda_{\min}/(2L_{\mathrm{sol}})\}$. We choose this radius to keep the inverse price feasible and the inactive slacks and active multipliers positive. For $h := \|\eta - \vartheta\|_F + \|b - \bar{c}\|_\infty \leq r_{\mathrm{bar}}$, the displayed bound gives $\|x' - \bar{x}^*\|_2 \leq L_{\mathrm{sol}} h$ and

$$\|B_\eta^{-1}(x' - \alpha_\eta) - \bar{p}^*\|_2 \leq \frac{2}{\kappa_B}\{\|x' - \bar{x}^*\|_2 + \|\eta - \vartheta\|_F(1 + \|\bar{p}^*\|_2)\} \leq C_p h \leq \delta_P/2.$$

Thus the restricted solution lies in $\mathcal{D}_\eta$ and the perturbed set is nonempty. For every inactive row,

$$b_i - A_i x' \geq s_{\min} - |b_i - \bar{c}_i| - \|A_i\|_2\|x' - \bar{x}^*\|_2 \geq s_{\min} - (1 + M_A L_{\mathrm{sol}})h \geq s_{\min}/2,$$

while $\|\lambda'_J - \bar{\lambda}_J\|_\infty \leq L_{\mathrm{sol}} h$ so $\lambda'_J \geq \lambda_{\min}/2$. Thus the restricted point satisfies the full KKT conditions on $K_\eta(b)$. Compact convexity and strict concavity give global optimality and uniqueness.

If $\bar{J} = \varnothing$, then $Q_\vartheta \bar{x}^* = r_\vartheta$ and all resource rows have slack at least $s_{\min}$. The restricted solution is the unconstrained stationary point $x' = Q_\eta^{-1} r_\eta$, with multiplier zero, and the same resolvent calculation

gives the displayed bound with $L_{\mathrm{sol}} = 2\mu^{-1}(L_r + L_H\|\bar{x}^*\|_2)$ and no capacity norm term. In this branch put $C_{p,0} := 2(L_{\mathrm{sol}} + 1 + \|\bar{p}^*\|_2)/\kappa_B$ and choose r_{bar} also below $\delta_P/(2C_{p,0})$ and $s_{\min}/[2(1 + M_A L_{\mathrm{sol}})]$. The same inverse-price calculation then puts x' in $\mathcal{D}_\eta$, while the row-wise slack bound puts it strictly below every capacity row, so the perturbed feasible set is nonempty. Stationarity alone is the full KKT system, and strict concavity again gives the unique maximizer.

The preceding argument with $\eta = \vartheta$ and $\|c_0 - \bar{c}\|_\infty \le K_C/T$ transfers the active set, slacks, multiplier and inverse-price margin to the finite-T optimizer $d_T^* = x^*$ once $T \ge T_0$ with $K_C/T_0 \le r_{\mathrm{bar}}/2$. It also gives $\|x^* - \bar{x}^*\|_2 \le L_{\mathrm{sol}} K_C/T$. Now take $\|\eta - \vartheta\|_F \le r_*$ and $\|b - c_0\|_\infty \le c_*$ with $r_* = c_* = r_{\mathrm{bar}}/4$. Since $\|c_0 - \bar{c}\|_\infty \le K_C/T \le r_{\mathrm{bar}}/2$, both (η, b) and (ϑ, c_0) lie within r_{bar} of $(\vartheta, \bar{c})$, so the preceding paragraphs apply to both: each problem has binding set $\bar{J}$, and its solution is the restricted KKT point, $K_\eta z_{\eta,b} = (r_\eta, b_J)$ and $K_\vartheta z^* = (r_\vartheta, c_{0,J})$ with $z = (x, \nu)$. Subtracting,

$$K_\eta(z_{\eta,b} - z^*) = (r_\eta - r_\vartheta,\ b_J - c_{0,J}) - (K_\eta - K_\vartheta)z^*,$$

and $\|K_\eta - K_\vartheta\| = \|Q_\eta - Q_\vartheta\| \le L_H\|\eta - \vartheta\|_F$. With $\|K_\eta^{-1}\| \le 2C_K$ and $\|z^*\|_2 \le R_1 := R_0 + L_{\mathrm{sol}} r_{\mathrm{bar}}/2$,

$$\|z_{\eta,b} - z^*\|_2 \le 2C_K\{(L_r + L_H R_1)\|\eta - \vartheta\|_F + \sqrt{|J|}\,\|b - c_0\|_\infty\}.$$

The full multipliers agree with $-\nu$ on $\bar{J}$ and vanish elsewhere, so $\|\lambda_{\eta,b} - \lambda^*\|_2 = \|\nu_{\eta,b} - \nu^*\|_2$, and $\|x_{\eta,b} - x^*\|_2 + \|\lambda_{\eta,b} - \lambda^*\|_2 \le \sqrt{2}\,\|z_{\eta,b} - z^*\|_2$. Hence (33) holds with $L_* := 2\sqrt{2}\,C_K(L_r + L_H R_1 + \sqrt{|J|})$. If $\bar{J} = \varnothing$, the same subtraction with $z = x$, $K = Q$ and $\|Q_\eta^{-1}\| \le 2\mu^{-1}$ gives $L_* := 2\sqrt{2}\,\mu^{-1}(L_r + L_H R_1)$ with $R_1 := \|\bar{x}^*\|_2 + L_{\mathrm{sol}} r_{\mathrm{bar}}/2$, and both multipliers are zero. The only use of fluid scaling is the explicit T_0 choice above.

Finally choose $r_{\mathrm{img}} > 0$ so that $r_{\mathrm{img}} < \delta_D/4$ and

$$\frac{2}{\kappa_B}\big(r_{\mathrm{img}} + L_{\mathrm{sol}} K_C/T_0 + r_*(1 + \|\bar{p}^*\|_2)\big) \le \delta_P.$$

For $\|y - x^*\|_2 \le r_{\mathrm{img}}$, the preceding inverse-price bound puts both inverse prices in $\mathcal{P}$, while the distance bound and $r_{\mathrm{img}} < \delta_D/4$ put y in both images. □

Lemma 10 (Post-exploration feasibility)**.** *Use the common setting, put $\delta_T := T^{-3}$ for $T \ge 3$, and let*

$$N := T - \ell_T, \qquad q^+ := c^{\ell_T + 1}/N, \qquad \Delta_{\ell_T} := \|q^+ - c_0\|_\infty,$$
$$r_s := C_{\mathrm{conf}}\sqrt{\frac{\log T}{\ell_T}}, \quad C_{\mathrm{conf}} := 5\sigma\sqrt{\frac{n(n + 1 + \log n + 3)}{\kappa_0}}.$$

Instantiate the geometry gate at freeze time $s := \ell_T$ with $h := h_T$, $s_J := s_*$, $r_0 := r_{\mathrm{img}}/2$, $C_\Delta = C_x := L_*$, $C_g := 2L_*$, $C_a := C_f$, $c_\Delta := \min\{c_*, 1\}$ and $r_{\mathrm{loc}} := r_*$, and put $\|A\|_{\infty,2} := \max_i \|A_i\|_2$, $C_A := 2\sigma\|A\|_{\infty,2}$,

$$C_\Delta^{\mathrm{load}} := 2\max\{\|\bar{c}\|_\infty + K_C + M_A, C_A\}, \qquad M_A := \max_i \sup_{p \in \mathcal{P}} (Af(p))_i,$$

where C_Δ^{load} is finite, deterministic and independent of T, and define

$$Q_{\ell_T} := \{e_{\ell_T} \le r_s\} \cap \left\{\left\|A\sum_{u \le \ell_T}(\boldsymbol{d}^u - f(\boldsymbol{p}^u))\right\|_\infty \le C_A\sqrt{\ell_T \log(m/\delta_T)}\right\}. \tag{34}$$

There is a primitive $T_1 \ge T_0$ such that for $T \ge T_1$, $\Pr(Q_{\ell_T}^c) \le 2\delta_T$, and on Q_{ℓ_T}:

(i) every exploration request is fully served and

$$\Delta_{\ell_T} \le C_\Delta^{\mathrm{load}} \frac{\ell_T + \sqrt{\ell_T \log(m/\delta_T)}}{T}; \tag{35}$$

(ii) the true and fitted programs at q^+ are nonempty, with $d_0 := x_{\vartheta,c_0} = x^*$, $v := x_{\vartheta,q^+}$, $x := x_{\widehat{\vartheta},q^+}$, and $J = \bar{J}$;

(iii) every clause of the geometry gate (22) holds, with the constants supplied by Lemma 9, and $\mathbb{B}_2(d_0, r_{\mathrm{img}})$ is common to both images with inverse prices in $\mathcal{P}$.

The constant C_Δ^{load} is primitive and independent of T. The schedule definition (6) gives $\ell_T = J_0\lceil c_L\sqrt{T}/J_0\rceil = c_L\sqrt{T} + O(1) = o(T)$ and $h_T = \lceil c_h \ell_T\rceil$. Hence the logarithmic gate holds.

Proof. Write $L := \ell_T$ and $\delta_T := T^{-3}$. The deterministic design and schedule make Q_L $\mathcal{F}^L$-measurable. Lemma 8 gives $\Pr\{e_L > r_s\} \le \delta_T$. For every nonzero row A_i, iterated conditional mgfs and Chernoff give

$$\mathbb{E}e^{\lambda A_i^\top \sum_{u=1}^L \xi_u} \le e^{\lambda^2\sigma^2\|A_i\|_2^2 L/2}, \qquad \Pr\left\{\left\|A\sum_{u=1}^L \xi_u\right\|_\infty > C_A\sqrt{L\log(m/\delta_T)}\right\} \le \delta_T.$$

Each nonzero row contributes at most $2(\delta_T/m)^2$, whose union is $2\delta_T^2/m \le \delta_T$ for $T \ge 3$. Zero rows vanish. Thus $\Pr(Q_L^c) \le 2\delta_T$.

Let $\bar{c}_{\min} := \min_i \bar{c}_i > 0$. For a deterministic T_1 large enough, $C_{T,i} \ge T\bar{c}_{\min}/2$ and $M_A L + C_A\sqrt{L\log(m/\delta_T)} \le T\bar{c}_{\min}/4$. Because $d^u \ge 0$ and $A \ge 0$, prefix loads are nondecreasing, so every exploration request is fully served. Consequently $q^+ - c_0 = [Lc_0 - A\sum_{u=1}^L f(p^u) - A\sum_{u=1}^L \xi_u]/N$ and, since $N \ge T/2$ and $\|c_0\|_\infty \le \|\bar{c}\|_\infty + K_C/T$,

$$\Delta_L \le \frac{2}{T}\left[(\|\bar{c}\|_\infty + K_C + M_A)L + C_A\sqrt{L\log(m/\delta_T)}\right] \le C_\Delta^{\mathrm{load}}\frac{L + \sqrt{L\log(m/\delta_T)}}{T}.$$

The schedule gives $r_s = O(T^{-1/4}\sqrt{\log T})$ and $\Delta_L = O(T^{-1/2} + T^{-3/4}\sqrt{\log T})$. For sufficiently large T_1, Lemma 9 applies at (ϑ, c_0), (ϑ, q^+) and $(\widehat{\vartheta}, q^+)$, giving $d_0 = x^*$, $\|v - d_0\|_2 \le L_*\Delta_L$, $\|x - v\|_2 \le 2L_*(e_L + \sqrt{\Delta_L})$ and $\|x - x^*\|_2 \le L_*(e_L + \sqrt{\Delta_L})$; it also gives nonemptiness, the common image ball, binding-set identity, strict inactive slacks, multiplier bounds and active-rank clauses. Since $x \in K_{\widehat{\vartheta}}(q^+)$, $c^{L+1} - NAx = N(q^+ - Ax) \ge 0$, and

$$\|g(x) - x\|_2 = \|f(\widehat{p}(x)) - \widehat{f}(\widehat{p}(x))\|_2 \le C_f e_L \le C_a(e_L + \sqrt{\Delta_L}).$$

For positive rows let $a_A := \min_{i:A_i\mathbf{1}>0} A_i\mathbf{1}$ and $\tau_b := \Delta_L/a_A$, with $\tau_b = 0$ if all rows are zero. Enlarging T_1 gives $2L_*(e_L + \sqrt{\Delta_L}) + \sqrt{n}\,\tau_b \le r_0/2$, $\|x - x^*\|_2 \le r_0/4$, $\Delta_L \le c_\Delta$ and $e_L \le r_{\mathrm{loc}}$. We now check the logarithmic geometry-gate condition:

$$4\log(N/h_T)\max\{C_H r_s, C_a(e_L + \sqrt{\Delta_L})\} = O(T^{-1/4}(\log T)^{3/2}) = o(1),$$

so the logarithmic geometry-gate condition holds. All conclusions are pathwise on Q_L. □

D.2 Main-text results

Proof of Lemma 1. Fix T and write $d := n + 1$, $s := \ell_T$, $N := T - s$, $h := h_T$ and $\mathcal{G}_t := \mathcal{H} \vee \mathcal{F}^t$. The design, schedule, A, C_T, Θ, $\mathfrak{T}$ and κ_0 are deterministic. Under the enlarged filtration, we have

$$\mathbb{E}[\xi_t \mid \mathcal{G}_{t-1}] = 0, \qquad \mathbb{E}[e^{\lambda v^\top \xi_t} \mid \mathcal{G}_{t-1}] \le e^{\lambda^2\sigma^2/2}$$

for every unit v and real λ. Bounded prices, affine means and mechanical service make $\mathcal{R}_T := V^{\mathrm{Fluid}}(C_T) - \sum_{t=1}^{T}(p^t)^\top \widehat{d}^t$ integrable.

For $u \in \mathbb{S}^{d-1}$ and a design row j, put $a_r = u^\top V_s^{-1/2} z_r$. These deterministic coefficients satisfy $\sum_r a_r^2 = 1$. Iterating the enlarged-filtration mgf by the tower property gives

$$\mathbb{E}\left[\exp\left\{\lambda \sum_{r=1}^{s} a_r \xi_{r,j} - \lambda^2\sigma^2/2\right\}\middle|\mathcal{H}\right] \le 1.$$

The net calculation and row union bound in the proof of Lemma 8 therefore give, almost surely,

$$\Pr\left\{e_s > 5\sigma\sqrt{\frac{n\{d + \log(n/\delta)\}}{\kappa_0 s}}\middle|\mathcal{H}\right\} \le \delta.$$

The conditional second-moment implication is $\mathbb{E}[\xi_{r,j}^2 \mid \mathcal{G}_{r-1}] \le \sigma^2$. If $r < q$, $\xi_{r,j} \in \mathcal{G}_{q-1}$, so conditional orthogonality gives $\mathbb{E}[e_s^2 \mid \mathcal{H}] \le C_e/s$ for a history-independent C_e.

Put $\delta_T = T^{-3}$ and let $\mathcal{Q}_s$ be (34). The preceding conditional OLS tail gives $\Pr(e_s > r_s \mid \mathcal{H}) \le \delta_T$. For each nonzero resource row, conditional Chernoff gives

$$\Pr\left\{\left|A_i^\top \sum_{r=1}^{s} \xi_r\right| > C_A\sqrt{s\log(m/\delta_T)}\middle|\mathcal{H}\right\} \le 2(\delta_T/m)^2.$$

Zero rows vanish, so $\Pr(\mathcal{Q}_s^c \mid \mathcal{H}) \le 2\delta_T$ a.s. These concentration bounds apply before restriction to $\mathcal{Q}_s$.

By Lemma 10, on $\mathcal{Q}_s$ exploration is fully served and the capacity/program, common-ball, strict-slack and binding-set conditions hold, with $\Delta_s \le \bar{\Delta}_T := C_\Delta^{\mathrm{load}}\{s + \sqrt{s\log(m/\delta_T)}\}/T$. Algorithm 2 is covered by that lemma. After exploration it solves the fitted program (8), whose capacity constraint is $A\boldsymbol{d} \le \boldsymbol{C}^+/N$ with $\boldsymbol{C}^+ = c^{s+1}$, that is, $b = c^{s+1}/N$. Thus its starting target is the fitted maximizer $x = \widehat{d}^*$. Its target is defined by (9): $q_t = \widehat{d}^* - \sum_{r=s+1}^{t-1} \widehat{\Delta}^r/(T-r)$ with $\widehat{\Delta}^r = D_r - (\widehat{\alpha}_s + \widehat{B}_s p^r)$. Hence $q_{t+1} = q_t - \widehat{\Delta}^t/(T-t)$. While active it posts $p^t = \widehat{p}(q_t) = \widehat{B}_s^{-1}(q_t - \widehat{\alpha}_s)$, so $\widehat{\alpha}_s + \widehat{B}_s p^t = q_t$, its residual is $\widehat{\Delta}^t = D_t - q_t$ and its target obeys $q_{t+1} = q_t - (D_t - q_t)/(T-t)$, which is (25). The algorithm is active whenever $t \le T - h$ and $q_t \in \widehat{\mathcal{D}}$, and $q_t \in \widehat{\mathcal{D}}$ implies $\widehat{p}(q_t) \in \mathcal{P}$. Hence, in every period before the first failure of the active test (24), the algorithm posts the same price as the lemma's controller. From that period on it posts either $\widehat{p}(q_t) \in \mathcal{P}$ or the fallback $p^{\mathrm{F}} \in \mathcal{P}$, both $\mathcal{G}_{t-1}$-measurable, which is the class of controllers the lemma covers. Apply Lemma 14 at s with fallback p^{F}, estimator $\widehat{\Theta}_s$ and $\mathcal{B}_s = \mathcal{Q}_s$. The lemma's virtual recursion is driven by the actual innovations ξ_t, which satisfy the centering and mgf bounds above given $\mathcal{G}_{t-1}$. Virtual and actual prices and requests coincide only before the first active failure. The request-law claim for virtual quantities applies only on that interval. The first active failure is a bounded stopping time and optional sampling is valid because the stopped martingale has second moment at most $V\sum_{r=1}^{T-1}(T-r)^{-2} < \infty$. Thus each innovation is conditioned on its preceding history, and

$$\begin{aligned}\mathbf{1}_{\mathcal{Q}_s}\mathbb{E}[\mathcal{R}_T \mid \mathcal{G}_s] \le \mathbf{1}_{\mathcal{Q}_s}\Bigg[&C_1\{s + h + \log(N/h)\} + C_\lambda N\Delta_s \\ &+ C_2 N(e_s + \sqrt{\Delta_s})^2 + C_3 h(e_s + \sqrt{\Delta_s})\sum_{k=h}^{N-1} k^{-1} + C_{\mathrm{exit}} N\sqrt{\pi_{\mathrm{ex},s}}\Bigg].\end{aligned}$$

Lemma 14 gives, with

$$\bar{\pi}_T := \begin{cases}\min\{1, 2nN\exp[-r_0^2 h/(1024 n\sigma^2)]\}, & \sigma > 0,\\ 0, & \sigma = 0,\end{cases}$$

that $\pi_{\mathrm{ex},s} \le \bar{\pi}_T$. Since $V^{\mathrm{Fluid}}(C_T) \le T\bar{R}$ and revenue from sales is nonnegative, $\mathcal{R}_T \le T\bar{R}$ pathwise. The tower property and $Q_s \in \mathcal{G}_s$ therefore give

$$\mathbb{E}[\mathcal{R}_T \mid \mathcal{H}] \le \mathbb{E}[\mathbf{1}_{Q_s}\mathbb{E}[\mathcal{R}_T \mid \mathcal{G}_s] \mid \mathcal{H}] + 2T\bar{R}\delta_T.$$

Using $\mathbb{E}[e_s^2 \mid \mathcal{H}] \le C_e/s$, Cauchy–Schwarz, $\Delta_s \le \bar{\Delta}_T$ on Q_s, and $\sum_{k=h}^{N-1} k^{-1} \le 2\log(N/h)$ gives

$$\begin{aligned}\mathbb{E}[\mathcal{R}_T \mid \mathcal{H}] \le{}& C_1\{s + h + \log(N/h)\} + C_\lambda N\bar{\Delta}_T \\ &+ C_2 N\left(\frac{2C_e}{s} + 2\bar{\Delta}_T\right) + 2C_3 h\sqrt{\frac{2C_e}{s} + 2\bar{\Delta}_T}\log(N/h) \\ &+ C_{\mathrm{exit}} N\sqrt{\bar{\pi}_T} + 2T\bar{R}\delta_T.\end{aligned}$$

With $s = c_L\sqrt{T} + O(1)$, $h = \lceil c_h s\rceil$, $N\bar{\Delta}_T = O(s)$, $N/s = O(T/s)$, $h\sqrt{s^{-1} + \bar{\Delta}_T}\log(N/h) = O(hs^{-1/2}\log(N/h))$ and $N\sqrt{\bar{\pi}_T} = o(1)$, this is at most

$$C\left\{s + \frac{T}{s} + h + \frac{h}{\sqrt{s}}\log\frac{N}{h} + \log\frac{N}{h}\right\} + o(1) \le C_{\mathrm{NI}}\sqrt{T}.$$

Complete fallback covers $s + h \ge T$ and finitely many smaller horizons after enlarging C_{NI}. Taking the complement of the union of the countably many integer-horizon null sets gives one full-measure set, and the left side is exactly (10). □

Proof of Lemma 2. Take $\delta = T^{-1/4}/16$ and $f_\eta(p) = 1/2 - p/4 + \eta(p-1)$, $\eta = \pm\delta$. These are Proposition 5's Bernoulli instances, box $[1/2, 3/2]$, capacity T and common anchor $(1, 1/4)$. A no-information algorithm ignores that anchor. Every request is served and the benchmark agrees. The proposition gives $\sqrt{T}/4096$ for every integer horizon. □

Proof of Theorem 2. Let π be Algorithm 2. Its posted prices and its fallback p^{F} lie in $[L, U]^n$. By definition, $\mathrm{Reg}_T^{\mathrm{fl}}(\pi) = V^{\mathrm{Fluid}}(C_T) - \mathbb{E}\sum_{t=1}^T (\boldsymbol{p}^t)^\top \hat{\boldsymbol{d}}^t$. Apply Lemma 1 with the trivial sigma-field $\mathcal{H} = \{\varnothing, \Omega\}$. Then $\mathcal{G}_t = \mathcal{F}^t$, so the conditional form of Assumption 4(a) is Assumption 4(a) itself, and the remaining hypotheses of the lemma are those of the theorem. The conditional expectation in (10) is then the unconditional one, so $\mathrm{Reg}_T^{\mathrm{fl}}(\pi) \le C_{\mathrm{NI}}\sqrt{T}$ for every integer horizon. □

E Informed controller and proof

This appendix gives the controller, finite-horizon budgets, and proofs for Theorem 3 and its two corollaries. Here $D_t = \boldsymbol{d}^t$, $\widehat{D}_t = \hat{\boldsymbol{d}}^t$, $p_t = \boldsymbol{p}^t$ and $c_t = \boldsymbol{c}^t$ are the model quantities. All requests below are potential requests. Served requests are denoted by $\widehat{D}_t = \Phi(D_t, c_t)$ as in Definition 1.

E.1 Controller and plan-centred budgets

Definition 10 (Post-selection filtration)**.** Let $\mathcal{G}_0$ be an initial sigma-field containing the selected class, plan, multiplier and pre-deployment data, put $\mathcal{F}^t := \sigma\{D_s, \widehat{D}_s, c_{s+1} : s \le t\}$, and set $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{F}^t$. Prices are $\mathcal{G}_{t-1}$-measurable. D_t, $\xi_t = D_t - f(p_t)$ and $\widehat{D}_t$ are $\mathcal{G}_t$-measurable, and Assumption 4(a) holds relative to this filtration, conditionally on $\mathcal{G}_{t-1}$.

Definition 11 (Conditional initial-fluid loss)**.** For any complete policy π adapted to this filtration, define

$$L_T(\mathcal{G}_0; \pi) := V^{\mathrm{Fluid}}(T\rho) - \mathbb{E}\left[\sum_{t=1}^T (p_t^\pi)^\top \widehat{D}_t^\pi \,\middle|\, \mathcal{G}_0\right]. \tag{36}$$

When $\mathcal{G}_0$ is trivial this is the expected signed loss $\mathrm{Reg}_T^{\mathrm{fl}}$.

The initial-capacity comparator is

$$r^{\mathrm{fl}} := \max_{v \in \mathcal{P}} \{v^\top f(v) : Af(v) \le \rho\}, \qquad V^{\mathrm{Fluid}}(T\rho) = Tr^{\mathrm{fl}}. \tag{37}$$

Definition 12 (Inputs of the informed-price algorithm)**.** Adopt Definitions 4 and 10. Fix $T, m, n \ge 1$, $\mathcal{P} = [L, U]^n$, $0 \le L < U$, nonnegative $A \in \mathbb{R}_+^{m \times n}$, $\rho \in \mathbb{R}_+^m$, $p^0 \in \mathcal{P}$ and $d^0 \in \mathbb{R}^n$, all fixed or $\mathcal{G}_0$-measurable. Before deployment, fix a class containing the truth, with known $0 < \mu \le M$. The class, plan, multipliers, radius, budgets and every controller input are $\mathcal{G}_0$-measurable:

$$\mathfrak{C} \subseteq \{(B', e') : B' = B'^\top,\ \mu I \preceq -B' \preceq MI\}, \qquad p_{B',e'}(q) = p^0 + (B')^{-1}(q - d^0 - e'). \tag{38}$$

The true mean is $f(p) = d^0 + B(p - p^0) + e$. Write $p_B = p_{B,e}$. Suppose that Assumption 4 holds with respect to the post-selection filtration: requests are pathwise bounded, and each predictable price has conditionally centered innovation with covariance at most $\sigma^2 I$, $\sigma \ge 0$, given $\mathcal{G}_{t-1}$. Service follows Definition 1 and (1).

Choose pre-deployment $Ad \le \rho$, $\lambda \ge 0$, $0 < r < \min_j d_j$, and set $u = A^\top \lambda$, $s = \lambda^\top(\rho - Ad)$. In order to ensure feasible inverse prices, verify

$$p_{B',e'}(q) \in \mathcal{P} \quad \big((B', e') \in \mathfrak{C},\ \|q - d\|_2 \le r\big). \tag{39}$$

Use $q_- := \min_j d_j - r$, $q_+ := \max_j d_j + r$, $t_0 = \lceil 8M/\mu \rceil$ and integer $h \ge \max\{t_0 + 1, \lceil D_{\max}/q_- \rceil\}$.

Definition 13 (State and stopping rule of the informed controller)**.** For a selected plan d, radius r, and terminal constant h, set

$$\begin{aligned} N_t &:= T - t + 1 \quad (1 \le t \le T), \qquad N_{T+1} := 0, \qquad y_1 := Td, \qquad q_t := y_t / N_t, \\ \tau &:= \min\left(\{t \in \{1, \dots, T\} : N_t \le h \text{ or } \|q_t - d\|_2 > r\} \cup \{T + 1\}\right), \\ I_t &:= \mathbf{1}\{t < \tau\} \quad (1 \le t \le T), \qquad I_{T+1} := 0, \qquad y_{t+1} := y_t - I_t D_t. \end{aligned} \tag{40}$$

The two stopping tests use only y_t, known before period t. After the first fallback, y_t remains fixed. If $\tau \le T$, periods $\tau, \dots, T$ form a suffix of exactly $N_\tau = T - \tau + 1$ fallback periods. If neither test fires, $\tau = T + 1$, $N_\tau = 0$, and the fallback suffix is empty. Under the theorem's required $h \ge t_0 + 1$, this convention only handles the no-fallback case, because $N_T = 1 \le h$ forces $\tau \le T$. When $T = 1$, the controller starts in fallback and no expression with $N_t - 1$ is evaluated.

Definition 14 (Active/fallback informed controller)**.** Fix a $\mathcal{G}_0$-measurable initial candidate $p_1^{\mathrm{cand}} \in \mathcal{P}$ and a predictable fallback sequence $p_t^{\mathrm{fb}} \in \mathcal{P}$. In period t, post $p_t = p_t^{\mathrm{cand}}$ if $I_t = 1$ and $p_t = p_t^{\mathrm{fb}}$ otherwise. Observe the full potential request D_t and use, on every period, the service and capacity recursion

$$\widehat{D}_t := \Phi(D_t, c_t), \qquad c_1 := T\rho, \qquad c_{t+1} := c_t - A\widehat{D}_t. \tag{41}$$

On an active period update

$$y_{t+1} := y_t - D_t. \tag{42}$$

If $I_t = 1$ and $t < T$, compute

$$p_{t+1}^{\mathrm{cand}} := \Pi_{\mathcal{P}}\{p_t + \gamma_t(D_t - q_t)\}, \qquad \gamma_t := \frac{2}{\mu}\left(\frac{1}{t + t_0} + \frac{1}{N_t - 1}\right). \tag{43}$$

Here $\Pi_{\mathcal{P}}$ is Euclidean projection onto $\mathcal{P}$. The candidate is posted only if $I_{t+1} = 1$. Otherwise, the predictable fixed price is posted. Because an active period has $N_t > h \ge t_0 + 1$, $N_t - 1 \ge 1$ in every evaluation of (43). At fallback the auxiliary state y freezes and its price update stops, but the fixed rule Φ and physical capacity recursion (41) continue through period T.

Use the selected plan d, radius r, and constants q_-, q_+, t_0, h specified in Theorem 3.
Under Assumption 4(b), define

$$H_0 := 0, \qquad H_k := \sum_{j=1}^{k} \frac{1}{j} \quad (k \in \mathbb{Z}_{\geq 1}), \qquad V := n\sigma^2, \qquad R_{\max} := \bar{R} \leq nUD_{\max}. \tag{44}$$

The remaining plan-centred budgets are

$$\begin{aligned} K_d &:= \max\{n(U-L)^2(1+t_0), 9V/\mu^2\}, \\ E^{\mathrm{tr}}_{T,d} &:= 2K_d H_T, & A^{\mathrm{tr}}_{T,d} &:= M^2 E^{\mathrm{tr}}_{T,d}, \\ Q^{\mathrm{tr}}_{T,d} &:= 12A^{\mathrm{tr}}_{T,d} + 2VH_{T-1}, \\ P^{\mathrm{tr}}_{T,d} &:= 2E^{\mathrm{tr}}_{T,d} + 2Q^{\mathrm{tr}}_{T,d}/\mu^2, \\ J^{\mathrm{tr}}_{T,d} &:= h + 8A^{\mathrm{tr}}_{T,d}/r^2 + 8VH_T/r^2. \end{aligned} \tag{45}$$

E.2 Proof of the generic theorem

Proof. [Proof of Theorem 3] Fix a true pair in the class and an admissible law on the $\mathcal{G}_0$-event A_0 where the hypotheses hold, and write $\mathbb{E}_0, \mathbb{P}_0$ for conditional expectation and probability. States and tests are predictable. Initially $c_1 - Ay_1 = T(\rho - Ad) \geq 0$. On an active period $N_t \geq h+1$, $q_{tj} \geq q_-$ and $D_{tj} \leq D_{\max} < (h+1)q_- \leq y_{tj}$, so nonnegative A and work conservation give full service. Induction preserves

$$c_t = Ay_t + T(\rho - Ad) \quad (1 \leq t \leq \tau). \tag{46}$$

Thus $c_\tau \geq 0$, and mechanical service keeps capacity nonnegative in fallback.

We now derive the tracking budgets. Put $x_t = p_t - p_B(q_t)$ on the active prefix and extend $I_t x_t$ by zero. For $1 \leq t \leq T-h$, let $u_t = t + t_0$, $k_t = N_t - 1 \geq h$. Active transitions give $q_{t+1} = q_t - (Bx_t + \xi_t)/k_t$ and $p_B(q_{t+1}) = p_B(q_t) - (x_t + B^{-1}\xi_t)/k_t$. Feasible inverse prices and projection give

$$I_{t+1}\|x_{t+1}\|^2 \leq I_t \|F_t x_t + C_t \xi_t\|^2, \quad F_t = I + \gamma_t B + I/k_t,\ C_t = \gamma_t I + B^{-1}/k_t. \tag{47}$$

If $I_{t+1} = 0$ this is immediate. For an eigenvalue $-a$ of B, $a \in [\mu, M]$, the corresponding eigenvalue of F_t is

$$\phi_t(a) = 1 - \frac{2a}{\mu u_t} - \frac{2a/\mu - 1}{k_t}.$$

Since $u_t, k_t \geq t_0 + 1$ and $t_0 = \lceil 8M/\mu \rceil$, $0 \leq \phi_t(a) \leq 1 - 2/u_t - 1/k_t$, hence $\|F_t\|_{\mathrm{op}} \leq 1 - 2/u_t - 1/k_t$.

$$\|C_t\|_{\mathrm{op}} \leq \frac{2}{\mu}(1/u_t + 1/k_t) + \frac{1}{\mu k_t} \leq \frac{3}{\mu}(1/u_t + 1/k_t).$$

Because I_t, x_t, F_t, C_t are $\mathcal{G}_{t-1}$-measurable, conditional centering removes the cross term using only past information. Thus $z_t = \mathbb{E}_0[I_t\|x_t\|^2]$ satisfies

$$z_{t+1} \leq (1 - 2/u_t - 1/k_t)z_t + (9V/\mu^2)(1/u_t + 1/k_t)^2. \tag{48}$$

For $g_t = (t+t_0)^{-1} + N_t^{-1}$, $g_t = 1/u_t + 1/(k_t+1)$ and $g_{t+1} = 1/(u_t+1) + 1/k_t$. Direct subtraction yields $g_{t+1} - (1 - 2/u_t - 1/k_t)g_t - (1/u_t + 1/k_t)^2 = \frac{1}{u_t^2(u_t+1)} + \frac{k_t - 1}{k_t^2(k_t+1)} + \frac{k_t - 1}{u_t k_t (k_t+1)} \geq 0$. Since $z_1 \leq K_d g_1$, induction and harmonic summation give $E^{\mathrm{tr}}_{T,d}$ and $A^{\mathrm{tr}}_{T,d}$ (with $z_t = 0$ for $t > T - h$). When $T \leq h$ every period is fallback and no recursion is used.

We now bound deviations from the plan. Define $\mathcal{H}_t := \mathcal{G}_{t-1}, \Delta\mathsf{M}_s := I_s\xi_s/(T-s), \mathsf{D}_t := \sum_{s<t} I_s B x_s/(T-s)$ and $\mathsf{M}_t := \sum_{s<t} \Delta\mathsf{M}_s$. Bounded requests and means make each increment square-integrable. I_s is $\mathcal{H}_s$-measurable and $\mathbb{E}[\Delta\mathsf{M}_s \mid \mathcal{H}_s] = 0$, so $(\mathsf{M}_t, \mathcal{H}_t)$ is a square-integrable martingale. Predictability gives $\{\tau \le t\} \in \mathcal{H}_t$, hence τ is a stopping time and the stopped process remains square-integrable. Iterating from $q_1 = d$ gives, for $t \le \tau \wedge T$,

$$q_t - d = -(\mathsf{D}_t + \mathsf{M}_t), \qquad t \le \tau \wedge T, \tag{49}$$

All transitions before τ are active. At a deviation stop with $N_\tau > h$, the unmultiplied identity holds with $I_\tau = 0$. On active periods it gives the multiplied form. For arbitrary vectors v_s, set $w_k = v_{T-k}$ and $j = T - t + 1$. Then weighted Cauchy–Schwarz gives

$$\left\| \sum_{s<t} \frac{v_s}{T-s} \right\|_2^2 \le \frac{3}{\sqrt{j}} \sum_{k=j}^{T-1} \frac{\|w_k\|_2^2}{\sqrt{k}}.$$

Reversing sums and using $\sum_{j \le k} j^{-1/2} \le 2\sqrt{k}$ proves

$$\sum_{t=1}^{T} \left\| \sum_{s<t} \frac{v_s}{T-s} \right\|^2 \le 6 \sum_{s=1}^{T-1} \|v_s\|^2. \tag{50}$$

Martingale orthogonality gives

$$\mathbb{E}_0 \sum_{t=1}^{T} \|\mathsf{M}_t\|_2^2 = \sum_{s=1}^{T-1} \frac{\mathbb{E}_0[I_s \|\xi_s\|_2^2]}{T-s} \le V H_{T-1}.$$

Therefore,

$$\mathbb{E}_0 \sum_{t=1}^{T} I_t \|q_t - d\|^2 \le 2\mathbb{E}_0 \sum_t \|\mathsf{D}_t\|^2 + 2\mathbb{E}_0 \sum_t \|\mathsf{M}_t\|^2 \le 12 A^{\mathrm{tr}}_{T,d} + 2V H_{T-1} = Q^{\mathrm{tr}}_{T,d}. \tag{51}$$

We now bound the expected number of fallback periods. If $N_\tau > h$, one of $\|\mathsf{D}_\tau\|, \|\mathsf{M}_\tau\|$ exceeds $r/2$, and $N_\tau \|\mathsf{D}_\tau\|^2 \le 2\sum_s \|I_s B x_s\|^2$. Let σ_κ be the first passage of $\mathsf{M}_{t\wedge\tau}$ above κ. For deterministic K, $\sigma_\kappa \wedge K$ is a bounded $\mathcal{H}$-stopping time. Conditional centering after the passage (at $\mathcal{H}_{\sigma_\kappa} = \mathcal{G}_{\sigma_\kappa - 1}$) removes the cross term and

$$\kappa^2 \Pr_0\left(\max_{t \le K} \|\mathsf{M}_{t\wedge\tau}\|_2 \ge \kappa \right) \le \mathbb{E}_0 \|\mathsf{M}_{K\wedge\tau}\|_2^2.$$

This uses only the stopped process and remains valid on A_0. We have the deterministic-horizon inclusion

$$\{N_\tau \ge j,\ \|\mathsf{M}_\tau\| \ge r/2\} \subseteq \left\{ \max_{t \le T-j+1} \|\mathsf{M}_{t\wedge\tau}\| \ge r/2 \right\}.$$

Since $N_\tau \ge j$ implies $\tau \le T - j + 1$, taking $K = T - j + 1$ and $\kappa = r/2$ gives the stopped-martingale variance bound

$$\mathbb{E}_0 \|\mathsf{M}_{K\wedge\tau}\|_2^2 \le V \sum_{s=1}^{K-1} (T-s)^{-2} = V \sum_{k=j}^{T-1} k^{-2}.$$

Hence

$$\Pr_0\{N_\tau \ge j,\ \|\mathsf{M}_\tau\| \ge r/2\} \le \frac{4V}{r^2} \sum_{k=j}^{T-1} k^{-2} \le \frac{8V}{r^2 j}.$$

Let $\mathcal{D} = \{N_\tau > h, \|\mathsf{D}_\tau\| \geq r/2\}$ and $\mathcal{M} = \{N_\tau > h, \|\mathsf{M}_\tau\| \geq r/2\}$. Then $\{N_\tau > h\} \subseteq \mathcal{D} \cup \mathcal{M}$. $N_\tau \mathbf{1}_{\mathcal{D}} \leq 8 \sum_s \|I_s B x_s\|^2/r^2$ and the tail display gives $\mathbb{E}_0[N_\tau \mathbf{1}_{\mathcal{M}}] \leq 8VH_T/r^2$. Hence $\mathbb{E}_0 N_\tau \leq h + 8A^{\mathrm{tr}}_{T,d}/r^2 + 8VH_T/r^2 = J^{\mathrm{tr}}_{T,d}$. Finally $p_t - p_B(d) = x_t + B^{-1}(q_t - d)$ on active periods and $\|B^{-1}\| \leq 1/\mu$ give

$$\mathbb{E}_0 \sum_t I_t \|p_t - p_B(d)\|^2 \leq 2E^{\mathrm{tr}}_{T,d} + 2Q^{\mathrm{tr}}_{T,d}/\mu^2 = P^{\mathrm{tr}}_{T,d}. \tag{52}$$

We now derive the fluid comparison. For $u = A^\top \lambda$ let $\ell_u(v) = (v-u)^\top f(v)$. At $\bar{p} = p_B(d)$, $\nabla \ell_u(\bar{p}) = g_B$ and $\nabla^2 \ell_u = 2B \preceq -2\mu I$, so $\max_v \ell_u(v) - \ell_u(\bar{p}) \leq \|g_B\|^2/(4\mu)$. Weak duality and the quadratic expansion with Young's inequality give

$$r^{\mathrm{fl}} - p^\top f(p) \leq \lambda^\top \{\rho - A f(p)\} + (M + \mu)\|p - \bar{p}\|^2 + \|g_B\|^2/(2\mu). \tag{53}$$

Predictability and full active service permit centering requests. On the active prefix, $y_\tau = Td - \sum_{t<\tau} D_t$, so

$$\sum_{t<\tau} \lambda^\top(\rho - AD_t) = Ts + u^\top y_\tau - N_\tau \lambda^\top \rho.$$

If $\tau = 1$, $y_{\tau j} = Td_j \leq N_\tau q_+$. Otherwise, $q_{\tau-1,j} \leq d_j + r \leq q_+$ and $y_{\tau j} \leq (N_\tau + 1)q_+ \leq 2N_\tau q_+$. Thus the expected load term is at most $Ts + 2q_+\|u\|_1 \mathbb{E}_0 N_\tau$. Fallback occupies N_τ periods and earns nonnegative revenue, so its loss is at most $R_{\max}\mathbb{E}_0 N_\tau$. Substitution of the tracking, price and exit budgets proves (11). The direct-gradient bound (12) replaces $\|g_B\|$ by η. □

Proof. [Proof of Corollary 1] Under exact alignment, $u = p^0$, $s = 0$, and $g_B(d^*, \lambda^0) = -\boldsymbol{e}$, so substitution in the generic bound gives the first displayed inequality. If $\boldsymbol{e} = 0$, ℓ_{p^0} is strictly concave and has zero gradient at $p_B(d^*)$. In this case, weak duality and $Ad^* = \rho$ then make this price the unique initial-fluid solution. For the residual specialization, the triangle inequality and $\|B'\|_{op} \leq M$ give $\|2d - d^0 + B'(p^0 - A^\top\lambda) - \boldsymbol{e}'\|_2 \leq \|2d - d^0\|_2 + M\|p^0 - A^\top\lambda\|_2 + \bar{\epsilon} = \eta_{\mathrm{al}}$. The generic uniform bound then gives the residual statement, with the same plan-centred $P^{\mathrm{tr}}_{T,d}$ and $J^{\mathrm{tr}}_{T,d}$. □

Proof. [Proof of Corollary 2] Write $u_C = A^\top \lambda_C$. Interior nominal stationarity is $d_C + C(p_C - u_C) = 0$, while $d_C = d^0 + C(p_C - p^0)$. Therefore $g_C(d_C, \lambda_C) = 2d_C - d^0 + C(p^0 - u_C) = 0$. Complementary slackness gives $s(d_C, \lambda_C) = 0$. For every $(B', e') = (C + E, e') \in \mathfrak{C}_C$, $2d_C - d^0 + B'(p^0 - u_C) - e' = g_C(d_C, \lambda_C) + E(p^0 - u_C) - e' = E(p^0 - u_C) - e'$, so its norm is at most $\delta_B\|p^0 - u_C\|_2 + \bar{\epsilon} = \eta$. Taking the supremum gives $R_{\mathrm{grad}}(d_C, \lambda_C; \mathfrak{C}_C) \leq \eta$, and the truth belongs to $\mathfrak{C}_C$, so $\|g_B\|_2 \leq \eta$. The remaining hypotheses of Theorem 3 hold at (d_C, λ_C, r) by assumption, so its bound applies with $\|g_B\|_2^2 \leq \eta^2$. The argument allows non-strict complementarity and rank deficiency. □

E.3 Dense unknown-matrix illustration

A two-product example has $A = (1, 1)$, $p^0 = (3, 3)$, $d^0 = (4, 4)$, $\lambda^0 = 3$, $\rho = 4$, and $C = \begin{pmatrix} -1 & 1/5 \\ 1/5 & -1 \end{pmatrix}$. A small symmetric Frobenius neighborhood retains positive cross-effects and strict inverse-image margin, illustrating feasibility with unknown entries.

F Minimax Rate with an Exact Anchor

Proposition 3 (Minimax rate with an exact anchor and a shared resource)**.** *Let* $n = 2$, $A = (1, 1)$, $[L, U] = [1/2, 3/2]$, $C_T = 7T/16$, *and* $v = (1, -1)^\top$. *Fix the common exact anchor* $\boldsymbol{p}^0 = (7/8, 9/8)^\top$, $\boldsymbol{d}^0 = (55/256, 73/256)^\top$ *and the family* $f_\eta(\boldsymbol{p}) = \boldsymbol{d}^0 + B_\eta(\boldsymbol{p} - \boldsymbol{p}^0)$, $\quad B_\eta = -\frac{1}{4}I_2 - (\frac{1}{64} + \eta)vv^\top$, $\quad |\eta| \leq 1/128$.

Requests have conditionally independent Bernoulli coordinates with these means, and fulfillment follows any fixed rule in Definition 1. This fixed family satisfies the model assumptions uniformly, including a positive initial demand-image margin. Its response matrices are symmetric negative definite with strictly positive cross-price coefficients. Its unique fluid optima bind the shared resource, have common shadow price $1/4$*, and satisfy* $|p^*_{\eta,j} - p^0_j| \geq 1/16 \quad (j = 1, 2), \qquad \|\boldsymbol{p}^*_\eta - \boldsymbol{p}^0\|_2 \geq \sqrt{2}/8$*. When the policy knows this entire family and that the anchor is exact, for every integer* $T \geq 1$*,* $2^{-22}\sqrt{T} \leq \inf_\pi \sup_{|\eta|\leq 1/128} \mathrm{Reg}^{\mathrm{fl}}_T(\pi; f_\eta) \leq 7\sqrt{T}$*, where the infimum ranges over all randomized nonanticipating pricing policies. The constants and separation bounds are independent of the horizon.*

Use Proposition 3's fixed family, writing $\Theta = [-1/128, 1/128]$, $v = (1, -1)^\top$, $q = (1, 5/4)^\top$, $D_\eta = 9/8 + 8\eta$, $s_\eta = \eta/(2D_\eta)$ and $\lambda = 1/4$.

$$B_\eta = -I_2/4 - (1/64 + \eta)vv^\top, \qquad f_\eta(p) = d^0 + B_\eta(p - p^0). \tag{54}$$

Its cross effects lie in $[1/128, 3/128]$, eigenvalues are $-1/4, -9/32 - 2\eta$, and box means lie in $[15/512, 241/512]$. Independent Bernoulli coordinates satisfy Assumption 4(a) with proxy $1/4$. The corner-regressor Gram matrix is $G = ((1, 1, 1), (1, 5/4, 1), (1, 1, 5/4))$. The leading minors of $G - I/16$ are $15/16, 29/256, 39/4096$, proving identifiability. The common load is

$$Af_\eta(p) = 1 - (p_1 + p_2)/4. \tag{55}$$

Direct stationarity gives $p^*_\eta = q + s_\eta v$, the capacity constraint is binding at $7/16$, and $f_\eta(p^*_\eta) + B^\top_\eta(p^*_\eta - \lambda\mathbf{1}) = 0$. Its value is

$$r^*_\eta = 513/1024 + \eta^2/(8D_\eta). \tag{56}$$

Since $|s_\eta| \leq 1/272$, price and demand-image margins are at least $1/8, 1/32$. The anchor's price margin is $3/8$. The price separation satisfies

$$|p^*_{\eta,j} - p^0_j| \geq 1/16, \quad \|p^*_\eta - p^0\|^2 = 1/32 + 2s^2_\eta, \tag{57}$$

while both informative candidates remain blind:

$$f_\eta(p^0) = d^0, \quad f_\eta(q) = (47/256, 65/256)^\top. \tag{58}$$

All primitives are fixed uniformly in η, T. Revealing the family to the learner only strengthens the lower bound.

Proof of Proposition 3. Write $Q_\eta = -B_\eta \succeq I/4$, and let P_η be the policy's full observed-history law, R_η its sales-revenue fluid regret. We now relate regret to deviations from the fluid optimum using the exact identity

$$r^*_\eta - \lambda(7/16) - (p - \lambda\mathbf{1})^\top f_\eta(p) = (p - p^*_\eta)^\top Q_\eta(p - p^*_\eta). \tag{59}$$

Since $p - \lambda\mathbf{1} \geq 0$, $p_t^\top \widehat{d}_t \leq \lambda A\widehat{d}_t + (p_t - \lambda\mathbf{1})^\top d_t$. We now sum over time, use the bound C_T on total served load, and center predictable requests to obtain

$$R_\eta(T) \geq \mathbb{E}_\eta \sum_t (p_t - p^*_\eta)^\top Q_\eta(p_t - p^*_\eta) \geq \tfrac{1}{4}\mathbb{E}_\eta \sum_t \|p_t - p^*_\eta\|^2. \tag{60}$$

Bernoulli $\mathrm{kl}(u, w) \leq (u - w)^2/[w(1 - w)]$, the box mean bounds and $f_\eta(p) - f_0(p) = -\eta vv^\top(p - q)$ bound one-step KL by $512\eta^2\|p - q\|^2$. Common action kernels cancel in the adaptive chain rule. Stock and service add no information beyond observed requests. Thus

$$\mathrm{KL}(P_0, P_\eta) \leq 2048\eta^2 R_0(T). \tag{61}$$

Take $\delta = T^{-1/4}/128$ and $d = \|p^*_\delta - q\|$, with $d^2 \geq \delta^2/4$. If $R_0 \geq \sqrt{T}$ the claim is immediate. Otherwise KL is at most $1/8$, so total variation is at most $1/4$. An independent uniform time J and test $E = \{\|p_J - q\| \geq d/2\}$ give $R_0 + R_\delta \geq Td^2\{P_0(E) + P_\delta(E^c)\}/16 \geq T\delta^2/128$. Hence the maximum is at least $2^{-22}\sqrt{T}$.

For the upper bound, post $p^{\mathrm{E}} = (3/4, 3/2)$ for $m_T = \min\{T, \lceil\sqrt{T}\rceil\}$ periods, then $p^*_{\widehat{\eta}}$, where $\widehat{\eta} = \Pi_\Theta\{2(\bar{D}_1 - 65/256)\}$. This policy uses only observed requests. Projection gives $\mathbb{E}(\widehat{\eta} - \eta)^2 \leq 1/m_T$. Using $\sup|s'_\eta| = 144/289 < 1/2$ and $\|Q_\eta\| \leq 19/64$, we bound potential-revenue regret by $m_T + 19T/(128m_T)$. All posted prices have mean total requests $7/16$. For $S = \sum_t Ad_t$, any first truncation leaves less than two units, so lost requests are at most $(S - C_T)_+ + 2$. Conditional variance gives $\mathbb{E}(S - C_T)^2 \leq T/2$. We now use the price bound $3/2$ to bound the revenue loss from truncation, yielding

$$R_\eta(T) \leq m_T + 19T/(128m_T) + (3/2)(\sqrt{T/2} + 2) \leq 7\sqrt{T}. \tag{62}$$

□

The missing direction is explicit: $H(p^* - p^0) = 0$ and $AH = 0$ can leave operating demand and resource feasibility unchanged, while $H^\top(p^* - A^\top\lambda^*)$ has a nonzero feasible-tangent component. Here $H = -vv^\top$. Identifying this direction requires an additional primitive restriction or historical information beyond coordinate or norm separation.

G Historical Selector Inputs and Eligibility

Definition 15 (Historical selector inputs)**.** Fix the model's price set, $p^0 \in \mathcal{P}$, nonnegative A, ρ and Assumption 4(b). A pre-campaign record H takes values in a standard Borel space, $\mathcal{H} = \sigma(H) \subseteq \mathcal{F}^0$, with law $\mathbb{P}_H$, and a measurable set $\mathsf{H}_{\mathrm{dec}}$ of records. The selected $C \in \mathbb{S}^n$, $d^0, d \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$ and finite nonnegative $\delta_B, \bar{\epsilon}, \eta$ are Borel maps. Extend $d^0 = 0$ off $\mathsf{H}_{\mathrm{dec}}$, so $\epsilon_0 = \|d^0 - f(p^0)\|$ is always defined. Fix $0 < \mu_0 \leq M_0$, $r_0 > 0$, $0 < q_0 \leq q_1$ and $u_0 \geq 0$. Define $\mathfrak{C}(h) = \{(B', e') : B' = B'^\top, \|B' - C(h)\|_{\mathrm{op}} \leq \delta_B(h), \|e'\| \leq \bar{\epsilon}(h)\}$, with $e(h) = f(p^0) - d^0(h)$, $p_B(q; h) = p^0 + B^{-1}[q - d^0(h) - e(h)]$, and $p_C(q; h) = p^0 + C(h)^{-1}[q - d^0(h)]$ only after the spectral gates.

On the spectral-pass set S, select the unique maximizer d_C of $r_C(q) = p_C(q)^\top q$ over $\{q \geq 0 : Aq \leq \rho\}$ and a minimum-norm nonnegative complementary multiplier λ_C. We now set $(d, \lambda, \eta) = (d_C, \lambda_C, \eta_C)$ there and $(d^0, 0, 0)$ off S. Compactness, finite support enumeration for redundant rows, and fixed tie-breaking make these maps total and Borel. Initialize at $p_C(d)$ only after the spectral and price-image checks. Otherwise, use the fixed box-valued fallback p^{fb}. The initializer is Borel and predictable.

Definition 16 (Observable eligibility and fallback)**.** A record in $\mathsf{H}_{\mathrm{dec}}$ is eligible when

$$\begin{gathered}\lambda_{\min}(-C) - \delta_B \geq \mu_0, \quad \lambda_{\max}(-C) + \delta_B \leq M_0, \\ Ad \leq \rho, \quad \lambda \geq 0, \quad d_j - r_0 \geq q_0, \quad d_j + r_0 \leq q_1, \quad \|A^\top\lambda\|_1 \leq u_0,\end{gathered} \tag{63}$$

and the corrected margin (67), direct residual $R_{\mathrm{grad}}(d, \lambda; \mathfrak{C}) \leq \eta$, and $\mathcal{I}(d, r_0, \mathfrak{C})$ all pass with μ_0, M_0, t_{00}, h_0. Require the indicator of $\mathcal{I}$ to be Borel. The remaining finite continuous inequalities in Borel maps then make G_T historical-measurable. Set the predicate false off $\mathsf{H}_{\mathrm{dec}}$ and evaluate spectral tests before inverses. Algorithm 3 runs the complete informed controller only on gate pass and score selection. Otherwise, it runs NI from period one. Fixed finite $c_{\mathrm{ref}} > 0$ makes the NI score finite. Its finite design gives well-defined OLS, compact projection gives invertible $\widehat{B}_L$, and its fixed box-valued fallback completes every failed solve or image exit. Both branches fulfill mechanically throughout.

Definition 17 (Selected policy)**.** The record H is the historical sample of Definition 20 with $\delta = \delta_T$. Here $C, d^0, \delta_B, \bar{\epsilon}$ are its estimators and radii, and the maps and gates are those of Definitions 15–16. Write $\mathcal{H} = \sigma(H)$ and let Assumption 4(a) hold with $\mathcal{G}_t = \mathcal{H} \vee \mathcal{F}^t$ in place of $\mathcal{F}^t$.

Let the historical coverage event be

$$E_T := E_T^{\mathrm{hist}} \quad \text{from (68)}, \qquad 0 < \delta_T < 1, \qquad \Pr(E_T^c) \le \delta_T. \tag{64}$$

Let π_{NI} be Algorithm 2 with the fixed design and schedule of Theorem 2, and let π_{inf} be the informed controller with the constants in Definition 22. For a fixed $c_{\mathrm{ref}} > 0$, define $a_T = c_{\mathrm{ref}}(1 + \log T) + T\{s + \eta^2/(2\mu_0)\}$ on G_T and $a_T = \infty$ on G_T^c, and set $b_T = c_{\mathrm{ref}}\sqrt{T}$. The selected policy π^{sel} equals π_{inf} on $G_T \cap \{a_T \le b_T\}$ and equals π_{NI} otherwise.

H Correcting the Demand Plan with a Known Response Class

Definition 18 (Response class and plan check)**.** Fix $p^0, d^0, C = C^\top \succ 0$, $\delta_B, \bar{\epsilon} \ge 0$, with the common curvature bounds of the invoking theorem. Every comparator other than the oracle receives the center and radius. The default class is $\mathfrak{C}_{\mathrm{norm}} = \{(B, e) : B = B^\top = C + E,\ \|E\|_{op} \le \delta_B,\ \mu I \preceq -B \preceq MI,\ \|e\| \le \bar{\epsilon}\}$. For $Ad \le \rho$, $\lambda \ge 0$ and $u = A^\top\lambda$, put $g_C = 2d - d^0 + C(p^0 - u)$,

$$\eta_C = \|g_C\| + \delta_B\|p^0 - u\| + \bar{\epsilon}, \qquad s = \lambda^\top(\rho - Ad), \qquad R_{\mathrm{grad}} = R_{\mathrm{grad}}(d, \lambda; \mathfrak{C}) \tag{65}$$

with R_{grad} as in (12), and $p_C(q) = p^0 + C^{-1}(q - d^0)$.

Definition 19 (Inverse-image gate and budgets)**.** For a parameter set $\mathfrak{C}$, $\mathcal{I}(d, r, \mathfrak{C})$ means (39), $0 < r < \min_j d_j$ and the same q_-, q_+, t_0, h prescribed by Theorem 3, using the bounded-request constant $D_{\max}$. All tracking and exit budgets (45) use this plan and radius: write $P_T^{\mathrm{tr}} = P_{T,d}^{\mathrm{tr}}$, $J_T^{\mathrm{tr}} = J_{T,d}^{\mathrm{tr}}$. A failed pre-deployment check selects NI.

Proposition 4 (Response-class plan correction)**.** *Adopt Theorem 3's conditional hypotheses and complete controller, Definitions 18–19 and the same-class bound $R_{\mathrm{grad}} \le \eta < \infty$. For every instance in the class, horizon and admissible law, active requests are fully served and, on the $\mathcal{G}_0$-event where these hypotheses hold, the conditional loss $L_T(\mathcal{G}_0; \pi^{\mathrm{gpf}})$ obeys*

$$L_T(\mathcal{G}_0; \pi^{\mathrm{gpf}}) \le (M + \mu)P_T^{\mathrm{tr}} + T\{s + \eta^2/(2\mu)\} + (R_{\max} + 2q_+\|A^\top\lambda\|_1)J_T^{\mathrm{tr}}. \tag{66}$$

The norm class permits $\eta = \eta_C$. Any sharper bound and image gate must use the same smaller class and its own known bounds $(\delta_{\mathfrak{C}}, \bar{\epsilon}_{\mathfrak{C}}, \mu_{\mathfrak{C}})$ satisfying $\|B' - C\|_{\mathrm{op}} \le \delta_{\mathfrak{C}}$, $\|e'\|_2 \le \bar{\epsilon}_{\mathfrak{C}}$, and $\mu_{\mathfrak{C}} I \preceq -B'$ for every $(B', e') \in \mathfrak{C}$. Nominal interior KKT plans have $s = g_C = 0$ and the specialization in Corollary 2. Redundant and overlapping rows are allowed.

Proof. Write $u = A^\top\lambda$ and $\ell_u(p) := (p - u)^\top f(p)$. Since $p_C(d) = p^0 + C^{-1}(d - d^0)$, $d = d^0 + C\{p_C(d) - p^0\}$. For every $(B', e') \in \mathfrak{C}$ and every q,

$$\begin{aligned} p_{B',e'}(q) - p_C(d) &= p^0 + (B')^{-1}(q - d^0 - e') - p_C(d) \\ &= (B')^{-1}\left[q - d^0 - e' - B'\{p_C(d) - p^0\}\right] \\ &= (B')^{-1}\left[q - d - e' - (B' - C)\{p_C(d) - p^0\}\right]. \end{aligned}$$

For the default norm class, $-B' \succeq \mu I$, so $\|(B')^{-1}\|_{\mathrm{op}} \le 1/\mu$ and, whenever $\|q - d\|_2 \le r$, $\|p_{B',e'}(q) - p_C(d)\|_2 \le \{r + \bar{\epsilon} + \delta_B\|p_C(d) - p^0\|_2\}/\mu$. Each coordinate moves by at most this amount, so the computable margin

$$\min_j\{p_{Cj}(d) - L, U - p_{Cj}(d)\} \ge \{r + \bar{\epsilon} + \delta_B\|p_C(d) - p^0\|_2\}/\mu \tag{67}$$

implies $p_{B',e'}(q) \in \mathcal{P}$ for every class member and every such q, which is the image gate (39) required by $\mathcal{I}$ and Theorem 3. For a sharper class $\mathfrak{C}$ the same calculation holds with $(\delta_B, \bar{\epsilon}, \mu)$ replaced by $(\delta_{\mathfrak{C}}, \bar{\epsilon}_{\mathfrak{C}}, \mu_{\mathfrak{C}})$, and $\mu_{\mathfrak{C}}$ is the curvature constant μ passed to the theorem and its budgets.

We now bound the gradient norm. For the true pair, $f(p_B(d)) = d$ and $\nabla \ell_u(p_B(d)) = 2d - d^0 + B(p^0 - u) - e$. Writing $B = C + E$ gives $\nabla \ell_u(p_B(d)) = g_C + E(p^0 - u) - e$, whose norm is bounded by $R_{\mathrm{grad}}(d, \lambda; \mathfrak{C}) \le \eta$. For the norm class, every member (B', e') with $B' = C + E'$ gives $\|g_C + E'(p^0 - u) - e'\|_2 \le \|g_C\|_2 + \delta_B \|p^0 - u\|_2 + \bar{\epsilon} = \eta_C$, so $\eta = \eta_C$ is admissible. Also $s = \lambda^\top(\rho - Ad) \ge 0$. Thus all hypotheses of Theorem 3 hold with initial sigma-field $\mathcal{G}_0$, so its conditional bound gives (66). For a nominal interior KKT plan, stationarity gives $g_C = 0$ and complementary slackness gives $s = 0$. □

H.1 Historical response bounds

Definition 20 (Historical fixed-design coverage)**.** Fix a feasible $p^0 \in \mathcal{P}$, $\ell > 0$, set $x^0 = p^0$, and let $x^k = p^0 + \ell e_k \in \mathcal{P}$ for $k = 1, \dots, n$, and let $\mathcal{K} = \{0, 1, \dots, n\}$. Before observing data, choose $N \ge 1$ and $0 < \delta < 1$, collect independent requests $D^{k,r}$, $r = 1, \dots, N$, at each x^k, whose centered coordinates $D_j^{k,r} - f_j(x^k)$ obey, for a fixed $\sigma_{\mathrm{off}} > 0$ and every $\lambda \in \mathbb{R}$, $\mathbb{E} \exp\{\lambda(D_j^{k,r} - f_j(x^k))\} \le \exp\{\lambda^2 \sigma_{\mathrm{off}}^2/2\}$. Thus the mean of the N independent requests has proxy $\sigma_{\mathrm{off}}^2/N$ and $\Pr\{|\bar{D}_j^k - f_j(x^k)| > u\} \le 2\exp\{-u^2 N/(2\sigma_{\mathrm{off}}^2)\}$. Set $\bar{D}^k = N^{-1} \sum_r D^{k,r}$, $b_N := \sigma_{\mathrm{off}} \sqrt{2 \log\{2n(n+1)/\delta\}/N}$ and

$$E_T^{\mathrm{hist}} := \{\max_{k \in \mathcal{K}} \|\bar{D}^k - f(x^k)\|_\infty \le b_N\}. \tag{68}$$

The fixed-design response bound is $d^0 := \bar{D}^0$, $\widetilde{B}_{:j} := (\bar{D}^j - \bar{D}^0)/\ell$ and $C := (\widetilde{B} + \widetilde{B}^\top)/2$.

Definition 21 (Historical sampling fee)**.** Let $C_{\mathrm{hist}} \ge 0$ be the total fee for the historical request vectors used in Definition 20, including any bundled label cost, and let it be $\mathcal{H}$-measurable. The regrets below are gross of C_{hist}. Profit after fees subtracts C_{hist} once, in the same way as (19) subtracts its acquisition and computation cost.

Lemma 11 (Historical coverage implies class membership)**.** *Under Assumption 2(a) and the symmetry condition $B = B^\top$, the construction in Definition 20 satisfies $\Pr((E_T^{\mathrm{hist}})^c) \le \delta$ and, on E_T^{hist},*

$$\|B - C\|_{\mathrm{op}} \le \delta_B := \frac{2nb_N}{\ell}, \qquad \|f(p^0) - d^0\|_2 \le \bar{\epsilon} := \sqrt{n}\, b_N. \tag{69}$$

Proof. For every k, j, independence across r gives $\mathbb{E} \exp\{\lambda(\bar{D}_j^k - f_j(x^k))\} \le \exp\{\lambda^2 \sigma_{\mathrm{off}}^2/(2N)\}$ and $\Pr\{|\bar{D}_j^k - f_j(x^k)| > u\} \le 2e^{-u^2 N/(2\sigma_{\mathrm{off}}^2)}$. At $u = b_N$ the exponential bound equals $\delta/[n(n+1)]$, so each tail probability is at most that value. The union bound over the $n(n+1)$ pairs (k, j) therefore bounds the failure probability by δ. On the event, writing $\varepsilon^k = \bar{D}^k - f(x^k)$, the jth slope column has error $(\varepsilon^j - \varepsilon^0)/\ell$ and hence Euclidean norm at most $2\sqrt{n} b_N/\ell$. Thus the unsymmetrized slope error has Frobenius norm at most $2nb_N/\ell$. Since $B = B^\top$, symmetrization is an orthogonal projection, so it does not increase the Frobenius norm, and operator norm is no larger than Frobenius norm, proving the first bound. The anchor error has n coordinates bounded by b_N, so the second bound follows. □

The lemma supplies $\bar{\epsilon}$ and δ_B. Weyl supplies curvature, while optional stopping uses the separate coverage event.

H.2 Selecting by regret scales

Definition 22 (Uniform selector constants)**.** Use Definition 15's fixed positive $\mu_0, M_0, r_0, q_0, q_1, u_0$, the bounded-request $D_{\max}$, valid joint response coverage and the gates (63), (67). On covered eligible histories these imply $\mu_0 I \preceq -B \preceq M_0 I$ while B remains unobserved. Put $V = n\sigma^2$, $t_{00} = \lceil 8M_0/\mu_0 \rceil$, $h_0 = \max\{t_{00}+1, \lceil D_{\max}/q_0 \rceil\}$ and $K_0 = \max\{n(U-L)^2(1+t_{00}), 9V/\mu_0^2\}$. Set $p_0^{\mathrm{tr}} = 4K_0 + (48M_0^2K_0 + 4V)/\mu_0^2$, $j_0^{\mathrm{tr}} = (16M_0^2K_0 + 8V)/r_0^2$, and

$$C_{\log} = (M_0+\mu_0)p_0^{\mathrm{tr}} + (R_{\max} + 2q_1u_0)(h_0 + j_0^{\mathrm{tr}}). \tag{70}$$

Use the same fixed r_0, t_{00}, h_0 and bound $R_{\mathrm{grad}} \le \eta$ on every eligible history, including numerical optimization error. A is fixed and nonnegative. Redundant and overlapping rows are allowed.

Lemma 12 (Uniform conditional cost of the informed controller)**.** *Assume the true response belongs to the class in Definition 3, under Assumptions 2(a) and 4. Use Definitions 15–16 and 22. Assume the record H contains the full historical sample $(D^{k,r})_{k\in\mathcal{K}, r\le N}$ of Definition 20, and that C, d^0, δ_B and $\bar{\epsilon}$ are its estimators and radii computed from it. Put $\mathcal{H} := \sigma(H)$ and let $\mathfrak{C} := \mathfrak{C}(H)$ be the parameter set of Definition 15. Set*

$$E_T := E_T^{\mathrm{hist}} \quad \text{as defined in (68)}, \qquad G_T := \left\{ H \in \mathsf{H}_{\mathrm{dec}} : \begin{array}{c} \text{all inequalities in (63) and (67) hold,} \\ \mathcal{I}(d, r_0, \mathfrak{C}) \text{ holds and } R_{\mathrm{grad}}(d, \lambda; \mathfrak{C}) \le \eta \end{array} \right\}. \tag{71}$$

Both events are $\mathcal{H}$-measurable: E_T is a Borel function of the sample means $\bar{D}^k$, which are functions of the historical sample contained in H, while G_T is Borel by the requirement that the indicator of $\mathcal{I}$ be Borel in Definition 16. On E_T, Lemma 11 gives $\|B - C\|_{\mathrm{op}} \le \delta_B$ and $\|e\| \le \bar{\epsilon}$, so the true pair is in $\mathfrak{C} = \mathfrak{C}(H)$ on $E_T \cap G_T$. On G_T, the displayed inequalities supply resource feasibility, curvature via Weyl, the image margin, $\mathcal{I}$, and $R_{\mathrm{grad}} \le \eta$. Set $\mathcal{G}_t := \mathcal{H} \vee \mathcal{F}^t$. Suppose Assumption 4(a) holds conditionally on $\mathcal{G}_{t-1}$ with the same fixed bounds. Almost surely on $E_T \cap G_T$ with $s + \eta^2 < \infty$, the controller of Proposition 4, with $t_0 = t_{00}$ and terminal-window parameter h_0, has conditional loss $L_T(\mathcal{H}; \pi^{\inf})$ at most $C_{\log}(1 + \log T) + T\{s + \eta^2/(2\mu_0)\}$. The radius in both tracking and exit budgets is the same fixed r_0, and the constant in (70) is independent of the history and horizon.

Proof. Because H takes values in a standard Borel space and all selected quantities are Borel maps, $s(H)$, $\eta(H)$ and $s(H) + \eta(H)^2$ are $\mathcal{H} = \sigma(H)$-measurable. For $b \in \mathbb{N}$ let $B_b := E_T \cap G_T \cap \{s + \eta^2 \le b\} \in \mathcal{H}$. We work conditionally on $\mathcal{G}_0 := \mathcal{H}$ throughout. Proposition 4 applies directly to $\mathcal{G}_0$-measurable inputs and conditional losses.

On B_b, Lemma 11 puts the true pair in $\mathfrak{C} = \mathfrak{C}(H)$. Weyl and the spectral gates give, for every $(B', e') \in \mathfrak{C}$, $\lambda_{\min}(-B') \ge \lambda_{\min}(-C) - \|B' - C\|_{\mathrm{op}} \ge \mu_0$ and $\lambda_{\max}(-B') \le \lambda_{\max}(-C) + \|B' - C\|_{\mathrm{op}} \le M_0$. The remaining gates give $q_- := \min_j d_j - r_0 \ge q_0$, $q_+ := \max_j d_j + r_0 \le q_1$ and $\|A^\top \lambda\|_1 \le u_0$. Also $t_{00} = \lceil 8M_0/\mu_0 \rceil$, $h_0 \ge t_{00} + 1$ and $h_0 \ge \lceil D_{\max}/q_0 \rceil \ge \lceil D_{\max}/q_- \rceil$. These inputs are $\mathcal{H}$-measurable, and by hypothesis Assumption 4(a), including its conditional moment-generating-function bound, holds given $\mathcal{G}_{t-1} = \mathcal{H} \vee \mathcal{F}^{t-1}$ for every t. Hence Proposition 4 applies with $(\mu, M, r, t_0, h) = (\mu_0, M_0, r_0, t_{00}, h_0)$ and $\mathcal{G}_0 = \mathcal{H}$, and its conclusion holds almost surely on B_b. In particular the stopped martingales are conditionally square-integrable: for every deterministic $K \le T$, $\mathbb{E}[\|\mathsf{M}_{K\wedge\tau}\|_2^2 \mid \mathcal{H}] \le V \sum_{s=1}^{T-1} (T-s)^{-2} < \infty$, $V = n\sigma^2$, and conditional optional sampling gives

$$\mathbb{P}\{N_\tau \ge j, \|\mathsf{M}_\tau\|_2 \ge r_0/2 \mid \mathcal{H}\} \le \frac{4V}{r_0^2} \sum_{k=j}^{T-1} k^{-2} \le \frac{8V}{r_0^2 j}.$$

On B_b, $u := A^\top \lambda \ge 0$ and $0 \le \lambda^\top \rho = s + u^\top d \le b + q_1 \|u\|_1 \le b + u_0 q_1$, which supplies integrability.

We now bound the predecessor budgets: $P^{\mathrm{tr}}_{T,d} \le [4K_0 + (48M_0^2K_0 + 4V)/\mu_0^2]H_T = p_0^{\mathrm{tr}}H_T$ and $J^{\mathrm{tr}}_{T,d} \le h_0 + (16M_0^2K_0 + 8V)H_T/r_0^2 = h_0 + j_0^{\mathrm{tr}}H_T$. Substitution into (66), with $q_+ \le q_1$, $\|A^\top\lambda\|_1 \le u_0$ and $R_{\mathrm{grad}} \le \eta$, yields

$$L_T(\mathcal{H};\pi^{\mathrm{inf}}) \le C_{\log}H_T + T\left\{s + \frac{\eta^2}{2\mu_0}\right\} \le C_{\log}(1+\log T) + T\left\{s + \frac{\eta^2}{2\mu_0}\right\} \quad \text{a.s. on } B_b.$$

For $T \le h_0$, revenue from sales is nonnegative and $V^{\mathrm{Fluid}}(C_T) \le TR_{\max}$, so $L_T(\mathcal{H};\pi^{\mathrm{inf}}) \le R_{\max}h_0 \le C_{\log} \le C_{\log}(1+\log T)$, because (70) gives $C_{\log} \ge R_{\max}h_0$ and $1 + \log T \ge 1$. The constant in (70) is unchanged.

For each b, the bound fails only on a null set $\mathcal{N}_b$. The countable union $\mathcal{N} := \bigcup_{b\ge1}\mathcal{N}_b$ is null, and $E_T \cap G_T \cap \{s+\eta^2 < \infty\} = \bigcup_{b\ge1} B_b$. Thus one common null set and the same constants work for all eligible covered histories. □

Proof of Theorem 4. Fix T. All selected quantities, G_T, a_T and b_T are $\mathcal{H}$-measurable. Define $I_T := \mathbf{1}\{G_T \cap \{a_T \le b_T\}\}$. The branch is selected before period one, so

$$L_T(\mathcal{H};\pi^{\mathrm{sel}}) = I_T L_T(\mathcal{H};\pi_{\mathrm{inf}}) + (1-I_T)L_T(\mathcal{H};\pi_{\mathrm{NI}}).$$

Choose a common full-measure historical set for the two conditional bounds. On $E_T \cap G_T$, Lemma 12 gives

$$L_T(\mathcal{H};\pi_{\mathrm{inf}}) \le C_{\log}(1+\log T) + T\{s + \eta^2/(2\mu_0)\} \le Ka_T.$$

The conditional no-information lemma gives, including G_T^c,

$$L_T(\mathcal{H};\pi_{\mathrm{NI}}) \le C_{\mathrm{NI}}\sqrt{T} \le Kb_T.$$

If $I_T = 1$, the mixture is at most Ka_T. If $I_T = 0$, either G_T^c ($a_T = +\infty$) or G_T with $a_T > b_T$, so it is at most Kb_T. Thus on E_T,

$$L_T(\mathcal{H};\pi^{\mathrm{sel}}) \le K\min\{a_T, b_T\}.$$

On E_T^c, every branch remains box-valued and follows the mechanical-service rule. Pathwise,

$$0 \le p_t^\top \widehat{D}_t \le nUD_{\max} =: R_{\max}, \qquad 0 \le V^{\mathrm{Fluid}}(C_T) \le TR_{\max},$$

so $L_T(\mathcal{H};\pi^{\mathrm{sel}}) \le TR_{\max}$. The tower property gives

$$\mathrm{Reg}_T^{\mathrm{fl}}(\pi^{\mathrm{sel}}) \le K\mathbb{E}[\mathbf{1}_{E_T}\min\{a_T,b_T\}] + TR_{\max}\Pr(E_T^c) \le K\mathbb{E}\min\{a_T,b_T\} + TR_{\max}\delta_T.$$

This proves (13).

For the aligned specialization, assume $\Pr(E_T \cap G_T^c) = 0$. On E_T,

$$a_T \le \left(c_{\mathrm{ref}} + \frac{c_{\mathrm{samp}}^2}{\mu_0}\right)(1+\log T) + \frac{c_\eta^2}{\mu_0}\epsilon_0^2 T$$

by $(x+y)^2 \le 2x^2 + 2y^2$. With the stated C_{inf}/K,

$$a_T \le (C_{\mathrm{inf}}/K)\{\epsilon_0^2 T + 1 + \log T\}, \qquad b_T = c_{\mathrm{ref}}\sqrt{T} \le (C_{\mathrm{inf}}/K)\sqrt{T}.$$

Taking the minimum and using the preceding coverage decomposition proves (16). □

For the observable eligibility event G_T, $\mathbb{E}\min\{a_T,b_T\} = b_T\Pr(G_T^c) + \mathbb{E}[\mathbf{1}_{G_T}\min\{a_T,b_T\}]$. Thus, if a geometric check fails, the policy incurs the no-information cost even when all statistical bounds hold. Uniform regret constants over instances require common primitive NI bounds, including its initial demand-image margin.

I Rate for One Product

The one-product bound below complements the network construction of Appendix F. Unlike that construction, its hard instances let the distance between the anchor and the optimum vanish.

Proposition 5 (Minimax rate for one product)**.** *Fix the one-product class* $f_\eta(p) = 1/2 - p/4 + \eta(p-1)$, $|\eta| \le 1/16$, *on* $p \in [1/2, 3/2]$, *with Bernoulli requests, one unit of resource per request, and capacity* T. *Every model has the same exact informed observation* $(p^0, d^0) = (1, 1/4)$. *For every integer* $T \ge 1$, *put* $\delta = T^{-1/4}/16$. *Every randomized nonanticipating policy, even when told that the anchor error is exactly zero, satisfies* $\max_{\eta\in\{-\delta,\delta\}} \mathrm{Reg}_T^{\mathrm{fl}}(\pi; f_\eta) \ge \frac{\sqrt{T}}{4096}$. *The family is fixed independently of* T. *Only the hard pair varies with the horizon. Consequently the same lower bound holds for every supplied valid anchor-error radius, including zero.*

Proof. Write $f_v(p) = 1/2 - p/4 + v\delta(p-1)$ and $k_v = 1/4 - v\delta$ for $v = \pm 1$. Request means lie in $[3/32, 13/32]$, $k_v \in [3/16, 5/16]$, and capacity T serves every Bernoulli request. The optima $p_v^* = 1 + v\delta/(2k_v)$ are interior with uniform image margin, and $R_v = \mathbb{E}_v \sum_t k_v(p_t - p_v^*)^2$. Wrong-side prices cost at least $\delta^2/2$. Include parameter-independent policy randomization in the filtration. Common action kernels contribute no conditional KL. $\mathrm{kl}(x,y) \le (x-y)^2/[y(1-y)]$ gives $\mathrm{KL}(P_+, P_-) \le 64\delta^2\mathbb{E}_+ \sum_t (p_t - 1)^2 \le 1024\delta^2 R_+ + 1024\delta^4 T$. We now use this bound to prove the regret lower bound. If $R_+ \ge \delta^2 T$ the result follows. Otherwise $\delta^4 T = 16^{-4}$ gives KL at most $1/32$ and total variation below $1/4$ (affinity $a \ge e^{-\mathrm{KL}/2}$ and TV $\le \sqrt{1-a^2}$). For $E = \{\sum_t \mathbf{1}_{p_t \ge 1} \ge T/2\}$, each model's wrong-side count yields $R_- + R_+ \ge \delta^2 T\{P_-(E) + P_+(E^c)\}/4 \ge 3\delta^2 T/16$. Thus $\max_v R_v \ge \delta^2 T/16 = \sqrt{T}/4096$. The fixed family contains the horizon-dependent pair. Its common exact anchor makes every nonnegative error radius valid. □

J Supplementary Details for Surrogate-Assisted Learning

J.1 Supplementary Details for Section 6

Definition 23 (Surrogate policy and calibration)**.** Policy π^{surr} uses (17) only for exploration regression and freezes its projected $\widehat{\Theta}_L^{\mathrm{cv}}$ thereafter. All capacity, service, revenue and residual computations use actual requests.

The nuisance estimates are predictable and clipped as in Lemma 13. Let $\mathcal{E}_{\mathrm{cal}} \in \mathcal{F}^0$ be an ex ante calibration event with $\mathbb{P}(\mathcal{E}_{\mathrm{cal}}^c) \le \delta_{\mathrm{cal}} \le T^{-3}$. Assume that the pre-campaign calibration procedure supplies deterministic envelopes $\eta_{m,L}$ and $\eta_{\Gamma,L}$ such that, on $\mathcal{E}_{\mathrm{cal}}$, their errors at the exploration prices satisfy $\max_{s\le L_T} \|\mu_S(\boldsymbol{p}^s) - \widehat{m}_s(\boldsymbol{p}^s)\|_2 \le \eta_{m,L}$, $\frac{1}{L_T}\sum_{s\le L_T} \|\widehat{\Gamma}_s(\boldsymbol{p}^s) - \Gamma^\star(\boldsymbol{p}^s)\|^2_{\Sigma_S(\boldsymbol{p}^s)} \le \eta_{\Gamma,L}^2$. Here $\|G\|^2_{\Sigma_S(p)} = \mathrm{tr}(G\Sigma_S(p)G^\top)$. Let $D_{\mathfrak{T}} := \sup_{\Theta_1,\Theta_2\in\mathfrak{T}} \|\Theta_1 - \Theta_2\|_F < \infty$. Define $\bar{e}_L^2 := \frac{(1-\rho_\star^2)\sigma^2 + \eta_{\Gamma,L}^2}{L_T} + \eta_{m,L}^2 + D_{\mathfrak{T}}^2\delta_{\mathrm{cal}}$. Assume also that the corresponding deterministic calibration radius in (75) satisfies $r_{L,T}\log(T/h_T) \le c_{\mathrm{sur}}$, where $c_{\mathrm{sur}} > 0$ is the sufficiently small instance constant in Lemma 14.

Calibration and projection. The oracle covariance projection (Glasserman 2004, Ch. 4) gives

$$\mathrm{Var}(\xi_t^\star \mid p^t = p, \mathcal{F}^{t-1}) = \Sigma_d(p) - \Sigma_{dS}(p)\Sigma_S(p)^{-1}\Sigma_{Sd}(p). \tag{72}$$

This covariance matrix is bounded above by Σ_d in the positive semidefinite ordering. The effective-noise proxy is a separate assumption.

J.1.1 Fixed-Design Surrogate Calibration

Predictable coefficient and center estimates are evaluated at the covered design prices. Calibration errors, including regularization bias, enter the envelopes below. At exploration prices set $x_s = z(p^s)$, $V_L = \sum_s x_s x_s^\top \succeq \kappa_0 L_T I$ and form (17). Freeze

$$\widetilde{\Theta}_L^{\mathrm{cv}} = (\sum_s \widetilde{D}^s x_s^\top) V_L^{-1}, \qquad \widehat{\Theta}_L^{\mathrm{cv}} = \Pi_{\mathfrak{T}}(\widetilde{\Theta}_L^{\mathrm{cv}}). \tag{73}$$

Define $\delta_s^m = \mu_S(p^s) - \widehat{m}_s(p^s)$, $\delta_s^\Gamma = \widehat{\Gamma}_s(p^s) - \Gamma^\star(p^s)$ and $\xi_s^\star = \xi_s - \Gamma^\star(p^s) Z^s$. Then

$$\widetilde{D}^s - f(p^s) = \xi_s^\star - \delta_s^\Gamma Z^s - \widehat{\Gamma}_s(p^s)\delta_s^m. \tag{74}$$

The first two terms are centered. The last is bias. The controller uses physical requests.

Lemma 13 (Fixed-design surrogate estimation)**.** *Under Assumptions 2(a), 6(a) and 7, cycle the design in (5) for an integer multiple $L_T \geq J_0$. All design prices lie in $\mathcal{S}_{\mathrm{sur}}$. Use predictable nuisances with $\|\widehat{\Gamma}_s\|_{\mathrm{op}} \leq M_\Gamma$, exact Frobenius projection and class diameter $D_{\mathfrak{T}}$. An ex-ante event $\mathcal{E}_{\mathrm{cal}} \in \mathcal{F}^0$ has complement probability at most δ_{cal}. On this event deterministic envelopes satisfy $\max_{s \leq L_T} \|\delta_s^m\| \leq \eta_{m,L}$ and $L_T^{-1} \sum_{s \leq L_T} \|\delta_s^\Gamma\|^2_{\Sigma_S(p^s)} \leq \eta^2_{\Gamma,L}$. Here $\|G\|^2_{\Sigma_S(p)} = \mathrm{tr}(G\Sigma_S(p)G^\top)$ and $\upsilon_L^2 = (1-\rho_\star^2)\sigma^2 + \eta^2_{\Gamma,L}$. Put $\bar{e}_L^2 = \upsilon_L^2/L_T + \eta^2_{m,L} + D^2_{\mathfrak{T}}\delta_{\mathrm{cal}}$. Then $\mathbb{E}\|\widehat{\Theta}_L^{\mathrm{cv}} - \Theta\|_F^2 \leq C\bar{e}_L^2$, and for every $\delta \in (0, 1/2)$,*

$$r_L(\delta) = C\{\upsilon_L \sqrt{[2n + 1 + \log(n/\delta)]/L_T} + \eta_{m,L}\}, \qquad \Pr(\|\widehat{\Theta}_L^{\mathrm{cv}} - \Theta\|_F > r_L(\delta)) \leq \delta + \delta_{\mathrm{cal}}. \tag{75}$$

Write $r_{L,T} = r_L(T^{-3})$. C depends only on fixed dimensions, κ_0, $\sup_p \|z(p)\|$, M_Γ and primitive moments, and not on L_T. The calibration complement contributes at most $D^2_{\mathfrak{T}}\delta_{\mathrm{cal}}$.

Proof. Put $L = L_T$, $E = \mathcal{E}_{\mathrm{cal}}$, $b_s = V_L^{-1}x_s$, $C_z = \sup_p \|z(p)\|$, $\nu_0^2 = (1-\rho_\star^2)\sigma^2$ and $h_s = \xi_s^\star - \delta_s^\Gamma Z^s$. Then $\|b_s\| \leq C_z/(\kappa_0 L)$ and $\sum_s b_s b_s^\top = V_L^{-1}$. Oracle orthogonality gives $\mathbb{E}[\xi_s^\star (Z^s)^\top \mid \mathcal{F}^{s-1}] = 0$ and $\mathbb{E}[\|h_s\|^2 \mid \mathcal{F}^{s-1}] \leq n\nu_0^2 + \|\delta_s^\Gamma\|^2_{\Sigma_S(p^s)}$. Since $E \in \mathcal{F}^0$, cross-period terms vanish even after multiplying by $\mathbf{1}_E$. Summing diagonal terms yields $\mathbb{E}[\mathbf{1}_E \|\sum_s h_s b_s^\top\|_F^2] \leq C(\nu_0^2 + \eta^2_{\Gamma,L})/L$. The bias norm is at most $M_\Gamma C_z \eta_{m,L}/\kappa_0$. Projection cannot increase error to the true member and bounds squared error by $D^2_{\mathfrak{T}}$ on E^c, proving the moment claim.

We now derive the tail bound. Let σ_S^2 be a fixed conditional directional proxy for Z^s from Assumption 7(c). Conditional Cauchy–Schwarz gives $h_{s,i}$ proxy $w_{s,i} = 2\nu_0^2 + 2\sigma_S^2\|(\delta_s^\Gamma)^\top e_i\|^2$. For a deterministic unit $u \in \mathbb{R}^{n+1}$ set $S_i = \sum_s (u^\top b_s) h_{s,i}$ and $A_i = \sum_s (u^\top b_s)^2 w_{s,i}$. The average envelope and $\Sigma_S \succeq \lambda_S I$ give, on E,

$$A_i \leq v_L := \frac{2\nu_0^2}{\kappa_0 L} + \frac{2\sigma_S^2 C_z^2 \eta^2_{\Gamma,L}}{\lambda_S \kappa_0^2 L}.$$

Iterated conditional mgfs give $\mathbb{E}[\mathbf{1}_E \exp\{\lambda S_i - \lambda^2 A_i/2\}] \leq 1$. On $E \cap \{S_i \geq a\}$ the exponential is at least $\exp\{\lambda a - \lambda^2 v_L/2\}$. Markov inequality with $\lambda = a/v_L$ and the negative-sign argument give $\Pr(E \cap \{|S_i| > \sqrt{2v_L x}\}) \leq 2e^{-x}$. If $v_L = 0$, the moment calculation gives $S_i = 0$ on E. A $1/2$-net with at most 5^{n+1} points and n rows uses $x = \log(2n5^{n+1}/\delta)$. Row norms are at most twice the directional maximum. Add bias and project. Put $B_m = M_\Gamma C_z/\kappa_0$ and $C_v = \max\{2/\kappa_0, 2\sigma_S^2 C_z^2/(\lambda_S \kappa_0^2)\}$. A common constant in both bounds is

$$C = \max\{1, 2n(n+1)/\kappa_0, 2C_z^2/\kappa_0^2, 2B_m^2, 4\sqrt{nC_v}, B_m\}.$$

Indeed $v_L \leq C_v \upsilon_L^2/L$ and $\log(2n5^{n+1}/\delta) \leq 2\{2n + 1 + \log(n/\delta)\}$. Adding $\Pr(E^c)$ proves (75). □

Primitive sampling specialization. For $T \geq 2$, fix the same $J = J_0$ prices, $L = L_T = Jq$ and $F \geq 2$. Before sampling set $\ell_T = \log\{8(F+2)Jn^2T^3\}$, $k = \lceil L\ell_T \rceil$, $M = 2Fk$, $m_f = 2k$ and $C = k$ per price. Acquire independent request–score vectors in deterministic folds and an independent score-only center bank. Use their sample means. Paired request means equal campaign means. Center means equal campaign score means for online correction and historical score means for historical correction. Center and historical-request coordinates have proxies σ_c^2, σ_H^2. Historical and campaign score vectors have directional proxy σ_S^2. Put all samples and randomization in $\mathcal{F}^0$. The campaign conditional assumptions hold under this filtration, and matrices are clipped as before. Historical $\mathcal{T}_f$-measurable $G_{f,a}$ obeys $|G_{f,a}| \leq G_{\max}$ and excludes its held-out fold at all prices and every center observation. Learned pooling is unchanged.

Coordinate Chernoff with Jn terms gives $E_m = \{\max_j \|\bar{S}_{c,j} - \mu_{S,j}\| \leq \eta_{m,L}\} \in \mathcal{F}^0$, $\Pr(E_m^c) \leq 2Jne^{-\ell_T} \leq T^{-3}$, where

$$\eta_{m,L} = \sigma_c\sqrt{2n\ell_T/C}, \qquad \eta_{\Gamma,L}^2 = 4nM_\Gamma^2\sigma_S^2.$$

The coefficient bound follows from $\Sigma_S \preceq \sigma_S^2 I$ and $\|\widehat{\Gamma} - \Gamma^\star\|_F \leq 2\sqrt{n}M_\Gamma$. Clipping enforces it. Thus the radius below is $O(\sqrt{\log T/L})$ and satisfies the product gate eventually under the retained square-root schedule. The design is independent of the analysis-only c_{sur}. A numerical finite-horizon check requires valid proxy and geometric bounds.

For historical correction, require independent pricewise pairs, stable request means, an independent matching center bank, deterministic folds and clipped training-fold coefficients. Online requests obey the enlarged-filtration proxy. Historical correction accommodates unobserved online scores and singular score covariance. Take $g_a = 0$, $\gamma_{f,a} = G_{\max}$ on the whole sample space. Cauchy–Schwarz gives the paired covariance envelope $\mathrm{diag}(2\sigma_H^2, 2\sigma_S^2)$. Set $b_{0,a} = 0$, $v_{c,a} = \sigma_c^2$, $\delta_H = \delta_O = T^{-3}$. The existing risk formulas and the $Jn(F+2)$ historical and Jn online tail unions give

$$R_{\mathrm{lin},a} \leq R_*/k, \qquad R_* := \{\sigma_H/\sqrt{F} + G_{\max}\sqrt{\sigma_S^2 + \sigma_c^2}\}^2,$$
$$b_{H,a} \leq B_*/\sqrt{L}, \quad B_* := \sigma_H/\sqrt{F} + G_{\max}\sigma_S + \sqrt{2}G_{\max}\sigma_c.$$

Because $\Omega_a \leq (2FJ\ell_T)^{-1}$, the actual learned-weight radius obeys

$$r_{\mathrm{lin}}(\mathcal{H}) \leq \min\left\{D_{\mathfrak{T}}, \frac{k_{\mathrm{op}}\sqrt{Jn}}{\sqrt{L}}\left(B_* + \frac{\sigma}{F\sqrt{2J\ell_T}}\right)\right\}.$$

This gate also holds eventually. Charge JM pairs, JC centers and any newly bought request labels in (19). Only shared existing labels are sunk. Acquisition cost is $O(\sqrt{T}\log T)$ at fixed unit costs. Fixed center bias retains its nonvanishing term. The result uses positive q, the original controller and exact Frobenius parameter projection.

Frozen estimate.

Lemma 14 (Regret with a frozen estimate)**.** *Use Definitions 4 and 5, Assumptions 1, 2, 5 and 6(a), and Definition 1. At deterministic $s < T$, freeze a $\mathcal{G}_s$-measurable class-valued estimate and solve at $b = c^{s+1}/N$, $N = T - s$. Fix $1 \leq h < N$, the $\mathcal{G}_s$-measurable fixed price $p_s^{\mathrm{F}} \in \mathcal{P}$, and the actual controller in (25), or any controller that posts the same prices as that controller in every period before the first failure of that controller's active test (24) and, from that period on and in the terminal window $t > T - h$, posts arbitrary $\mathcal{G}_{t-1}$-measurable prices in $\mathcal{P}$ (the conclusion covers every such controller). Put*

$$e_s := \|\widehat{\Theta}_s - \Theta\|_F, \quad \Delta_s := \|c^{s+1}/N - c_0\|_\infty, \quad \varepsilon_s := e_s + \sqrt{\Delta_s}.$$

The lemma takes as inputs a nonnegative $\mathcal{G}_s$-measurable radius r_s, a deterministic $\Delta_{\max} \in [0,\infty)$, and a $\mathcal{G}_s$-measurable good event $\mathcal{B}_s$ on which all of the following hold:

$$e_s \le r_s, \quad \Delta_s \le \Delta_{\max} \wedge c_\Delta, \quad r_s \le r_{\mathrm{loc}},$$

$$A\sum_{t=1}^{s} d^t \le C_T, \quad \widehat{d}^t = d^t \ (t \le s), \quad \text{the geometry gate (22) holds,}$$

$$c^{s+1} = C_T - A\sum_{t=1}^{s} d^t, \qquad c^{s+1} - NAx \ge 0.$$

Let $\delta_s := \Pr(\mathcal{B}_s^c)$ be an input of the lemma and assume $\mathbb{E}e_s^2 \le M_s$ and $\Pr(e_s > r_s) \le \delta_{\mathrm{est},s}$. Define the realized initial-fluid loss

$$\mathcal{R}_T := V^{\mathrm{Fluid}}(C_T) - \sum_{t=1}^{T} (p^t)^\top \widehat{d}^t.$$

For the virtual process in (26), under the enlarged-filtration condition (23), let $\pi_{\mathrm{ex},s} := \Pr(\mathcal{E}_s \mid \mathcal{G}_s)$. With constants depending only on the fixed primitive bounds and the displayed gate,

$$\begin{aligned}\mathbf{1}_{\mathcal{B}_s}\mathbb{E}[\mathcal{R}_T \mid \mathcal{G}_s] \le \mathbf{1}_{\mathcal{B}_s}\Bigg[& C_1\{s + h + \log(N/h)\} + C_\lambda N\Delta_s + C_2 N\varepsilon_s^2 \\ & + C_3 h\varepsilon_s \sum_{k=h}^{N-1} k^{-1} + C_{\mathrm{exit}} N\sqrt{\pi_{\mathrm{ex},s}}\Bigg].\end{aligned} \tag{76}$$

Averaging (76) adds the complement charge $T\bar{R}\delta_s$. Since capacities and the gate make $0 \le \Delta_s \le \Delta_{\max}$,

$$\mathbb{E}[\mathbf{1}_{\mathcal{B}_s}\varepsilon_s^2] \le 2M_s + 2\Delta_{\max}, \qquad \mathbb{E}[\mathbf{1}_{\mathcal{B}_s}\varepsilon_s] \le \sqrt{2M_s + 2\Delta_{\max}}.$$

Define the integrated exit term

$$\overline{\pi}_{\mathrm{ex},s} := \mathbb{E}\left[\mathbf{1}_{\mathcal{B}_s}\sqrt{\Pr(\mathcal{E}_s \mid \mathcal{G}_s)}\right]. \tag{77}$$

Then averaging the conditional inequality gives the same display with $C_{\mathrm{exit}}N\overline{\pi}_{\mathrm{ex},s}$ in place of the conditional exit term, and with $\mathbb{E}[\mathbf{1}_{\mathcal{B}_s}\Delta_s]$, $\mathbb{E}[\mathbf{1}_{\mathcal{B}_s}\varepsilon_s^2]$ and $\mathbb{E}[\mathbf{1}_{\mathcal{B}_s}\varepsilon_s]$ in the corresponding terms. A deterministic upper bound on $\overline{\pi}_{\mathrm{ex},s}$ may be substituted when available.

Proof. Fix a freeze history in $\mathcal{B}_s$ and abbreviate $e = e_s$, $\Delta = \Delta_s$, $\varepsilon = \varepsilon_s$, $b = c^{s+1}/N$. The gate gives $\|x - x^*\|_2 \le r_0/4$ with $x^* = d_0$, so x lies in the common ball $B(d_0, r_0) \subseteq \mathcal{D}_\vartheta$ of Definition 4, and $c^{s+1} - NAx \ge 0$ gives $Ax \le b$. Hence $x \in K_\vartheta(b) \ne \varnothing$, and v is the true maximizer over $K_\vartheta(b)$. For positive rows set $a_A := \min_{i:A_i\mathbf{1}>0} A_i\mathbf{1}$ and $\tau_b := \Delta/a_A$. Set $\tau_b := 0$ if all rows are zero. For $\eta \in \{\vartheta, \widehat{\vartheta}\}$, $w \in K_\eta(b)$ gives $A_i(w - \tau_b\mathbf{1}) \le b_i - \tau_b A_i\mathbf{1} \le b_i - \Delta \le c_{0,i}$, while $w \in K_\eta(c_0)$ gives $A_i(w - \tau_b\mathbf{1}) \le c_{0,i} - \Delta \le b_i$. Zero rows vanish. Thus the shift maps each of the two capacity constraints into the other. Membership in the demand images is checked only at the two points below. The geometry gate gives $\|w - x\| + \sqrt{n}\,\tau_b \le r_0/2$ and $\|x - d_0\| \le r_0/4$ for $w \in \{x, v\}$, hence $\|w - \tau_b\mathbf{1} - d_0\| \le \|w - x\| + \sqrt{n}\,\tau_b + \|x - d_0\| \le 3r_0/4 < r_0$. By the common-ball condition, the shifted point lies in both images with both inverse prices in $\mathcal{P}$, which yields both comparisons (including zero rows) and gives $\|a\| \le C_a e + C_a\sqrt{\Delta} = C_a\varepsilon$.

We now show that the actual and virtual processes coincide before the terminal window when no exit occurs. Keep y_t for inventory and put $z_t := q_t - x$, where $q_t = y_t/(T - t + 1)$. On an active period, $D_t = g(q_t) + \xi_t = q_t + a + Hz_t + \xi_t$ and

$$z_{t+1} = \left(I - \frac{H}{T-t}\right) z_t - \frac{a + \xi_t}{T-t}, \quad z_{s+1} = 0, \tag{78}$$

while the virtual recursion (27) holds on every path. Starting from $y_{s+1} = Nx$, induction on an active prefix gives

$$\sum_{v=s+1}^{t} D_v = Nx - (T-t)q_{t+1}, \qquad s+1 \le t \le T-h, \tag{79}$$

and, with $\widetilde{c}^{t+1} := c^{s+1} - \sum_{v=s+1}^{t} AD_v$,

$$\widetilde{c}^{t+1} = (c^{s+1} - NAx) + (T-t)Aq_{t+1} \ge 0.$$

The first term is nonnegative by fitted feasibility and the second on $\mathcal{E}_s^c$ because the common image ball is nonnegative and $A \ge 0$. Hence $AD_t \le c^t$, so the fulfillment rule serves each active request completely and closes the induction. From $q_{s+1} = x = \widetilde{q}_{s+1}$, actual and virtual tests, prices, requests, loads, service and residuals coincide through $T-h$ on $\mathcal{E}_s^c$.

We now bound the moments of the virtual process and its exit probability. Let $S := \sum_{k=h}^{N-1} k^{-1} \le 2\log(N/h)$. The gate gives $\|H\|S \le \log 2$ and $2C_a\varepsilon S \le r_0/8$. For $k_t := T-t+1$, set

$$\Phi_{t,v} := \prod_{u=v+1}^{t-1} (I - H/(T-u)), \quad \Phi_{v+1,v} := I, \quad \|\Phi_{t,v}\|_{\mathrm{op}} \le \exp\{\|H\| \textstyle\sum_{k=k_t}^{N-1} k^{-1}\} \le 2,$$

$$\widetilde{z}_t = b_t + W_t := -\sum_{v=s+1}^{t-1} \frac{\Phi_{t,v}(a+\xi_v)}{T-v}.$$

For $s+1 \le t \le T-h+1$ (so $k_t \ge h$) and coordinate i, $w_{t,v,i} := -\Phi_{t,v}^{\top} e_i/(T-v)$ is $\mathcal{G}_{v-1}$-measurable and $\sum_v \|w_{t,v,i}\|_2^2 \le 4\sum_v (T-v)^{-2} \le 8/k_t$. The conditional Chernoff bound therefore gives each $W_{t,i}$ proxy $v_t := 4\sigma^2 \sum_v (T-v)^{-2} \le 8\sigma^2/k_t \le 8\sigma^2/h$, over the full path space, while $\|b_t\| \le 2C_a\varepsilon \sum_{k=k_t}^{N-1} k^{-1} \le r_0/8$. Hence, on all paths,

$$\mathbb{E}[\|\widetilde{z}_t\|^2 \mid \mathcal{G}_s] \le C\{\varepsilon^2 \log^2(N/k_t) + \sigma^2/k_t\}, \qquad \sum_{t=s+1}^{T-h} \mathbb{E}\|\widetilde{z}_t\|^2 \le C\{N\varepsilon^2 + \sigma^2 \log(N/h)\}.$$

Since $\|x - x^*\| \le r_0/4$, exit requires $\|W_t\| \ge r_0/8$. The coordinate threshold $r_0/(8\sqrt{n})$ gives

$$\frac{(r_0/(8\sqrt{n}))^2}{2(8\sigma^2/h)} = \frac{r_0^2 h}{1024 n\sigma^2}, \qquad \pi_{\mathrm{ex},s} \le \min\{1, 2nNe^{-r_0^2 h/(1024 n\sigma^2)}\},$$

with $\pi_{\mathrm{ex},s} = 0$ for $\sigma = 0$. Tail integration gives $\mathbb{E}[X^4 \mid \mathcal{G}_s] \le 16v^2$ for scalar v-sub-Gaussian X, so $\mathbb{E}[\|W_t\|^4 \mid \mathcal{G}_s] \le 16n^2 v_t^2$ and $\mathbb{E}[\|\widetilde{z}_t\|^4 \mid \mathcal{G}_s] \le C\{\varepsilon^4 \log^4(N/k_t) + \sigma^4/k_t^2\}$. The same integration gives $\mathbb{E}[(w^{\top}\xi_t)^4 \mid \mathcal{G}_{t-1}] \le 16\sigma^4$ and $\mathbb{E}\|\xi_t\|^4 \le 16n^2\sigma^4$. Compactness and $\|B'^{-1}\| \le 2/\kappa_B$ give $\|\widetilde{p}_t\|, \|g(\widetilde{q}_t)\| \le C_{\mathrm{aff}}(1 + \|\widetilde{z}_t\|)$. With $Z_4 := 27[1 + (r_0/8)^4 + 16n^2(8\sigma^2)^2]$, conditional Cauchy–Schwarz gives $\mathbb{E}[(1 + \|\widetilde{z}_t\|)^2\|\xi_t\|^2 \mid \mathcal{G}_s] \le \sqrt{Z_4} \cdot 4n\sigma^2$. Thus $C_X^2 := 2C_{\mathrm{aff}}^4 Z_4 + 8n\sigma^2 C_{\mathrm{aff}}^2 \sqrt{Z_4}$ obeys $\mathbb{E}[|\widetilde{X}_t|^2 \mid \mathcal{G}_s] \le C_X^2$, and $\mathbb{E}[(\sum_{t=s+1}^{T-h} |\widetilde{X}_t|)^2 \mid \mathcal{G}_s] \le C_X^2 (N-h)^2$.

For actual revenue $Y_t := (p^t)^{\top}\widehat{d}^t$, nonnegative prices and mechanical service give $0 \le Y_t \le U\|d^t\|_1$ and $\mathbb{E}[Y_t^2 \mid \mathcal{G}_{t-1}] \le U^2\mathbb{E}[\|d^t\|_1^2 \mid \mathcal{G}_{t-1}] \le C_{\mathrm{act}}$. Thus its active-prefix L^2 norm is at most $C_{\mathrm{act}}^{1/2} N$. Second moments suffice for this bound. Define the virtual loss by replacing only active-prefix revenue in $\mathcal{R}_T$ by $\widetilde{X}_t$, retaining actual prefix and fallback/terminal service:

$$\mathcal{R}_T^{\mathrm{virt}} := V^{\mathrm{Fluid}}(C_T) - \sum_{t \le s} (p^t)^{\top}\widehat{d}^t - \sum_{t=s+1}^{T-h} \widetilde{X}_t - \sum_{t > T-h} (p^t)^{\top}\widehat{d}^t.$$

On $\mathcal{E}_s^c$, $\mathcal{R}_T = \mathcal{R}_T^{\mathrm{virt}}$, so $\mathcal{R}_T = \mathcal{R}_T^{\mathrm{virt}} + (\mathcal{R}_T - \mathcal{R}_T^{\mathrm{virt}})\mathbf{1}_{\mathcal{E}_s}$. Minkowski gives the two conditional L^2 norms at most $N\sqrt{C_{\mathrm{act}}}$ and NC_X. Conditional Cauchy–Schwarz gives

$$\mathbb{E}[|(\mathcal{R}_T - \mathcal{R}_T^{\mathrm{virt}})\mathbf{1}_{\mathcal{E}_s}| \mid \mathcal{G}_s] \le C_{\mathrm{exit}} N \sqrt{\pi_{\mathrm{ex},s}}, \qquad C_{\mathrm{exit}} := \sqrt{C_{\mathrm{act}}} + C_X. \tag{**}$$

We now bound the expected virtual loss. For the virtual process put $\widetilde{r}_t := g(\widetilde{q}_t)$ and $\ell(y) := (B^{-1}(y - \alpha))^\top y$. Since $\widehat{p}(q) = B^{-1}(g(q) - \alpha)$, $\mathbb{E}[\widetilde{X}_t \mid \mathcal{G}_{t-1}] = \ell(g(\widetilde{q}_t))$ on all paths. Expectations extend over the full path space; the only charge associated with $\mathcal{E}_s^c$ is (**). With $m_t := \mathbb{E}[x - \widetilde{q}_t \mid \mathcal{G}_s]$, recursion and centering give

$$\mathbb{E}[x - g(\widetilde{q}_t) \mid \mathcal{G}_s] = m_t - (T - t)(m_{t+1} - m_t),$$

$$\sum_{t=s+1}^{T-h} \mathbb{E}[x - g(\widetilde{q}_t) \mid \mathcal{G}_s] = -h\, m_{T-h+1}, \qquad \|m_{T-h+1}\| \le 2C_a \varepsilon S,$$

where $m_{s+1} = 0$. Put $Q := -\nabla^2 \ell/2 \succeq 0$. The KKT gate $A_J x^* = c_{0,J}$, $A_J x = b_J$ gives $\ell(x^*) - \ell(y) = -\lambda_J^\top A_J(y - x^*) + (y - x^*)^\top Q(y - x^*)$ and

$$\sum_{t=s+1}^{T-h} \mathbb{E}[\lambda_J^\top A_J(x^* - \widetilde{r}_t) \mid \mathcal{G}_s] = (N - h)\lambda_J^\top (c_{0,J} - b_J) - h\lambda_J^\top A_J m_{T-h+1}.$$

The first term is at most $C_\lambda N\Delta$, where, for $J \neq \varnothing$,

$$\lambda_J = (A_J A_J^\top)^{-1} A_J \nabla \ell(x^*), \qquad C_\lambda := \sqrt{m}\, \|\nabla \ell\|_{2,\max}/s_J \ge \|\lambda_J\|_1;$$

for $J = \varnothing$ set $\lambda_J = C_\lambda = 0$. The second term is at most $C_3 h\varepsilon S$. Pointwise and after summing,

$$\|g(\widetilde{q}_t) - x^*\|_2^2 \le 3\{\|g(\widetilde{q}_t) - \widetilde{q}_t\|_2^2 + \|\widetilde{q}_t - x\|_2^2 + \|x - x^*\|_2^2\}, \quad \sum_t \mathbb{E}\|g(\widetilde{q}_t) - \widetilde{q}_t\|_2^2 \le C\{N\varepsilon^2 + \log(N/h)\}.$$

Indeed the left sum is bounded by $2NC_a^2\varepsilon^2 + 2C_H^2 e^2 \sum_t \mathbb{E}\|\widetilde{z}_t\|_2^2$. Together with the all-path second-moment bound and $\|x - x^*\| \le C_x\varepsilon$, the quadratic KKT term is at most $C_2 N\varepsilon^2 + C\log(N/h)$. Since $V^{\mathrm{Fluid}}(C_T) = T\ell(x^*)$ and $Y_t \ge 0$, the omitted prefix and last periods contribute at most $(s + h)\ell(x^*) \le (s + h)\bar{R}$. Adding these terms and (**) proves (76). All virtual moments were established on the full path space. Finally, $Y_t \ge 0$ and $V^{\mathrm{Fluid}}(C_T) \le T\bar{R}$ give $\mathcal{R}_T \le T\bar{R}$ on $\mathcal{B}_s^c$, so averaging charges $T\bar{R}\delta_s$ and the displayed $\Delta_{\max}$ inequalities give the stated averaged bounds. □

Deterministic exploration specialization. Under the primitive hypotheses of Lemma 14, use Algorithm 2's fixed design, $s = L_T$, $h = h_T$, a frozen $\mathcal{G}_{L_T}$-measurable estimator with $\mathbb{E}e_L^2 \le C\bar{e}_L^2$, and the stated geometry, KKT, radius and enlarged-filtration gates. Physical requests, service and (9) are unchanged. Pseudo-observations enter only estimation. The load and estimator tails give $\Pr(\mathcal{B}_{L_T}^c) \le T^{-3} + \delta_{\mathrm{est}}$, and the virtual exit probability is

$$\overline{\pi}_{\mathrm{ex},L_T} \le \min\{1, CnTe^{-ch_T}\}. \tag{80}$$

Averaging the bound of Lemma 14 gives $C\{L_T + h_T + \log(T/h_T) + T\bar{e}_L^2 + h_T\bar{e}_L \log(T/h_T)\}$ plus $T\bar{R}(T^{-3} + \delta_{\mathrm{est}})$ and the displayed exit charge. Complete fallback and a larger instance constant cover $L_T + h_T \ge T$ and the finitely many smaller horizons.

J.1.2 Historical-corrected linear estimator

At J identifying prices, partition M request–signal pairs into nonempty deterministic folds I_f with $m_f = |I_f|$, $w_f = m_f/M$. Let q online requests and C independent center observations be available, with $M, C, q > 0$ and $1 \le F \le M$. For $a = (j, i)$, $\mathcal{T}_f$ contains all observations outside fold f and pre-campaign randomization, and the $\mathcal{T}_f$-measurable $G_{f,a}$ obeys $|G_{f,a}| \le G_{\max}$ and is independent of its held-out fold and center sample. Define

$$H_a = \sum_f w_f\{\bar{D}_{f,a} - G_{f,a}(\bar{S}_{f,a} - \bar{S}_{c,a})\}, \qquad Y_a = q^{-1}\sum_{\ell=1}^{q} D_{a,\ell}^{\mathrm{on}}. \tag{81}$$

Take deterministic $|g_a| \le G_{\max}$, $\gamma_{f,a} \ge 0$, and an ex-ante $\mathcal{E}_{\mathrm{cal}}$ with failure at most δ_{cal}, on which $|G_{f,a} - g_a| \le \gamma_{f,a}$. Historical means are $(\mu_a, \mu_{S,a})$ with covariance dominated by $\left(\begin{smallmatrix} v_{D,a} & c_a \\ c_a & v_{S,a} \end{smallmatrix}\right)$. Each independent center observation has mean $\mu_{S,a} + b_a$, where $|b_a| \le b_{0,a}$, and variance at most $v_{c,a}$. These bounds hold in particular when the moments are known exactly. Set

$$\begin{aligned} R_{0,a} &:= \frac{v_{D,a} - 2g_a c_a + g_a^2 v_{S,a}}{M} + g_a^2\left(\frac{v_{c,a}}{C} + b_{0,a}^2\right), \\ \Delta_{\mathrm{cf},a} &:= \sum_f w_f \gamma_{f,a}^2\left(\frac{v_{S,a}}{m_f} + \frac{v_{c,a}}{C} + b_{0,a}^2\right), \\ R_{\mathrm{lin},a} &:= \left(\sqrt{R_{0,a}} + \sqrt{\Delta_{\mathrm{cf},a}}\right)^2. \end{aligned} \tag{82}$$

With centered residual, score and center proxies $s_a^2, \sigma_{S,a}^2, \sigma_{c,a}^2$, put $x_H = \log(2Jn(F+2)/\delta_H)$ and $x_O = \log(2Jn/\delta_O)$ for $\delta_H, \delta_O \in (0, 1/2)$ and set

$$\begin{aligned} b_{H,a} &:= s_a\sqrt{\frac{2x_H}{M}} + \sum_f w_f \gamma_{f,a}\sigma_{S,a}\sqrt{\frac{2x_H}{m_f}} + G_{\max}\left(b_{0,a} + \sigma_{c,a}\sqrt{\frac{2x_H}{C}}\right), \\ b_O &:= \sigma\sqrt{\frac{2x_O}{q}}. \end{aligned} \tag{83}$$

The conditional Chernoff inequality and the $Jn(F+2)$ and Jn unions give these radii with failures δ_H and δ_O.

Before observing history fix $\eta_{\mathrm{pool}}, a_{\min} \in [0, 1]$ and use the historical-measurable $\Omega_a \in [\omega_-, \omega_+]$, where $\omega_+ = q/(M+q)$, $\omega_- = qa_*/(M+q)$ and $a_* := 1 - \eta_{\mathrm{pool}} + \eta_{\mathrm{pool}} a_{\min}$. Put $\mu_a^{\mathrm{pool}} = (1-\Omega_a)H_a + \Omega_a Y_a$. Let K be the deterministic fixed-design least-squares map, with $k_{\mathrm{op}} = \|K\|_{\mathrm{op}}$, $k_F = \|K\|_F$, and let $D_{\mathfrak{T}}$ be the diameter of $\mathfrak{T}$. Set

$$\begin{aligned} \mathcal{M}_L &:= \min\left\{D_{\mathfrak{T}}^2,\ k_{\mathrm{op}}^2(1-\omega_-)^2\sum_a R_{\mathrm{lin},a} + \frac{\sigma^2}{q}\omega_+^2 k_F^2 + D_{\mathfrak{T}}^2\delta_{\mathrm{cal}}\right\}, \\ r_{\mathrm{lin}}(\mathcal{H}) &:= \min\left\{D_{\mathfrak{T}},\ k_{\mathrm{op}}\left[\sum_a\{(1-\Omega_a)b_{H,a} + \Omega_a b_O\}^2\right]^{1/2}\right\}. \end{aligned} \tag{84}$$

The fitted parameter is the exact Frobenius projection of the fixed-design fit based on μ_a^{pool}; with $x_j = z(\boldsymbol{p}^{(j)})$, $\widetilde{\Theta}_{L_T}^{\mathrm{lin}} := (\sum_j \boldsymbol{\mu}_j^{\mathrm{pool}} x_j^\top)(\sum_j x_j x_j^\top)^{-1}$ and $\widehat{\Theta}_{L_T}^{\mathrm{lin}} := \Pi_{\mathfrak{T}}(\widetilde{\Theta}_{L_T}^{\mathrm{lin}})$. Put $\delta_{\mathrm{est}} := \delta_{\mathrm{cal}} + \delta_H + \delta_O$.

Theorem 6 (Regret with a historical-corrected linear surrogate estimator)**.** *Assume all hypotheses of Theorem 2 and use its fixed-design schedule. In addition, assume:*

1. *Historical request–signal vectors are independent across observations and prices, products may be dependent within a vector, the center sample is independent of the historical sample, and each $G_{f,a}$ is $\mathcal{T}_f$-measurable with the bound above. At coordinate $a = (j,i)$ the historical request mean is the campaign mean $\mu_a = f_i(\boldsymbol{p}^{(j)})$. The center shift b_a is relative to the historical score mean. There are positive integer counts M, C, q and complete online design cycles with $L_T = Jq$. These are the stable-history and positive-online-repeat conditions of this theorem.*

2. *The calibration event and deterministic proxy bounds used in (82)–(83) hold with failure probability δ_{cal}.*

3. *With $\mathcal{H}$ as above and $\mathcal{G}_t = \mathcal{H} \vee \mathcal{F}_t$, every online innovation satisfies $\mathbb{E}[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}] = 0$, $\mathbb{E}[\exp\{\lambda v^\top \boldsymbol{\xi}_t\} \mid \mathcal{G}_{t-1}] \le \exp(\lambda^2\sigma^2/2)$ for every unit v and every $\lambda \in \mathbb{R}$.*

4. *The projection is exact onto $\mathfrak{T}$, and the pooled estimator and remaining capacity are $\mathcal{G}_{L_T}$-measurable.*

5. *The radius gate is $r_{\mathrm{lin}}(\mathcal{H}) \log(T/h_T) \le c_{\mathrm{sur}}$. Use the freeze-time good-event bound $\delta_{L_T} := \Pr(\mathcal{B}_{L_T}^c)$ and integrated virtual-exit bound $\overline{\pi}_{\mathrm{ex},L_T}$ from (77). Their meanings are those in Lemma 14. For the prescribed schedule, the explicit load and estimator tails bound δ_{L_T} and (80) supplies the exit term for all sufficiently large horizons.*

For $L_T + h_T < T$, let $\bar{R}$ be a uniform per-period fluid-revenue bound, C_{exit} the exit constant from Lemma 14, and C_1, C_2, C_3 its finite primitive-dependent localization constants. Then

$$\mathrm{Reg}_T^{\mathrm{fl}}(\pi^{\mathrm{lin}}) \le C_1\left(L_T + h_T + \log\frac{T}{h_T}\right) + C_2 T\mathcal{M}_L + C_3 h_T\sqrt{\mathcal{M}_L}\log\frac{T}{h_T} + T\bar{R}\delta_{L_T} + C_{\mathrm{exit}}(T - h_T)\overline{\pi}_{\mathrm{ex},L_T}. \tag{85}$$

For $L_T + h_T \ge T$, the complete fallback has regret at most $T\bar{R}$.

Proof. Replace each fold coefficient by its reference g_a, producing error $Z_{0,a}$ with $\mathbb{E}Z_{0,a}^2 \le R_{0,a}$. For the purpose of analysis, clip $G_{f,a}$ further to $[g_a - \gamma_{f,a}, g_a + \gamma_{f,a}] \cap [-G_{\max}, G_{\max}]$. This copy remains training-fold measurable and agrees with the actual coefficient on $\mathcal{E}_{\mathrm{cal}}$. Its mean differs from the reference by $-\sum_f w_f(G_{f,a}^\dagger - g_a)(\bar{S}_{f,a} - \bar{S}_{c,a})$. Weighted Cauchy–Schwarz and independence from each held-out fold and center sample give mean square at most $\Delta_{\mathrm{cf},a}$. The L^2 triangle inequality gives $R_{\mathrm{lin},a}$ for the copy, unconditionally.

Apply (28) and a $Jn(F+2)$-term union to reference residual, held-out score and center means. This gives $b_{H,a}$ with failure δ_H. Conditional on $\mathcal{H}$, the online mgf gives b_O with a Jn-term union and failure δ_O. On calibration and both tail events, the pooled error is bounded by $(1-\Omega_a)b_{H,a} + \Omega_a b_O$. Projection and $\|K\|$ give (84) and δ_{est}.

Retain learned $W = \mathrm{diag}(\Omega_a)$, copied historical error $e_H^\dagger$ and online error u. Conditional centering eliminates only the historical–online cross term: $\mathbb{E}\|K\{(I-W)e_H^\dagger + Wu\}\|^2 = \mathbb{E}\|K(I-W)e_H^\dagger\|^2 + \mathbb{E}\,\mathrm{tr}\{KW\mathcal{V}(\mathcal{H})W^\top K^\top\}$. Conditional orthogonality across disjoint fixed-price time blocks gives $\mathcal{V}(\mathcal{H}) \preceq (\sigma^2/q)I_{Jn}$. The two terms are thus at most $k_{\mathrm{op}}^2(1-\omega_-)^2\sum_a R_{\mathrm{lin},a}$ and $(\sigma^2/q)\omega_+^2 k_F^2$, retaining weight–history dependence. Projection and the diameter charge calibration failure, giving $\mathcal{M}_L$. Apply Lemma 14 with this unconditional moment and history-measurable radius. Absorbing $N\Delta_L \le CL_T$ and the harmonic factor of two yields (85), including its failure and exit terms. □

Proof of Theorem 5. The predictable projected estimator is $\mathcal{F}^{L_T}$-measurable. Lemma 13 gives $\mathbb{E}e_L^2 \le C\bar{e}_L^2$ and $\Pr(e_L > r_{L,T}) \le T^{-3} + \delta_{\mathrm{cal}}$. The calibration record lies in $\mathcal{F}^0$, so the conditional noise assumptions and the stated radius gate permit frozen-estimator transfer. This proves the displayed bound, including its

$T\delta_{\mathrm{cal}}$ term. Under the specified rates, $T\bar{e}_L^2 = O(T/L_T) + O(T^{-2})$ and $h_T\bar{e}_L \log(T/h_T) = O(T^{1/4}\log T)$. Equation (80) and estimator failure give $\Delta_{\mathrm{surr}}(T) \le CT(2T^{-3} + \delta_{\mathrm{cal}}) + CT\sqrt{CnTe^{-ch_T}} = O(T^{-2}) + o(1)$. The square-root schedule therefore gives total $O(\sqrt{T})$, with small horizons covered by bounded fluid value and nonnegative revenue. Finally subtracting σ^2/L_T from the defined $\bar{e}_L^2$ gives exactly the strict statistical-envelope condition, including calibration failure. This comparison orders only envelopes. □

J.2 Actual regret comparison

Definition 24 (Objects for the actual-regret comparison)**.** Let

$$z(p) := (1, p^\top)^\top, \qquad z_j := z(p^{(j)}), \qquad G_{\mathrm{design}} := \frac{1}{J_0}\sum_{j=1}^{J_0} z_j z_j^\top \succeq \kappa_0 I_{n+1},$$

and let $q_T := \lceil c_L\sqrt{T}/J_0 \rceil$, $L_T := J_0 q_T$, $N_T := T - L_T$ and $h_T := \lceil c_h(1+\log T)\rceil$. Let $x^* = \bar{x}^*$ and $J = \bar{J}$ be the limiting optimizer and stable binding set from Assumption 3. For $E = (u, \dot{B}) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}$, use column-wise vectorization and define $H_* := B^{-1} + B^{-\top}$, $Q_0 := -\frac{1}{2}H_* \succeq 0$,

$$\dot{H}(E) := -B^{-1}\dot{B}B^{-1} - B^{-\top}\dot{B}^\top B^{-\top}, \qquad \dot{r}(E) := -B^{-1}\dot{B}B^{-1}\alpha + B^{-1}u.$$

Let $\mathrm{Sens}(E)$ be the first component of

$$\begin{pmatrix} H_* & -A_J^\top \\ A_J & 0 \end{pmatrix}\begin{pmatrix} \mathrm{Sens}(E) \\ \dot{\lambda}(E) \end{pmatrix} = \begin{pmatrix} \dot{r}(E) - \dot{H}(E)x^* \\ 0 \end{pmatrix}.$$

Write

$$\mathrm{Sens}(E) = \mathsf{S}\,\mathrm{vec}(E), \qquad p^* := B^{-1}(x^* - \alpha), \qquad G(E) := -u - \dot{B}p^* = \mathsf{G}\,\mathrm{vec}(E),$$

and define $Q_{\mathrm{dyn}} := \mathsf{S}^\top Q_0 \mathsf{S} + \mathsf{G}^\top Q_0 \mathsf{G}$. Let

$$\Sigma_d := \mathbb{E}[\xi_t\xi_t^\top \mid \mathcal{G}_{t-1}], \quad \Sigma_S := \mathbb{E}[Z_t Z_t^\top \mid \mathcal{G}_{t-1}], \quad \Sigma_{dS} := \mathbb{E}[\xi_t Z_t^\top \mid \mathcal{G}_{t-1}],$$

where these are fixed over prices and time, $\Sigma_S \succeq \lambda_S I$, and $\Sigma_{\mathrm{eff}} := \Sigma_d - \Sigma_{dS}\Sigma_S^{-1}\Sigma_{dS}^\top$. Define

$$V_{\mathrm{raw}}^0 := (G_{\mathrm{design}}^{-1} \otimes I_n)\left\{\frac{1}{J_0}\sum_{j=1}^{J_0} z_j z_j^\top \otimes \Sigma_d\right\}(G_{\mathrm{design}}^{-1} \otimes I_n),$$

$$V_{\mathrm{cv}}^0 := (G_{\mathrm{design}}^{-1} \otimes I_n)\left\{\frac{1}{J_0}\sum_{j=1}^{J_0} z_j z_j^\top \otimes \Sigma_{\mathrm{eff}}\right\}(G_{\mathrm{design}}^{-1} \otimes I_n),$$

and $\gamma_Q := \mathrm{tr}[Q_{\mathrm{dyn}}(V_{\mathrm{raw}}^0 - V_{\mathrm{cv}}^0)]$.

The filtration $\mathcal{G}_t$, the independent calibration samples (pairs of requests and scores, and observations of scores, at each design price), the nuisance estimates $\widehat{\Gamma}_j$ and $\widehat{m}_j$, and the known closed convex set $\mathcal{G}_\Gamma$ onto which the coefficient estimate is projected are defined in Section EC.1; $\xi_s^\star := \xi_s - \Gamma^\star Z_s$.

Theorem 7 (Actual regret comparison)**.** *Suppose Assumptions 1, 2, 3, 4, 5, 6(a) and 7 hold, together with the mechanical service rule of Definition 1. Let* $d := n+1$, $G_{\mathrm{design}} \succeq \kappa_0 I_d$, $C_T = T\bar{c}$, *and* $\mathrm{dist}(\Theta, \partial\mathfrak{T}) \ge r_{\mathfrak{T}} > 0$. *Assume the conditional centering, the joint exploration mgf, the post-exploration request mgf with proxy* $\bar{\sigma} = \max\{\sigma, K_{\mathrm{camp}}\}$, *and the fixed covariance blocks* $\Sigma_d, \Sigma_S, \Sigma_{dS}$ *stated in Section EC.1, with known* λ_S *such that* $\Sigma_S \succeq \lambda_S I$. *Use the independent calibration samples of Section EC.1 and a known closed convex set* $\mathcal{G}_\Gamma$

that contains every $\Gamma^\star(p^{(j)})$*, satisfies* $\sup_{\Gamma\in\mathcal{G}_\Gamma}\|\Gamma\|_{\mathrm{op}} \le M_\Gamma$ *and is used with Frobenius projection. No other calibration event is assumed. For fixed* $K_\Gamma, K_m > 0$*, take* $M_{\Gamma,T} \ge K_\Gamma(1+\log T)^2$*,* $C_{m,T} \ge K_m L_T(1+\log T)^2$*, and* $T \ge T_{\mathrm{cal}}(K_\Gamma, K_m)$*, so that the calibration event in Lemma 15 has* $\delta_{\mathrm{cal}}(T) := 4T^{-4}$ *and* $\Pr(C_T) \ge 1-\delta_{\mathrm{cal}}(T)$*. Both policies use the same design, service rule, certainty-equivalent policy and last periods* $h_T = \lceil c_h(1+\log T)\rceil$ *with* $c_h > 2/c_{\mathrm{ex}}$*, where* c_{ex} *is the constant in Section EC.1. Any predictable active test that passes on the good event of Section EC.1 when* $\|q_t - x^*\|_2 \le r_0/2$ *gives the same result, for example the test* $q_t \in \widehat{\mathcal{D}}$ *of Algorithm 2. If* $\gamma_Q \ge \underline{\gamma}_Q > 0$*, then*

$$\mathrm{Reg}_T^{\mathrm{fl}}\left(\pi_T^{\mathrm{raw}}\right) - \mathrm{Reg}_T^{\mathrm{fl}}\left(\pi_T^{\mathrm{cv}}\right) = \frac{N_T}{L_T}\gamma_Q + o(\sqrt{T}),$$

so the difference is positive for all sufficiently large T*.*

Proof outline. During exploration both policies use the same prices, so they have the same exploration revenue and the same remaining capacity. By the frozen-fit lemma and a second-order expansion of the benchmark, the regret of each policy is $(N_T/L_T)\,\mathrm{tr}(Q_{\mathrm{dyn}}V^0)$, with $V^0 = V^0_{\mathrm{raw}}$ or V^0_{cv}, plus a common term and $o(\sqrt{T})$. The contribution of exits, the last periods and the failure events to the regret of each policy is $o(\sqrt{T})$. The theorem follows from the difference of the two expressions. The complete proof is in Section EC.1. □

The condition $\gamma_Q > 0$ holds when $\Sigma_d \succ 0$, $\rho_\star > 0$ and $Q_{\mathrm{dyn}} \ne 0$. Indeed, $V^0_{\mathrm{raw}} - V^0_{\mathrm{cv}} = (G^{-1}_{\mathrm{design}} \otimes I_n)\{G_{\mathrm{design}} \otimes (\Sigma_d - \Sigma_{\mathrm{eff}})\}(G^{-1}_{\mathrm{design}} \otimes I_n)$, and $\Sigma_d - \Sigma_{\mathrm{eff}} = \Sigma_{dS}\Sigma_S^{-1}\Sigma_{dS}^\top \succeq \rho_\star^2\Sigma_d \succ 0$ by Assumption 7(b). Hence $V^0_{\mathrm{raw}} - V^0_{\mathrm{cv}} \succ 0$, and $Q_{\mathrm{dyn}} \succeq 0$ with $Q_{\mathrm{dyn}} \ne 0$ gives $\gamma_Q > 0$.

E-Companion: Proofs and Numerical Details

This e-companion contains the complete proof of Theorem 7 (Section EC.1) and the experimental designs, seeds and complete comparisons (Sections EC.2–EC.6).

EC.1 Complete Proof of Theorem 7

This section proves Theorem 7. The proof uses the same fixed design, service rule, certainty-equivalent controller, and terminal fallback for the raw and surrogate-assisted policies. The policies differ only through the exploration estimator.

The proof has six steps. Lemma 15 bounds the errors of the coefficient and center estimates. Lemma 16 gives the second-order sensitivity of the fitted fluid solution on the active set. Lemma 17 bounds the load and estimator failure probabilities and shows that, on the good events, the fitted problem has active set J and every request is fully served through time $T - h_T$ unless the exit event occurs. Lemma 18 computes the covariance of each estimator. Lemma 19 couples each controller with a virtual process and charges exits. Lemma 20 expands the quadratic loss of each policy. The proof of the theorem removes the first-order loss by telescoping the residual controller against the active constraints, applies Lemma 20, and subtracts the two expansions.

Throughout,

$$C_T = T\bar{c}, \qquad \bar{c} \in \mathbb{R}^m_{++},$$

so that the true fluid solution, its active set, and all sensitivity matrices are independent of T. Let

$$d := n + 1, \qquad r_S := \dim(S_t), \qquad z(p) := (1, p^\top)^\top, \qquad z_j := z(p^{(j)}).$$

The deterministic exploration design cycles through the J_0 design prices, and

$$G_{\text{design}} := \frac{1}{J_0} \sum_{j=1}^{J_0} z_j z_j^\top \succeq \kappa_0 I_d,$$

$$q_T := \left\lceil \frac{c_L \sqrt{T}}{J_0} \right\rceil, \qquad L_T := J_0 q_T, \qquad N_T := T - L_T, \qquad h_T := \lceil c_h (1 + \log T) \rceil,$$

where $c_h > 2/c_{\text{ex}}$, with c_{ex} defined below. We consider sufficiently large T so that $L_T + h_T < T$. For $1 \le s \le L_T$, let $j(s)$ be the unique design index satisfying

$$p_s = p^{(j(s))}.$$

The fixed-design Gram matrix is

$$V_L := \sum_{s=1}^{L_T} z_s z_s^\top = L_T G_{\text{design}}.$$

All vectorizations below are column-wise. Thus, for $E \in \mathbb{R}^{n \times d}$,

$$\text{vec}(E) \in \mathbb{R}^{nd}, \qquad \text{vec}(\xi z^\top) = (z \otimes I_n)\xi.$$

Let $\mathcal{F}^0$ contain all pre-campaign randomization and all pre-campaign calibration observations. Define the enlarged filtration

$$\mathcal{G}_t := \mathcal{F}^0 \vee \sigma(D_s : 1 \le s \le t;\ S_s : 1 \le s \le \min\{t, L_T\}).$$

Thus scores are included in the filtration only during exploration. Write

$$D_t = f(p_t) + \xi_t.$$

For $t \le L_T$, define

$$Z_t := S_t - \mu_S(p_t).$$

The exploration innovations satisfy

$$\mathbb{E}[\xi_t \mid \mathcal{G}_{t-1}] = 0, \qquad \mathbb{E}[Z_t \mid \mathcal{G}_{t-1}] = 0, \qquad t \le L_T, \tag{EC.1}$$

and, for every $u \in \mathbb{R}^n$, $v \in \mathbb{R}^{rs}$, and $\theta \in \mathbb{R}$,

$$\mathbb{E}\big[\exp\{\theta(u^\top \xi_t + v^\top Z_t)\}\big|\mathcal{G}_{t-1}\big] \le \exp\left\{\frac{\theta^2 K_{\text{camp}}^2}{2}\big(\|u\|_2^2 + \|v\|_2^2\big)\right\}, \qquad t \le L_T. \tag{EC.2}$$

After exploration, only request innovations are used. Put

$$\bar{\sigma} := \max\{\sigma, K_{\text{camp}}\}.$$

For $t > L_T$,

$$\mathbb{E}[\xi_t \mid \mathcal{G}_{t-1}] = 0, \qquad \mathbb{E}\Big[e^{\theta u^\top \xi_t}\Big|\mathcal{G}_{t-1}\Big] \le \exp\left\{\frac{\theta^2 \bar{\sigma}^2 \|u\|_2^2}{2}\right\} \quad (u \in \mathbb{R}^n,\ \theta \in \mathbb{R}). \tag{EC.3}$$

The use of $\bar{\sigma}$ in (EC.3) allows one conditional mgf bound to be used uniformly in the exit calculation.

The conditional covariance blocks during exploration are homoscedastic:

$$\Sigma_d := \mathbb{E}[\xi_t \xi_t^\top \mid \mathcal{G}_{t-1}], \qquad \Sigma_S := \mathbb{E}[Z_t Z_t^\top \mid \mathcal{G}_{t-1}], \qquad \Sigma_{dS} := \mathbb{E}[\xi_t Z_t^\top \mid \mathcal{G}_{t-1}], \tag{EC.4}$$

and the matrices in (EC.4) do not depend on t, on the design price, or on the past history. The known constant $\lambda_S > 0$ satisfies

$$\Sigma_S \succeq \lambda_S I_{rs}.$$

Define

$$\Sigma_{\text{eff}} := \Sigma_d - \Sigma_{dS}\Sigma_S^{-1}\Sigma_{dS}^\top.$$

By Assumption 4(b),

$$\max_{1 \le j \le n} D_{t,j} \le D_{\max}, \qquad \|D_t\|_2 \le \sqrt{n}\, D_{\max} \qquad \text{almost surely for every } t.$$

Let $r_{\text{img}} > 0$ be the common image-ball radius supplied by Lemma 9, after shrinking it so that $r_{\text{img}} \le \delta_D$, where δ_D is the true demand-image margin. Set

$$r_0 := \frac{r_{\text{img}}}{2}, \qquad c_{\text{ex}} := \frac{r_0^2}{1024 n \bar{\sigma}^2}.$$

The case $\bar{\sigma} = 0$ is excluded by the strict covariance-gap condition in Theorem 7.

Calibration estimators

For each design price $p^{(j)}$, let

$$\left(D^{0,j,\ell}, S^{0,j,\ell}\right)_{1\le \ell \le M_{\Gamma,T}}$$

be iid calibration pairs at price $p^{(j)}$. Define

$$\xi^{0,j,\ell} := D^{0,j,\ell} - f(p^{(j)}), \qquad Z^{0,j,\ell} := S^{0,j,\ell} - \mu_S(p^{(j)}).$$

The calibration pairs are independent of the campaign and independent across j. The centered pairs $(\xi^{0,j,\ell}, Z^{0,j,\ell})$ have mean zero and covariance blocks $(\Sigma_d, \Sigma_S, \Sigma_{dS})$. By Assumption 7(a),

$$\Gamma^\star(p^{(j)}) = \Sigma_{dS}\Sigma_S^{-1}.$$

They satisfy

$$\mathbb{E}\exp\left\{\theta(u^\top \xi^{0,j,\ell} + v^\top Z^{0,j,\ell})\right\} \le \exp\left\{\frac{\theta^2 K_{\text{camp}}^2}{2}\left(\|u\|_2^2 + \|v\|_2^2\right)\right\}.$$

For every design price $p^{(j)}$, let

$$\left(S^{c,j,\ell}\right)_{1\le \ell \le C_{m,T}}$$

be iid center-bank observations with mean $\mu_S(p^{(j)})$, independent of the calibration pairs and independent of the campaign. Each coordinate of

$$S^{c,j,\ell} - \mu_S(p^{(j)})$$

is sub-Gaussian with fixed proxy K_c.

Define the sample means

$$\bar{D}_j^0 := \frac{1}{M_{\Gamma,T}}\sum_{\ell=1}^{M_{\Gamma,T}} D^{0,j,\ell}, \qquad \bar{S}_j^0 := \frac{1}{M_{\Gamma,T}}\sum_{\ell=1}^{M_{\Gamma,T}} S^{0,j,\ell},$$

the centered score covariance estimator

$$\widehat{\Sigma}_{S,j} := \frac{1}{M_{\Gamma,T}}\sum_{\ell=1}^{M_{\Gamma,T}} (S^{0,j,\ell} - \bar{S}_j^0)(S^{0,j,\ell} - \bar{S}_j^0)^\top,$$

and the centered request–score cross-covariance estimator

$$\widehat{\Sigma}_{dS,j} := \frac{1}{M_{\Gamma,T}}\sum_{\ell=1}^{M_{\Gamma,T}} (D^{0,j,\ell} - \bar{D}_j^0)(S^{0,j,\ell} - \bar{S}_j^0)^\top.$$

Put

$$\widetilde{\Gamma}_j := \begin{cases} \widehat{\Sigma}_{dS,j}\widehat{\Sigma}_{S,j}^{-1}, & \lambda_{\min}(\widehat{\Sigma}_{S,j}) \ge \lambda_S/2, \\ 0, & \text{otherwise}, \end{cases}$$

and let $\mathcal{G}_\Gamma$ be a known closed convex set satisfying

$$\Gamma^\star(p^{(j)}) \in \mathcal{G}_\Gamma \qquad \text{for every } j,$$

and

$$\sup_{\Gamma\in\mathcal{G}_\Gamma} \|\Gamma\|_{\text{op}} \le M_\Gamma.$$

The coefficient estimate is the Frobenius projection

$$\widehat{\Gamma}_j := \Pi_{\mathcal{G}_\Gamma}^{\|\cdot\|_F}(\widetilde{\Gamma}_j).$$

The center estimate is

$$\widehat{m}_j := \frac{1}{C_{m,T}} \sum_{\ell=1}^{C_{m,T}} S^{c,j,\ell}.$$

All $\widehat{\Gamma}_j$ and $\widehat{m}_j$ are $\mathcal{F}^0$-measurable.

Let

$$C_{\mathrm{est}} := 5 \max\left\{\sigma,\ K_{\mathrm{camp}}(1 + M_\Gamma^2)^{1/2}\right\} \sqrt{\frac{n(d + \log n + 4)}{\kappa_0}} + 1.$$

Lemma 15 (Calibration rates)**.** *Let $K_\Gamma, K_m > 0$ be any fixed positive constants. There is $T_{\mathrm{cal}} = T_{\mathrm{cal}}(K_\Gamma, K_m)$, depending also on M_Γ, κ_0, C_{est}, J_0, r_S and the design, such that, for $T \geq T_{\mathrm{cal}}$,*

$$M_{\Gamma,T} \geq K_\Gamma(1 + \log T)^2, \qquad C_{m,T} \geq K_m L_T (1 + \log T)^2$$

imply the existence of an $\mathcal{F}^0$-measurable event C_T and deterministic envelopes

$$\eta_{\Gamma,T} = C_\Gamma \sqrt{\frac{1 + \log T}{M_{\Gamma,T}}}, \qquad \eta_{m,T} = C_{\mathrm{ctr}} \sqrt{\frac{1 + \log T}{C_{m,T}}}.$$

The constants C_Γ and C_{ctr} depend only on $n, r_S, K_{\mathrm{camp}}, K_c, \lambda_S$. For use in Lemma 17, enlarge $T_{\mathrm{cal}}(K_\Gamma, K_m)$, if needed, so that

$$(1 + \log T) \geq \frac{2 M_\Gamma C_{\mathrm{ctr}} \max_j \|z_j\|_2}{\kappa_0 C_{\mathrm{est}} \sqrt{K_m}}.$$

Then, with $\delta_{\mathrm{cal}}(T) := 4T^{-4}$,

$$\Pr(C_T^c) \leq \delta_{\mathrm{cal}}(T),$$

and, on C_T,

$$\max_{1 \leq j \leq J_0} \|\widehat{\Gamma}_j - \Gamma^\star(p^{(j)})\|_F \leq \eta_{\Gamma,T},$$

$$\max_{1 \leq j \leq J_0} \|\widehat{m}_j - \mu_S(p^{(j)})\|_2 \leq \eta_{m,T}.$$

Under the stated sample sizes, the envelopes satisfy

$$\eta_{\Gamma,T}^2 + L_T \eta_{m,T}^2 \leq \left(\frac{C_\Gamma^2}{K_\Gamma} + \frac{C_{\mathrm{ctr}}^2}{K_m}\right) \frac{1}{1 + \log T}.$$

Proof of Lemma 15. For every fixed pair of coordinates, the product of two centered sub-Gaussian variables is sub-exponential. Bernstein's inequality therefore yields, for a constant $C_0 = C_0(K_{\mathrm{camp}})$,

$$\Pr\left(\left|\frac{1}{M_{\Gamma,T}} \sum_{\ell=1}^{M_{\Gamma,T}} Z_a^{0,j,\ell} Z_b^{0,j,\ell} - \mathbb{E}[Z_a Z_b]\right| > C_0 \sqrt{\frac{1 + \log T}{M_{\Gamma,T}}}\right) \leq 2T^{-6},$$

after increasing T_{cal}. The same inequality applies to $\xi_a^{0,j,\ell} Z_b^{0,j,\ell}$. The difference between the sample-mean-centered products and their uncentered versions is

$$\bar{Z}_j^0 (\bar{Z}_j^0)^\top \quad \text{or} \quad (\bar{D}_j^0 - f(p^{(j)}))(\bar{Z}_j^0)^\top.$$

The sub-Gaussian sample-mean bound implies

$$\|\bar{Z}_j^0\|_2 + \|\bar{D}_j^0 - f(p^{(j)})\|_2 \le C_1\sqrt{\frac{1+\log T}{M_{\Gamma,T}}}$$

outside a probability $O(T^{-6})$ event. A finite union over j and coordinates gives

$$\max_j \left\{\|\widehat{\Sigma}_{S,j} - \Sigma_S\|_{\mathrm{op}} + \|\widehat{\Sigma}_{dS,j} - \Sigma_{dS}\|_{\mathrm{op}}\right\} \le C_2\sqrt{\frac{1+\log T}{M_{\Gamma,T}}}$$

outside a probability $2T^{-4}$ event.

For $T \ge T_{\mathrm{cal}}$, the right-hand side is at most $\lambda_S/2$. Hence

$$\lambda_{\min}(\widehat{\Sigma}_{S,j}) \ge \lambda_S/2,$$

and

$$\widehat{\Sigma}_{S,j}^{-1} - \Sigma_S^{-1} = -\widehat{\Sigma}_{S,j}^{-1}(\widehat{\Sigma}_{S,j} - \Sigma_S)\Sigma_S^{-1}.$$

Consequently,

$$\|\widetilde{\Gamma}_j - \Gamma^\star(p^{(j)})\|_F \le C_3\sqrt{\frac{1+\log T}{M_{\Gamma,T}}}.$$

Since $\Gamma^\star(p^{(j)}) \in \mathcal{G}_\Gamma$, Frobenius projection is nonexpansive relative to $\Gamma^\star(p^{(j)})$, so

$$\|\widehat{\Gamma}_j - \Gamma^\star(p^{(j)})\|_F \le \|\widetilde{\Gamma}_j - \Gamma^\star(p^{(j)})\|_F.$$

For the center-bank sample, every coordinate is centered sub-Gaussian with proxy K_c. Thus

$$\Pr\left(\left|\frac{1}{C_{m,T}}\sum_{\ell=1}^{C_{m,T}}(S_a^{c,j,\ell} - \mu_{S,a}(p^{(j)}))\right| > C_4K_c\sqrt{\frac{1+\log T}{C_{m,T}}}\right) \le 2T^{-6}.$$

A finite union over prices and coordinates gives the center estimate bound outside a probability $2T^{-4}$ event. Intersecting the coefficient and center events proves the lemma. □

Lemma 16 (Active-set C^2 sensitivity)**.** *Let $\Theta = (\alpha, B)$. For $\vartheta = (a, M)$ near Θ, define*

$$H_\vartheta := M^{-1} + M^{-\top}, \qquad r_\vartheta := M^{-1}a, \qquad \ell_\vartheta(x) := x^\top M^{-1}(x - a).$$

Let $X(\vartheta, c)$ be the unique fluid solution over the feasible polytope, and write

$$x^* := X(\Theta, \bar{c}), \qquad J := \{i : A_i x^* = \bar{c}_i\}.$$

Define $\mathcal{D}_\Theta := \alpha + B\mathcal{P}$ and, for each parameter error E, $\mathcal{D}_{\Theta+E} := \alpha_E + B_E\mathcal{P}$, where $\Theta + E = (\alpha_E, B_E)$. Let λ_J^* denote the true active multiplier from the fluid KKT conditions, restricted to J, and set $\lambda_J^* = 0$ when $J = \varnothing$. Under Assumptions 2(b), 3 and 5, there exist $r_1, r_2, C_{\mathrm{sen}} > 0$ such that, whenever

$$\|\vartheta - \Theta\|_F \le r_1, \qquad \|c - \bar{c}\|_\infty \le r_2,$$

the active set remains J, inactive constraints retain positive slack, the optimizer and its active multiplier are C^2 in (ϑ, c), and, with

$$s_\lambda := \frac{\lambda_{\min}}{2},$$

$$\lambda_J(\vartheta, c) \geq s_\lambda \mathbf{1},$$

whenever $J \neq \varnothing$. If $J = \varnothing$, the multiplier block is absent.

For $E = (u, \dot{B})$, define

$$\dot{H}(E) = -B^{-1}\dot{B}B^{-1} - B^{-\top}\dot{B}^\top B^{-\top},$$
$$\dot{r}(E) = -B^{-1}\dot{B}B^{-1}\alpha + B^{-1}u.$$

The matrix

$$H_* := B^{-1} + B^{-\top} = B^{-1}(B + B^\top)B^{-\top} \prec 0.$$

The parameter sensitivity and capacity sensitivity are defined by

$$\begin{pmatrix} H_* & -A_J^\top \\ A_J & 0 \end{pmatrix} \begin{pmatrix} \mathrm{Sens}(E) \\ \dot{\lambda}(E) \end{pmatrix} = \begin{pmatrix} \dot{r}(E) - \dot{H}(E)x^* \\ 0 \end{pmatrix}, \tag{EC.5}$$

and

$$\begin{pmatrix} H_* & -A_J^\top \\ A_J & 0 \end{pmatrix} \begin{pmatrix} S_c v \\ \Lambda_c v \end{pmatrix} = \begin{pmatrix} 0 \\ v \end{pmatrix}, \qquad v \in \mathbb{R}^{|J|}. \tag{EC.6}$$

Here P_J restricts a vector to its coordinates in J. If $J = \varnothing$, the bordered matrix is H_* and $S_c = 0$. Uniformly on the neighborhood,

$$\|X(\vartheta, c) - x^* - \mathrm{Sens}(\vartheta - \Theta) - S_c P_J(c - \bar{c})\|_2 \leq C_{\mathrm{sen}}\big(\|\vartheta - \Theta\|_F + \|c - \bar{c}\|_\infty\big)^2. \tag{EC.7}$$

Proof of Lemma 16. The active KKT system is

$$F(\vartheta, c, x, \lambda) = \begin{pmatrix} H_\vartheta x - r_\vartheta - A_J^\top \lambda \\ A_J x - c_J \end{pmatrix} = 0.$$

The map F is C^2 because matrix inversion is C^∞ on the open set of nonsingular matrices and the primitive objective and demand maps are C^2.

At the truth, the Jacobian in (x, λ) is

$$D_{(x,\lambda)}F = \begin{pmatrix} H_* & -A_J^\top \\ A_J & 0 \end{pmatrix}.$$

Suppose that (v, w) belongs to its nullspace. Then $A_J v = 0$ and

$$H_* v - A_J^\top w = 0.$$

Multiplying by $v^\top$ gives

$$v^\top H_* v = 0.$$

Since $H_* \prec 0$, $v = 0$. Full row rank of A_J then implies $w = 0$. Thus the Jacobian is nonsingular. The implicit function theorem gives a unique C^2 solution map

$$(\vartheta, c) \mapsto (X(\vartheta, c), \lambda_J(\vartheta, c))$$

in a neighborhood of $(\Theta, \bar{c})$.

Assumption 3 supplies the inactive-row slack margin $s_{\min}$, the active-multiplier margin $\lambda_{\min}$, the active-row rank margin, and the price-box margin δ_P. Assumption 5 supplies the interior demand-image margin δ_D. By continuity, after shrinking the neighborhood, inactive slacks remain positive, inverse prices remain in the price set, and every active multiplier remains at least $\lambda_{\min}/2 = s_\lambda$. Thus the restricted KKT solution is the full KKT solution and is the unique optimizer.

If $J = \varnothing$, the optimizer is locally

$$X(\vartheta, c) = H_{\vartheta}^{-1} r_{\vartheta},$$

which is C^2 and independent of c, so $S_c = 0$.

Differentiating the KKT equation in the direction E gives (EC.5). Differentiating in the capacity direction v gives (EC.6). The multivariate Taylor formula with integral remainder gives

$$X(\vartheta, c) = x^* + \mathrm{Sens}(\vartheta - \Theta) + S_c P_J (c - \bar{c}) + R,$$

where

$$\|R\|_2 \le \frac{1}{2} \sup_{\mathcal{N}} \|D^2 X\| \left(\|\vartheta - \Theta\|_F + \|c - \bar{c}\|_2\right)^2.$$

Since $\|c - \bar{c}\|_2 \le \sqrt{m}\|c - \bar{c}\|_\infty$, the factor $\sqrt{m}$ is absorbed into C_{sen}. □

Because the sensitivity map is linear, there is a matrix $\mathsf{S} \in \mathbb{R}^{n \times nd}$ such that

$$\mathrm{Sens}(E) = \mathsf{S}\,\mathrm{vec}(E).$$

Let

$$p^* := B^{-1}(x^* - \alpha), \qquad G(E) := -u - \dot{B} p^* = \mathsf{G}\,\mathrm{vec}(E),$$

where $\mathsf{G} \in \mathbb{R}^{n \times nd}$. Define

$$Q_0 := -\frac{1}{2} H_* \succeq 0, \qquad Q_{\mathrm{dyn}} := \mathsf{S}^\top Q_0 \mathsf{S} + \mathsf{G}^\top Q_0 \mathsf{G}.$$

The matrices in Definition 24 are

$$V_{\mathrm{raw}}^0 = (G_{\mathrm{design}}^{-1} \otimes I_n) \left\{ \frac{1}{J_0} \sum_{j=1}^{J_0} z_j z_j^\top \otimes \Sigma_d \right\} (G_{\mathrm{design}}^{-1} \otimes I_n),$$

$$V_{\mathrm{cv}}^0 = (G_{\mathrm{design}}^{-1} \otimes I_n) \left\{ \frac{1}{J_0} \sum_{j=1}^{J_0} z_j z_j^\top \otimes \Sigma_{\mathrm{eff}} \right\} (G_{\mathrm{design}}^{-1} \otimes I_n),$$

and

$$\gamma_Q = \mathrm{tr}\left[Q_{\mathrm{dyn}} \left(V_{\mathrm{raw}}^0 - V_{\mathrm{cv}}^0 \right) \right].$$

The two compared policies use the same design, mechanical service rule, actual request observations for capacity and residual updates, frozen estimate controller, fixed price, and logarithmic last periods. Their exploration estimators are

$$\widetilde{\Theta}_{L_T}^{\mathrm{raw}} := \left(\sum_{s=1}^{L_T} D_s z_s^\top \right) V_L^{-1}, \qquad \widehat{\Theta}_{L_T}^{\mathrm{raw}} := \Pi_{\mathfrak{T}}^{\|\cdot\|_F} (\widetilde{\Theta}_{L_T}^{\mathrm{raw}}),$$

and

$$\widetilde{D}_s := D_s - \widehat{\Gamma}_{j(s)} \{ S_s - \widehat{m}_{j(s)} \},$$

$$\widetilde{\Theta}_{L_T}^{\mathrm{cv}} := \left(\sum_{s=1}^{L_T} \widetilde{D}_s z_s^\top \right) V_L^{-1}, \qquad \widehat{\Theta}_{L_T}^{\mathrm{cv}} := \Pi_{\mathfrak{T}}^{\|\cdot\|_F} (\widetilde{\Theta}_{L_T}^{\mathrm{cv}}).$$

The parameter class has fixed interior radius

$$\mathrm{dist}(\Theta, \partial \mathfrak{T}) \ge r_{\mathfrak{T}} > 0.$$

After period L_T, both policies freeze their projected estimate, define

$$b_T := \frac{c^{L_T+1}}{N_T},$$

solve the fitted fluid problem at capacity b_T, and run the residual controller until $T - h_T$ unless the active test fails. We write the policies as π_T^{raw} and π_T^{cv}.

$$r_T := C_{\mathrm{est}}\sqrt{\frac{1+\log T}{L_T}}, \qquad \bar{D}_A := \max_{1\le i\le m}\sum_{j=1}^{n} A_{ij}D_{\max},$$

$$C_{\mathrm{load}} := 2\left(\|\bar{c}\|_\infty + \bar{D}_A\right).$$

Define

$$\delta c_T := b_T - \bar{c}, \qquad \Delta_T := \|\delta c_T\|_\infty, \qquad \overline{\Delta}_T := C_{\mathrm{load}}\frac{L_T}{T},$$

where C_{load} is the deterministic load constant defined above. For $k \in \{\mathrm{raw}, \mathrm{cv}\}$, define

$$E_{k,T} := \widehat{\Theta}^k_{L_T} - \Theta, \qquad \varepsilon_{k,T} := \|E_{k,T}\|_F, \qquad x_k := X(\Theta + E_{k,T}, b_T),$$

$$\widehat{\Theta}^k_{L_T} = (\widehat{\alpha}_k, \widehat{B}_k), \qquad \widehat{p}_k(v) := \widehat{B}_k^{-1}(v - \widehat{\alpha}_k), \qquad p_k(v) := \widehat{p}_k(v) \quad \text{for every } v \in \mathbb{R}^n,$$

$$g_k(v) := \alpha + B\widehat{p}_k(v), \qquad M_k := B\widehat{B}_k^{-1} - I, \qquad \lambda_{k,J} := \lambda_J(\Theta + E_{k,T}, b_T).$$

Here X is the map of Definition 5: the fluid maximizer when the feasible set is nonempty and the fixed point otherwise. The multiplier $\lambda_{k,J}$ is defined on the neighborhood of Lemma 16 and set to 0 elsewhere. For the surrogate estimator, define

$$\delta_j^{\Gamma} := \widehat{\Gamma}_j - \Gamma^\star(p^{(j)}), \qquad \delta_j^{m,+} := \widehat{m}_j - \mu_S(p^{(j)}),$$

$$\xi_s^\star := \xi_s - \Gamma^\star(p^{(j(s))})Z_s, \qquad \widetilde{E}_{c,T} := \widetilde{\Theta}^{\mathrm{cv}}_{L_T} - \Theta.$$

For sufficiently large T, $\overline{\Delta}_T$ is smaller than the local capacity radius in Lemma 16.

The calibration event used below is the event from Lemma 15:

$$E_{\mathrm{cal}} := C_T, \qquad \Pr(C_T^c) \le \delta_{\mathrm{cal}}(T).$$

For $k \in \{\mathrm{raw}, \mathrm{cv}\}$, let

$$\mathcal{E}_{\mathrm{est},k,T} := \{\varepsilon_{k,T} \le r_T\},$$

and define $\mathcal{E}_{\mathrm{load},T}$ to be the event that all exploration requests are fully served and

$$\Delta_T \le \overline{\Delta}_T.$$

Let the constants L_*, c_*, r_* be those supplied by Lemma 9, and set

$$C_\Delta := L_*, \qquad C_g := 2L_*, \qquad c_\Delta := \min\{c_*, 1\}, \qquad r_{\mathrm{loc}} := r_*,$$

and

$$C_f := 1 + \sup_{p\in\mathcal{P}}\|p\|_2, \qquad C_H := \frac{2}{\kappa_B}, \qquad C_a := C_f, \qquad C_x := C_\Delta + C_g.$$

For $k \in \{\mathrm{raw}, \mathrm{cv}\}$, put

$$v_k := X(\Theta, b_T),$$

which does not depend on k. For every t and v in the post-exploration window, define

$$\Phi_{t,v}^{(k)} := \prod_{u=v+1}^{t-1} \left(I - \frac{M_k}{T-u} \right).$$

Let $a_A := \min_{i:A_i\mathbf{1}>0} A_i\mathbf{1}$ when there is a positive row and set $a_A := 1$ if every row of A is zero. Put $\tau_{b,T} := \Delta_T / a_A$ when there is a positive row and $\tau_{b,T} := 0$ otherwise. The actual controller state $q_{t,k}$ is initialized by $q_{L_T+1,k} = x_k$ and, while its predictable active test (24) passes, evolves according to (25).

The frozen geometry gate is the gate in Definition 5, instantiated at $s = L_T$, $h = h_T$, with the constants supplied by Lemma 9. Every clause is verified directly in the proof of Lemma 17. Write $\mathcal{E}_{\mathrm{geom},k,T}$ for the following full conjunction:

$$\begin{gathered}
\|v_k - x^*\|_2 \le C_\Delta \sqrt{\Delta_T}, \qquad \|x_k - v_k\|_2 \le C_g(\varepsilon_{k,T} + \sqrt{\Delta_T}), \\
\|g_k(x_k) - x_k\|_2 \le C_a(\varepsilon_{k,T} + \sqrt{\Delta_T}), \qquad \|x_k - x^*\|_2 \le C_x(\varepsilon_{k,T} + \sqrt{\Delta_T}), \\
\|x_k - x^*\|_2 \le r_0/4, \qquad c^{L_T+1} - N_T A x_k \ge 0, \qquad A_J x_k = (b_T)_J, \qquad A_{J^c} x_k < (b_T)_{J^c}, \\
\nabla \ell(x^*) = A_J^\top \lambda_J^*, \qquad \lambda_{k,J} > 0 \text{ if } J \ne \varnothing, \qquad \sigma_{\min}(A_J) \ge s_J > 0 \text{ if } J \ne \varnothing, \\
\mathbb{B}_2(x_k, r_0) \subseteq \mathcal{D}_\Theta, \qquad \mathbb{B}_2(x_k, r_0) \subseteq \mathcal{D}_{\Theta + E_{k,T}}, \\
\lambda_{k,J} \ge s_\lambda \mathbf{1}, \qquad \Delta_T \le c_\Delta, \qquad \varepsilon_{k,T} \le r_{\mathrm{loc}}, \\
4 \log \frac{N_T}{h_T} \max\{C_H \varepsilon_{k,T}, C_a(\varepsilon_{k,T} + \sqrt{\Delta_T})\} \le \min\{\log 2, r_0/8\}, \\
C_g(\varepsilon_{k,T} + \sqrt{\Delta_T}) + \sqrt{n}\, \tau_{b,T} \le r_0/2, \qquad \|\Phi_{t,v}^{(k)}\|_{\mathrm{op}} \le 2 \quad (L_T + 1 \le v < t \le T - h_T + 1).
\end{gathered}$$

Here $s_J := a_{\min}$, the active-row rank margin of Assumption 3. The constants L_*, c_*, r_* are those of Lemma 9. By (33), the two bounds used to instantiate the shift clauses are $\|v_k - x^*\|_2 \le L_* \Delta_T \le L_* \sqrt{\Delta_T}$ for $\Delta_T \le 1$ and $\|x_k - v_k\|_2 \le 2L_*(\varepsilon_{k,T} + \sqrt{\Delta_T})$. Thus $C_\Delta = L_*$ and $C_g = 2L_*$. The image inclusion follows from $\mathbb{B}_2(x_k, r_0) \subseteq \mathbb{B}_2(x^*, 5r_0/4) \subseteq \mathbb{B}_2(x^*, r_{\mathrm{img}}) \subseteq \mathcal{D}_\Theta$ and the common image-ball property gives the corresponding inclusion in $\mathcal{D}_{\Theta+E_{k,T}}$. In particular, $\mathbb{B}_2(x^*, r_0/2) \subseteq \mathbb{B}_2(x_k, r_0) \subseteq \mathcal{D}_{\Theta+E_{k,T}}$, so the fitted-image active test passes on this ball. The lower bound on $\lambda_{k,J}$ follows from Lemma 16.

For each k, define the good freeze event

$$\mathcal{B}_{k,T} := C_T \cap \mathcal{E}_{\mathrm{load},T} \cap \mathcal{E}_{\mathrm{est},k,T} \cap \mathcal{E}_{\mathrm{geom},k,T}.$$

Define the global unstopped process by

$$\widetilde{q}_{L_T+1,k}^{\,u} := \mathbf{1}_{\mathcal{B}_{k,T}} x_k + \mathbf{1}_{\mathcal{B}_{k,T}^c} x^*, \qquad \widetilde{q}_{t+1,k}^{\,u} := \widetilde{q}_{t,k}^{\,u} - \frac{g_k(\widetilde{q}_{t,k}^{\,u}) + \xi_t - \widetilde{q}_{t,k}^{\,u}}{T-t}, \quad L_T + 1 \le t \le T - h_T,$$

which is well defined because $p_k(v) = \widehat{B}_k^{-1}(v - \widehat{\alpha}_k)$ is defined on all of $\mathbb{R}^n$. Let

$$\tau_{\mathrm{ex},k} := \inf\{t \in \{L_T + 1, \dots, T - h_T + 1\} : \|\widetilde{q}_{t,k}^{\,u} - x^*\|_2 > r_0/2\}, \qquad \inf \varnothing := T - h_T + 2,$$

where $\tau_{\mathrm{ex},k}$ is an analysis stopping time. Define $\mathcal{E}_{\mathrm{ex},k} := \{\tau_{\mathrm{ex},k} \le T - h_T + 1\}$ and the clipped virtual process

$$\widetilde{q}_{t,k} := \widetilde{q}_{t,k}^{\,u} \quad (t < \tau_{\mathrm{ex},k}), \qquad \widetilde{q}_{t,k} := x^* \quad (t \ge \tau_{\mathrm{ex},k}).$$

The analysis below applies to any predictable active test that passes whenever $\{\|q_t - x^*\|_2 \le r_0/2\} \cap \mathcal{B}_{k,T}$ holds, including Algorithm 2's test $q_t \in \widehat{\mathcal{D}}$.

Lemma 17 (Post-exploration feasibility for the logarithmic window)**.** *For all sufficiently large T, the event $\mathcal{E}_{\mathrm{load},T}$ has probability one. Moreover,*

$$\Pr(\mathcal{E}^c_{\mathrm{est,raw},T}) \le T^{-4}, \qquad \Pr(\mathcal{E}^c_{\mathrm{est,cv},T}) \le T^{-4} + \delta_{\mathrm{cal}}(T).$$

On

$$\mathcal{E}_{\mathrm{load},T} \cap \mathcal{E}_{\mathrm{est},k,T} \cap C_T,$$

the fitted problem has active set J, satisfies the geometry gate, and every request is fully served through time $T - h_T$ unless the exit event occurs. In particular,

$$\mathbf{1}_{\mathcal{B}_{k,T}} \Pr(\mathcal{E}_{\mathrm{ex},k} \mid \mathcal{G}_{L_T}) \le 2nN_T e^{-c_{\mathrm{ex}} h_T} \quad a.s.,$$

and

$$\bar{\pi}_{\mathrm{ex},k} := \mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}} \sqrt{\Pr(\mathcal{E}_{\mathrm{ex},k} \mid \mathcal{G}_{L_T})}\Big] \le \sqrt{2nN_T}\, e^{-c_{\mathrm{ex}} h_T/2}.$$

Consequently,

$$N_T \bar{\pi}_{\mathrm{ex},k} = o(\sqrt{T}).$$

Proof of Lemma 17. The bounded-potential condition gives

$$A \sum_{u=1}^{s} D_u \le s\bar{D}_A \mathbf{1} \le L_T \bar{D}_A \mathbf{1}.$$

Since $L_T = O(\sqrt{T})$ and $C_T = T\bar{c}$, for large T,

$$L_T \bar{D}_A \le \frac{1}{2} T \min_i \bar{c}_i.$$

Thus every exploration request is fully served and

$$c^{L_T+1} = T\bar{c} - A \sum_{s=1}^{L_T} D_s.$$

Therefore

$$b_T - \bar{c} = \frac{L_T \bar{c} - A \sum_{s=1}^{L_T} D_s}{N_T}.$$

Since $N_T \ge T/2$,

$$\Delta_T \le \frac{2L_T}{T} \left(\|\bar{c}\|_\infty + \bar{D}_A\right) = \overline{\Delta}_T.$$

We now bound the raw estimator tail using only the exploration mgf (EC.2). Set

$$x := d \log 5 + \log(2n) + 4 \log T.$$

For each row i and unit vector $w \in \mathbb{R}^d$,

$$e_i^\top (\widetilde{\Theta}^{\mathrm{raw}}_{L_T} - \Theta) w = \sum_{s=1}^{L_T} (w^\top V_L^{-1} z_s) \xi_{s,i}.$$

The conditional mgf follows from (EC.2) with $v = 0$. A union over the n rows and a $1/2$-net of S^{d-1} with 5^d points gives

$$\Pr\left(\|\widetilde{\Theta}^{\mathrm{raw}}_{L_T} - \Theta\|_F > 2K_{\mathrm{camp}} \sqrt{\frac{2nx}{\kappa_0 L_T}} \right) \le T^{-4}.$$

For $T \geq e^2$, the definition of C_{est} implies

$$2K_{\mathrm{camp}}\sqrt{\frac{2nx}{\kappa_0 L_T}} \leq \frac{r_T}{2}.$$

The Frobenius projection is nonexpansive and $\Theta \in \mathfrak{T}$, so

$$\Pr(\mathcal{E}^c_{\mathrm{est,raw},T}) \leq T^{-4}.$$

For the surrogate estimator, write

$$\widetilde{E}_{c,T} = U_T + W_T,$$

where

$$U_T := \left(\sum_{s=1}^{L_T} [\xi_s^\star - \delta^\Gamma_{j(s)} Z_s] z_s^\top\right) V_L^{-1},$$

$$W_T := \left(\sum_{s=1}^{L_T} \widehat{\Gamma}_{j(s)} \delta^{m,+}_{j(s)} z_s^\top\right) V_L^{-1}.$$

For each row i and unit vector $w \in \mathbb{R}^d$, use $u = e_i$ and $v = -\widehat{\Gamma}^\top_{j(s)} e_i$ in (EC.2). The bound is valid for $\mathcal{F}^0$-measurable random (u, v) by the regular conditional distribution argument (equivalently, by continuity and a countable dense set). Since $\|\widehat{\Gamma}_j\|_{\mathrm{op}} \leq M_\Gamma$,

$$\mathbb{E}\left[e^{\theta u^\top(\xi_s^\star - \delta^\Gamma_{j(s)} Z_s)} \middle| \mathcal{F}^0 \vee \mathcal{G}_{s-1}\right] \leq \exp\left\{\frac{\theta^2 K^2_{\mathrm{camp}}}{2}(1 + M_\Gamma^2)\right\}.$$

The same finite-net argument gives

$$\Pr\left(\|U_T\|_F > \frac{r_T}{2}\right) \leq T^{-4}.$$

Since the design cycles exactly,

$$\begin{aligned}\|W_T\|_F &\leq \|V_L^{-1}\| \left(\sum_{s=1}^{L_T} \|\widehat{\Gamma}_{j(s)} \delta^{m,+}_{j(s)}\|_2^2\right)^{1/2} \left(\sum_{s=1}^{L_T} \|z_s\|_2^2\right)^{1/2} \\ &\leq \frac{M_\Gamma}{\kappa_0} \max_j \|z_j\|_2 \max_j \|\delta_j^{m,+}\|_2,\end{aligned}$$

which is at most $M_\Gamma \eta_{m,T} \max_j \|z_j\|_2/\kappa_0$ on C_T. The choice $C_{m,T} \geq K_m L_T (1 + \log T)^2$ and Lemma 15 make this bound at most $r_T/2$ on C_T for $T \geq T_{\mathrm{cal}}$. Hence

$$\Pr(\mathcal{E}^c_{\mathrm{est,cv},T}) \leq T^{-4} + \delta_{\mathrm{cal}}(T).$$

On either estimator event, the parameter error is at most r_T, and

$$\Delta_T \leq \overline{\Delta}_T.$$

Therefore Lemma 16, the local-margin assumption, and Lemma 9 imply that the active set is J, all inactive constraints retain slack, and

$$\lambda_{k,J} \geq \lambda_{\min}/2 = s_\lambda.$$

The constants in the geometry gate are those obtained by instantiating Definition 5 at $(s, h) = (L_T, h_T)$ with

$$C_\Delta = L_*, \qquad C_g = 2L_*, \qquad c_\Delta = \min\{c_*, 1\}, \qquad r_{\mathrm{loc}} = r_*.$$

For the drift clause,

$$\widehat{\alpha}_k + \widehat{B}_k \widehat{p}_k(x_k) = x_k.$$

Thus

$$\begin{aligned} g_k(x_k) - x_k &= f(\widehat{p}_k(x_k)) - (\widehat{\alpha}_k + \widehat{B}_k \widehat{p}_k(x_k)) \\ &= (\Theta - \widehat{\Theta}^k_{L_T}) z(\widehat{p}_k(x_k)), \end{aligned}$$

and compactness of the price set and the inverse-matrix bounds give

$$\|g_k(x_k) - x_k\|_2 \le C_f \varepsilon_{k,T}.$$

The matrix product bound is

$$\|M_k\|_{\mathrm{op}} = \|B\widehat{B}_k^{-1} - I\|_{\mathrm{op}} \le C_H \varepsilon_{k,T}, \qquad C_H = \frac{2}{\kappa_B}.$$

Since

$$r_T = O(T^{-1/4}\sqrt{1+\log T}), \qquad \overline{\Delta}_T = O(T^{-1/2}),$$

and

$$\log(N_T/h_T) = O(\log T),$$

all local-radius and image-shift clauses hold for large T.

It remains to prove full service. The demand-image margin gives $\mathbb{B}_2(x^*, \delta_D) \subseteq \mathcal{D}_\Theta$, and $\mathcal{D}_\Theta \subseteq \mathbb{R}^n_+$ because $f \ge 0$ on $\mathcal{P}$ (Assumption 2(b)). Hence

$$x^*_j \ge \delta_D \qquad \text{for every } j.$$

The actual and virtual states coincide before exit. On $\mathcal{B}_{k,T}$, $q_{L_T+1,k} = x_k = \widetilde{q}^u_{L_T+1,k}$. If $q_{t,k} = \widetilde{q}^u_{t,k}$ and $t < \tau_{\mathrm{ex},k} \wedge (T - h_T + 1)$, then $\|q_{t,k} - x^*\|_2 \le r_0/2$, so the active test passes, $\widehat{p}_k(q_{t,k}) \in \mathcal{P}$, and $D_t = f(\widehat{p}_k(q_{t,k})) + \xi_t = g_k(q_{t,k}) + \xi_t$. Then (25) gives $q_{t+1,k} = \widetilde{q}^u_{t+1,k}$. Hence the two states coincide up to and including $\tau_{\mathrm{ex},k} \wedge (T - h_T + 1)$. Since $r_0/2 < \delta_D/2$, on the geometry gate and before exit,

$$q_{t,k,j} \ge \delta_D/2.$$

The residual controller satisfies

$$D_t = (T - t + 1)q_{t,k} - (T - t)q_{t+1,k}.$$

Because $D_{t,j} \le D_{\max}$ and $T - t \ge h_T$,

$$q_{t+1,k,j} \ge \frac{\delta_D}{2} - \frac{D_{\max}}{h_T} > 0$$

for large T. Define

$$I_t := c^t - (T - t + 1)Aq_{t,k}.$$

At the freeze time, x_k is feasible for the fitted problem at b_T, so $Ax_k \le b_T = c^{L_T+1}/N_T$ and

$$I_{L_T+1} = c^{L_T+1} - N_T A x_k \ge 0.$$

If $I_t \ge 0$, then

$$\begin{aligned} c^t - AD_t &= c^t - A\big((T - t + 1)q_{t,k} - (T - t)q_{t+1,k}\big) \\ &= I_t + (T - t)Aq_{t+1,k} \ge 0. \end{aligned}$$

Thus the request is fully served and

$$I_{t+1} = c^{t+1} - (T-t)Aq_{t+1,k} = I_t \geq 0.$$

Induction proves feasibility through the first exit.

For the exit bound, we now use the recursion product $\Phi_{t,v}^{(k)}$. For $L_T + 1 \leq v < t \leq T - h_T + 1$,

$$\sum_{u=v+1}^{t-1} \frac{1}{T-u} \leq \sum_{r=h_T}^{N_T-1} \frac{1}{r} \leq 2 \log \frac{N_T}{h_T}$$

once $N_T/h_T \geq e$. The logarithmic clause of the gate therefore gives

$$\|\Phi_{t,v}^{(k)}\|_{\mathrm{op}} \leq \exp\Big\{2C_H \varepsilon_{k,T} \log \frac{N_T}{h_T}\Big\} \leq e^{(\log 2)/2} \leq 2,$$

and the deterministic drift is at most $2C_a(\varepsilon_{k,T} + \sqrt{\Delta_T}) \cdot 2\log(N_T/h_T) \leq r_0/8$. Since $\|x_k - x^*\|_2 \leq r_0/4$, an exit requires the noise part to exceed $r_0/2 - r_0/4 - r_0/8 = r_0/8$ in norm. Hence an exit implies that for some coordinate i and some $t \leq T - h_T + 1$,

$$\left| \sum_{v=L_T+1}^{t-1} e_i^\top \Phi_{t,v}^{(k)} \frac{\xi_v}{T-v} \right| \geq \frac{r_0}{8\sqrt{n}}.$$

The squared coefficient sum is at most

$$\sum_{v=L_T+1}^{t-1} \frac{\|e_i^\top \Phi_{t,v}^{(k)}\|_2^2}{(T-v)^2} \leq 4 \sum_{r=h_T}^{\infty} \frac{1}{r^2} \leq \frac{8}{h_T}.$$

The coefficients $e_i^\top \Phi_{t,v}^{(k)}/(T-v)$ are $\mathcal{G}_{L_T}$-measurable, and by the same regular conditional distribution argument (EC.3) holds for $\mathcal{G}_{v-1}$-measurable u. Using (EC.3) with $\bar{\sigma}$ as the common proxy, Chernoff's inequality gives

$$\mathbf{1}_{\mathcal{B}_{k,T}} \Pr\left(\left| \sum_{v=L_T+1}^{t-1} e_i^\top \Phi_{t,v}^{(k)} \frac{\xi_v}{T-v} \right| \geq \frac{r_0}{8\sqrt{n}} \middle| \mathcal{G}_{L_T} \right) \leq 2e^{-c_{\mathrm{ex}} h_T} \quad \text{a.s.}$$

A union bound over n coordinates and N_T times gives

$$\mathbf{1}_{\mathcal{B}_{k,T}} \Pr(\mathcal{E}_{\mathrm{ex},k} \mid \mathcal{G}_{L_T}) \leq 2nN_T e^{-c_{\mathrm{ex}} h_T} \quad \text{a.s.}$$

Taking square roots gives

$$\bar{\pi}_{\mathrm{ex},k} \leq \sqrt{2nN_T}\, e^{-c_{\mathrm{ex}} h_T/2}.$$

Finally,

$$N_T \bar{\pi}_{\mathrm{ex},k} \leq C T^{3/2 - c_{\mathrm{ex}} c_h/2} = o(\sqrt{T})$$

because $c_h > 2/c_{\mathrm{ex}}$. □

Lemma 18 (Fixed-design covariances and fourth moments)**.** *Define*

$$\widetilde{E}_{r,T} := \widetilde{\Theta}_{L_T}^{\mathrm{raw}} - \Theta, \qquad \widetilde{E}_{c,T} := \widetilde{\Theta}_{L_T}^{\mathrm{cv}} - \Theta,$$

$$\widetilde{e}_{r,T} := \mathrm{vec}(\widetilde{E}_{r,T}), \qquad \widetilde{e}_{c,T} := \mathrm{vec}(\widetilde{E}_{c,T}),$$

and

$$e_{k,T} := \mathrm{vec}(\widehat{\Theta}^{k}_{L_T} - \Theta).$$

Then

$$L_T \mathbb{E}[\widetilde{e}_{r,T}\widetilde{e}_{r,T}^{\top} \mid \mathcal{F}^0] = V^0_{\mathrm{raw}}.$$

For the surrogate estimator,

$$\widetilde{E}_{c,T} = U_T + W_T,$$

where

$$U_T = \left(\sum_{s=1}^{L_T} [\xi^{\star}_s - \delta^{\Gamma}_{j(s)} Z_s] z_s^{\top}\right) V_L^{-1},$$

$$W_T = \left(\sum_{s=1}^{L_T} \widehat{\Gamma}_{j(s)} \delta^{m,+}_{j(s)} z_s^{\top}\right) V_L^{-1}.$$

Conditional on $\mathcal{F}^0$*,*

$$L_T \mathbb{E}[\widetilde{e}_{c,T}\widetilde{e}_{c,T}^{\top} \mid \mathcal{F}^0] = V^0_{\mathrm{cv}} + R_{\Gamma,T} + R_{m,T},$$

where

$$R_{\Gamma,T} := (G^{-1}_{\mathrm{design}} \otimes I_n) \left\{ \frac{1}{J_0} \sum_{j=1}^{J_0} z_j z_j^{\top} \otimes \delta^{\Gamma}_j \Sigma_S (\delta^{\Gamma}_j)^{\top} \right\} (G^{-1}_{\mathrm{design}} \otimes I_n),$$

and

$$R_{m,T} := L_T \,\mathrm{vec}(W_T)\,\mathrm{vec}(W_T)^{\top}.$$

Moreover,

$$\mathbb{E}\|\widetilde{e}_{r,T}\|_2^4 + \mathbb{E}\|\widetilde{e}_{c,T}\|_2^4 + \mathbb{E}\|e_{\mathrm{raw},T}\|_2^4 + \mathbb{E}\|e_{\mathrm{cv},T}\|_2^4 \le \frac{C}{L_T^2}.$$

Let

$$A_{\mathrm{cv}} := \left\{ \Pi_{\mathfrak{T}}^{\|\cdot\|_F}(\widetilde{\Theta}^{\mathrm{cv}}_{L_T}) \neq \widetilde{\Theta}^{\mathrm{cv}}_{L_T} \right\} \subseteq \{\|\widetilde{E}_{c,T}\|_F > r_{\mathfrak{T}}\}.$$

Let

$$A_{\mathrm{raw}} := \left\{ \Pi_{\mathfrak{T}}^{\|\cdot\|_F}(\widetilde{\Theta}^{\mathrm{raw}}_{L_T}) \neq \widetilde{\Theta}^{\mathrm{raw}}_{L_T} \right\} \subseteq \{\|\widetilde{E}_{r,T}\|_F > r_{\mathfrak{T}}\}.$$

Then

$$\Pr(A_{\mathrm{cv}}) \le T^{-4} + \delta_{\mathrm{cal}}(T),$$

and

$$\left\| L_T \mathbb{E}[e_{\mathrm{raw},T} e_{\mathrm{raw},T}^{\top}] - V^0_{\mathrm{raw}} \right\|_{\mathrm{op}} \le C T^{-2}, \tag{EC.8}$$

$$\left\| L_T \mathbb{E}[e_{\mathrm{cv},T} e_{\mathrm{cv},T}^{\top}] - V^0_{\mathrm{cv}} \right\|_{\mathrm{op}} \le C \left(\eta^2_{\Gamma,T} + L_T \eta^2_{m,T} + T^{-2} + \Pr(A_{\mathrm{cv}})^{1/2} \right). \tag{EC.9}$$

Proof of Lemma 18. Column-wise vectorization gives

$$\widetilde{e}_{r,T} = (V_L^{-1} \otimes I_n) \sum_{s=1}^{L_T} (z_s \otimes I_n) \xi_s.$$

For $s < u$, the enlarged-filtration centering gives

$$\mathbb{E}[\xi_u \xi_s^{\top} \mid \mathcal{F}^0] = \mathbb{E}\big[\mathbb{E}[\xi_u \mid \mathcal{G}_{u-1}] \xi_s^{\top} \big| \mathcal{F}^0\big] = 0.$$

Therefore

$$\mathbb{E}[\widetilde{e}_{r,T}\widetilde{e}_{r,T}^{\top} \mid \mathcal{F}^0] = (V_L^{-1} \otimes I_n)\left\{\sum_{s=1}^{L_T} z_s z_s^{\top} \otimes \Sigma_d\right\}(V_L^{-1} \otimes I_n)$$
$$= (V_L^{-1} \otimes I_n)(V_L \otimes \Sigma_d)(V_L^{-1} \otimes I_n).$$

Since $V_L = L_T G_{\text{design}}$, multiplication by L_T yields V_{raw}^0.

The pseudo-observation identity is

$$\widetilde{D}_s - f(p^{(j(s))}) = \xi_s - \widehat{\Gamma}_{j(s)}\{Z_s - \delta_{j(s)}^{m,+}\}$$
$$= \xi_s^{\star} - \delta_{j(s)}^{\Gamma} Z_s + \widehat{\Gamma}_{j(s)}\delta_{j(s)}^{m,+}.$$

Because

$$\Gamma^{\star}(p^{(j)}) = \Sigma_{dS}\Sigma_S^{-1},$$

we have

$$\mathbb{E}[\xi_s^{\star} Z_s^{\top} \mid \mathcal{G}_{s-1}, \mathcal{F}^0] = \Sigma_{dS} - \Gamma^{\star}(p^{(j(s))})\Sigma_S = 0.$$

Consequently,

$$\mathbb{E}[(\xi_s^{\star} - \delta_{j(s)}^{\Gamma} Z_s)(\xi_s^{\star} - \delta_{j(s)}^{\Gamma} Z_s)^{\top} \mid \mathcal{G}_{s-1}, \mathcal{F}^0]$$
$$= \Sigma_{\text{eff}} + \delta_{j(s)}^{\Gamma}\Sigma_S(\delta_{j(s)}^{\Gamma})^{\top}.$$

The martingale cross terms between distinct exploration periods vanish. The cross term between U_T and the $\mathcal{F}^0$-measurable matrix W_T also vanishes. This proves the conditional covariance decomposition.

For the fourth-moment bound, let $w \in \mathbb{R}^{nd}$ be a unit vector. Then

$$w^{\top}\widetilde{e}_{r,T} = \sum_{s=1}^{L_T} a_s^{\top}\xi_s, \qquad a_s := (z_s^{\top}V_L^{-1} \otimes I_n)w.$$

The conditional mgf iteration and (EC.2) with $v = 0$ imply that this scalar is sub-Gaussian with proxy

$$K_{\text{camp}}^2 \sum_{s=1}^{L_T} \|a_s\|_2^2 = K_{\text{camp}}^2 w^{\top}(V_L^{-1} \otimes I_n)w \le \frac{K_{\text{camp}}^2}{\kappa_0 L_T}.$$

If X is centered sub-Gaussian with proxy v, then

$$\mathbb{E}[X^4] = \int_0^{\infty} 4t^3 \Pr(|X| > t)\, dt \le 16v^2.$$

Using

$$\|x\|_2^4 \le nd \sum_{a=1}^{nd} x_a^4$$

gives

$$\mathbb{E}\|\widetilde{e}_{r,T}\|_2^4 \le \frac{C}{L_T^2}.$$

For the centered surrogate term, the joint mgf bound with $v = -\Gamma^{\top}u$, $\|\Gamma\|_{\text{op}} \le M_{\Gamma}$, gives a conditional sub-Gaussian proxy bounded by

$$CK_{\text{camp}}^2(1 + M_{\Gamma}^2).$$

The same argument gives

$$\mathbb{E}\| \operatorname{vec}(U_T)\|_2^4 \le \frac{C}{L_T^2}.$$

For the center term, the proof of Lemma 17 gives the deterministic bound $\|W_T\|_F \le (M_\Gamma/\kappa_0)\max_j \|z_j\|_2 \max_j \|\delta_j^{m,+}\|_2$. Because the fourth moment of a sample mean of independent centered sub-Gaussian vectors is $O(C_{m,T}^{-2})$,

$$\mathbb{E}\|W_T\|_F^4 \le \frac{C}{C_{m,T}^2} \le \frac{C}{L_T^2},$$

and therefore

$$\mathbb{E}\|\widetilde{e}_{c,T}\|_2^4 \le \frac{C}{L_T^2}.$$

Projection onto a closed convex set containing Θ is nonexpansive, so the same fourth-moment bounds hold for the projected errors.

We now use the fourth-moment bounds to control the contributions from calibration failure and projection. The event A_{cv} is contained in $\{\|\widetilde{E}_{c,T}\|_F > r_{\mathfrak{T}}\}$. On C_T, the conditional finite-net argument in the proof of Lemma 17 gives

$$\Pr(A_{\mathrm{cv}} \mid C_T) \le T^{-4}.$$

Thus

$$\Pr(A_{\mathrm{cv}}) \le T^{-4} + \delta_{\mathrm{cal}}(T).$$

Let

$$p_{\mathrm{cal}} := \Pr(C_T^c).$$

The signed projection decomposition is

$$\begin{aligned} & L_T\mathbb{E}[e_{\mathrm{cv},T}e_{\mathrm{cv},T}^\top] - V_{\mathrm{cv}}^0 \\ &= \mathbb{E}[\mathbf{1}_{C_T}R_C] - p_{\mathrm{cal}}V_{\mathrm{cv}}^0 \\ &\quad + L_T\mathbb{E}\Big[\mathbf{1}_{C_T}\left(e_{\mathrm{cv},T}e_{\mathrm{cv},T}^\top - \widetilde{e}_{c,T}\widetilde{e}_{c,T}^\top\right)\Big] + L_T\mathbb{E}[\mathbf{1}_{C_T^c}e_{\mathrm{cv},T}e_{\mathrm{cv},T}^\top], \end{aligned} \tag{EC.10}$$

where R_C contains the conditional covariance remainders on C_T. The coefficient-estimation part satisfies

$$\|\mathbb{E}[\mathbf{1}_{C_T}R_C]\|_{\mathrm{op}} \le C\eta_{\Gamma,T}^2,$$

and the coherent center contribution satisfies

$$\text{on } C_T, \qquad \|R_{m,T}\|_{\mathrm{op}} \le L_T\|W_T\|_F^2 \le CL_T\eta_{m,T}^2.$$

The calibration-failure subtraction satisfies

$$p_{\mathrm{cal}}\|V_{\mathrm{cv}}^0\|_{\mathrm{op}} \le CT^{-4}.$$

Since $\Theta \in \mathfrak{T}$ and the projection onto $\mathfrak{T}$ is nonexpansive, $\|e_{\mathrm{cv},T}\|_2 \le \|\widetilde{e}_{c,T}\|_2$, so it suffices to bound the term for $\widetilde{e}_{c,T}$. Fourth-moment Cauchy–Schwarz gives

$$\begin{aligned} L_T\mathbb{E}[\|\widetilde{e}_{c,T}\|_2^2\mathbf{1}_{C_T^c}] &\le L_T\big(\mathbb{E}\|\widetilde{e}_{c,T}\|_2^4\big)^{1/2}p_{\mathrm{cal}}^{1/2} \\ &\le Cp_{\mathrm{cal}}^{1/2}. \end{aligned}$$

Likewise, on the projection-active event,

$$\begin{aligned} L_T \left\| \mathbb{E}\Big[\mathbf{1}_{A_{\mathrm{cv}}} \left(e_{\mathrm{cv},T} e_{\mathrm{cv},T}^\top - \widetilde{e}_{c,T} \widetilde{e}_{c,T}^\top \right) \Big] \right\|_{\mathrm{op}} \\ \le L_T \left(\mathbb{E}\|e_{\mathrm{cv},T}\|_2^4 + \mathbb{E}\|\widetilde{e}_{c,T}\|_2^4 \right)^{1/2} \Pr(A_{\mathrm{cv}})^{1/2} \\ \le C \Pr(A_{\mathrm{cv}})^{1/2}. \end{aligned}$$

Combining these bounds yields (EC.9).

For the raw estimator, the interior-radius assumption and the raw tail imply that the projection-active event A_{raw} has probability at most T^{-4}. Therefore

$$L_T \mathbb{E}[\|\widetilde{e}_{r,T}\|_2^2 \mathbf{1}_{A_{\mathrm{raw}}}] \le L_T \left(\mathbb{E}\|\widetilde{e}_{r,T}\|_2^4 \right)^{1/2} \Pr(A_{\mathrm{raw}})^{1/2} \le C T^{-2}.$$

This proves (EC.8). □

Lemma 19 (Good freeze event, virtual process and exit charge)**.** *For $k \in \{\mathrm{raw}, \mathrm{cv}\}$, the good freeze event $\mathcal{B}_{k,T}$, the analysis stopping time $\tau_{\mathrm{ex},k}$, the exit event $\mathcal{E}_{\mathrm{ex},k}$, and the clipped process $\widetilde{q}_{t,k}$ are the globally defined objects above. On $\mathcal{B}_{k,T}$, the unstopped process $\widetilde{q}^u_{t,k}$ starts at x_k. On its complement it starts at x^*. The clipped process equals $\widetilde{q}^u_{t,k}$ before $\tau_{\mathrm{ex},k}$ and equals x^* thereafter, so it has a full-path L^2 bound.*

Define the clipped virtual revenue used in the assembly

$$R_{k,T}^{\mathrm{virt}} := \sum_{t=L_T+1}^{T-h_T} p_k(\widetilde{q}_{t,k})^\top \{ g_k(\widetilde{q}_{t,k}) + \xi_t \},$$

the truncated virtual revenue

$$R_{k,T}^{\mathrm{tr,virt}} := \sum_{t=L_T+1}^{T-h_T} \mathbf{1}_{\mathcal{B}_{k,T}} \mathbf{1}_{\{t<\tau_{\mathrm{ex},k}\}} p_k(\widetilde{q}_{t,k})^\top \{ g_k(\widetilde{q}_{t,k}) + \xi_t \},$$

and the actual post-exploration revenue

$$R_{k,T}^{\mathrm{act}} := \sum_{t=L_T+1}^{T-h_T} (p_t^k)^\top \widehat{D}_t^k.$$

Define the virtual regret by

$$\mathrm{Reg}_{T,k}^{\mathrm{virt}} := \mathbf{1}_{\mathcal{B}_{k,T}} \sum_{t=L_T+1}^{T-h_T} \left[\ell(x^*) - p_k(\widetilde{q}_{t,k})^\top \{ g_k(\widetilde{q}_{t,k}) + \xi_t \} \right].$$

There is a constant C such that

$$\Pr(\mathcal{B}_{k,T}^c) \le C T^{-3} + \delta_{\mathrm{cal}}(T),$$

and

$$\mathbb{E}[(R_{k,T}^{\mathrm{virt}})^2] + \mathbb{E}[(R_{k,T}^{\mathrm{tr,virt}})^2] + \mathbb{E}[(R_{k,T}^{\mathrm{act}})^2] \le C N_T^2.$$

Furthermore,

$$\mathbb{E}\left[|R_{k,T}^{\mathrm{act}} - R_{k,T}^{\mathrm{tr,virt}}| \right] \le C N_T \left\{ \Pr(\mathcal{B}_{k,T}^c)^{1/2} + \bar{\pi}_{\mathrm{ex},k} \right\} = o(\sqrt{T}),$$

and

$$\mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}} |R_{k,T}^{\mathrm{virt}} - R_{k,T}^{\mathrm{tr,virt}}| \Big] \le C N_T \bar{\pi}_{\mathrm{ex},k} = o(\sqrt{T}).$$

Proof of Lemma 19. The probability bound for $\mathcal{B}_{k,T}^c$ is the union of the calibration failure, estimator failure and load failure bounds from Lemmas 15 and 17. On the intersection of these three events, the geometry gate holds for all sufficiently large T by Lemma 17.

On $\mathcal{B}_{k,T}$, put

$$z_{t,k}^u := \widetilde{q}_{t,k}^u - x_k, \qquad a_k := g_k(x_k) - x_k.$$

The unstopped recursion gives, for all $t \le T - h_T$,

$$z_{t+1,k}^u = \left(I - \frac{M_k}{T-t}\right) z_{t,k}^u - \frac{a_k + \xi_t}{T-t}.$$

By definition, $\widetilde{q}_{t,k} = \widetilde{q}_{t,k}^u$ before $\tau_{\mathrm{ex},k}$ and is x^* after that time. The exit event is the one in Lemma 17, which gives the stated bounds on $\Pr(\mathcal{E}_{\mathrm{ex},k} \mid \mathcal{G}_{L_T})$ and $\bar{\pi}_{\mathrm{ex},k}$.

Because the virtual path is clipped at x^* after exit, the virtual price, mean demand and noise terms have uniformly bounded second moments. In particular,

$$\mathbb{E}[(R_{k,T}^{\mathrm{virt}})^2] \le N_T \sum_{t=L_T+1}^{T-h_T} \mathbb{E}\left[p_k(\widetilde{q}_{t,k})^\top \{g_k(\widetilde{q}_{t,k}) + \xi_t\}\right]^2 \le C N_T^2.$$

The same bound holds for $R_{k,T}^{\mathrm{tr,virt}}$. The bounded-potential condition gives the corresponding bound for the actual revenue.

On $\mathcal{B}_{k,T} \cap \mathcal{E}_{\mathrm{ex},k}^c$, the actual and virtual paths coincide through time $T - h_T$. Indeed, by induction on t, if $q_{t,k} = \widetilde{q}_{t,k}$, then $\|q_{t,k} - x^*\|_2 \le r_0/2$, so the active test passes, Lemma 17 gives full service, and the two recursions take the same step. Hence

$$R_{k,T}^{\mathrm{act}} = R_{k,T}^{\mathrm{tr,virt}}.$$

On $\mathcal{B}_{k,T} \cap \mathcal{E}_{\mathrm{ex},k}$, both revenues are pathwise bounded by CN_T, hence

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}} \mathbf{1}_{\mathcal{E}_{\mathrm{ex},k}} |R_{k,T}^{\mathrm{act}} - R_{k,T}^{\mathrm{tr,virt}}|\Big] \\
&\quad \le C N_T \, \mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}} \Pr(\mathcal{E}_{\mathrm{ex},k} \mid \mathcal{G}_{L_T})\big] \\
&\quad \le C N_T \bar{\pi}_{\mathrm{ex},k}.
\end{aligned}$$

On $\mathcal{B}_{k,T}^c$, Cauchy–Schwarz and the full-path L^2 bounds give

$$\mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}^c} |R_{k,T}^{\mathrm{act}} - R_{k,T}^{\mathrm{tr,virt}}|\Big] \le C N_T \Pr(\mathcal{B}_{k,T}^c)^{1/2}.$$

This proves the actual-to-truncated bound.

Finally, on $\mathcal{B}_{k,T}$, the clipped and truncated virtual revenues differ only after $\tau_{\mathrm{ex},k}$. Since the virtual process is clipped after exit,

$$\left|R_{k,T}^{\mathrm{virt}} - R_{k,T}^{\mathrm{tr,virt}}\right| \le C N_T \mathbf{1}_{\mathcal{E}_{\mathrm{ex},k}}.$$

The tower argument used for the actual-to-truncated bound then gives $\mathbb{E}[\mathbf{1}_{\mathcal{B}_{k,T}} |R_{k,T}^{\mathrm{virt}} - R_{k,T}^{\mathrm{tr,virt}}|] \le C N_T \bar{\pi}_{\mathrm{ex},k}$. Since $c_h > 2/c_{\mathrm{ex}}$, every displayed bound is $o(\sqrt{T})$. □

Lemma 20 (Second-order expansion)**.** *For $k \in \{\mathrm{raw}, \mathrm{cv}\}$, define*

$$e_{k,T} := \mathrm{vec}(E_{k,T}), \qquad \varepsilon_{k,T} := \|E_{k,T}\|_F, \qquad z_{t,k}^u := \widetilde{q}_{t,k}^u - x_k.$$

On $\mathcal{B}_{k,T}$ before exit,

$$x_k - x^* = S_c P_J \delta c_T + \mathsf{S} e_{k,T} + R_{x,k},$$

with

$$\|R_{x,k}\|_2 \le C(\varepsilon_{k,T} + \Delta_T)^2,$$

and

$$a_k := g_k(x_k) - x_k = \mathsf{G} e_{k,T} + R_{a,k},$$

with

$$\|R_{a,k}\|_2 \le C(\varepsilon_{k,T}^2 + \varepsilon_{k,T}\Delta_T).$$

Define

$$\Phi_{t,v}^{(k)} := \prod_{u=v+1}^{t-1} \left(I - \frac{M_k}{T-u}\right),$$

$$\zeta_{t,k} := -\sum_{v=L_T+1}^{t-1} \frac{\Phi_{t,v}^{(k)} \xi_v}{T-v},$$

$$K_{t,k} := I - (I + M_k) \sum_{v=L_T+1}^{t-1} \frac{\Phi_{t,v}^{(k)}}{T-v}.$$

Then

$$g_k(x_k + z_{t,k}^u) - x^* = S_c P_J \delta c_T + \mathsf{S} e_{k,T} + K_{t,k} \mathsf{G} e_{k,T} + \rho_{t,k} + (I + M_k)\zeta_{t,k}, \tag{EC.11}$$

where

$$\|\rho_{t,k}\|_2 \le C\left[(\varepsilon_{k,T}^2 + \varepsilon_{k,T}\Delta_T)\eta_T + \Delta_T^2\right], \qquad \eta_T := 1 + \log \frac{N_T}{h_T}.$$

Moreover,

$$\mathbb{E}[\zeta_{t,k} \mid \mathcal{G}_{L_T}] = 0,$$

and

$$\sum_{t=L_T+1}^{T-h_T} \mathbb{E}[\|\zeta_{t,k}\|_2^2 \mid \mathcal{G}_{L_T}] \le C\eta_T.$$

Let

$$\mathcal{R}_{k,T}^{\mathrm{cov}} := C\frac{N_T}{L_T}\left((\eta_{\Gamma,T}^2 + L_T\eta_{m,T}^2)\mathbf{1}_{\{k=\mathrm{cv}\}} + T^{-2} + \Pr(A_k)^{1/2}\right) + CN_T L_T^{-1} \Pr(\mathcal{B}_{k,T}^c)^{1/2}.$$

Here A_k is A_{raw} for $k = \mathrm{raw}$ and A_{cv} for $k = \mathrm{cv}$. Then

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{B}_{k,T}} \sum_{t=L_T+1}^{T-h_T} \{g_k(\widetilde{q}_{t,k}) - x^*\}^\top Q_0 \{g_k(\widetilde{q}_{t,k}) - x^*\}\right] - \frac{N_T}{L_T}\operatorname{tr}(Q_{\mathrm{dyn}} V_k^0)\right| \\
&\quad \le C\left\{N_T L_T^{-3/2}\eta_T^3 + N_T\overline{\Delta}_T L_T^{-1/2}\eta_T^3 + N_T\overline{\Delta}_T^2\eta_T^3 + \eta_T^2 + \frac{(1+\log T)^3}{L_T} + \frac{h_T}{L_T}\right\} + \mathcal{R}_{k,T}^{\mathrm{cov}} + CN_T^2\eta_T^2 e^{-c_{\mathrm{ex}}h_T},
\end{aligned} \tag{EC.12}$$

where V_{raw}^0 is used for $k = \mathrm{raw}$ and V_{cv}^0 for $k = \mathrm{cv}$. Every term on the right-hand side is $o(\sqrt{T})$.

Proof of Lemma 20. Lemma 16 gives

$$x_k - x^* = S_c P_J \delta c_T + \mathsf{S} e_{k,T} + R_{x,k},$$

with the stated bound on $R_{x,k}$.

Define

$$a(\vartheta, c) := \alpha + BB_{\vartheta}^{-1}\{X(\vartheta, c) - \alpha_{\vartheta}\} - X(\vartheta, c).$$

At $(\Theta, \bar{c})$, $a(\Theta, \bar{c}) = 0$. Also, $a(\Theta, c) \equiv 0$ for nearby c, because the true demand map satisfies $g(x) = x$ at its fluid solution. Differentiating in the parameter direction $E = (u, \dot{B})$ gives

$$D_{\vartheta} a(\Theta, \bar{c})[E] = -u - \dot{B}p^* = G(E) = \mathsf{G}\,\mathrm{vec}(E).$$

The C^2 Taylor formula yields

$$a_k = \mathsf{G}e_{k,T} + R_{a,k},$$

with

$$\|R_{a,k}\|_2 \le C(\varepsilon_{k,T}^2 + \varepsilon_{k,T}\Delta_T).$$

The recursion for $z_{t,k}^u$ is

$$z_{t+1,k}^u = \left(I - \frac{M_k}{T-t}\right) z_{t,k}^u - \frac{a_k + \xi_t}{T-t}.$$

Variation of constants gives

$$z_{t,k}^u = -\sum_{v=L_T+1}^{t-1} \frac{\Phi_{t,v}^{(k)} a_k}{T-v} + \zeta_{t,k}.$$

Since

$$g_k(x_k + z) - x^* = (x_k - x^*) + a_k + (I + M_k)z,$$

substitution gives (EC.11) with

$$\rho_{t,k} = R_{x,k} + K_{t,k} R_{a,k}.$$

The geometry gate implies

$$\|\Phi_{t,v}^{(k)}\|_{\mathrm{op}} \le 2, \qquad \sum_{v=L_T+1}^{t-1} \frac{1}{T-v} \le \eta_T,$$

which proves the displayed bound on $\rho_{t,k}$.

Conditional centering gives

$$\mathbb{E}[\zeta_{t,k} \mid \mathcal{G}_{L_T}] = 0.$$

Martingale orthogonality and the product bound imply

$$\mathbb{E}[\|\zeta_{t,k}\|_2^2 \mid \mathcal{G}_{L_T}] \le C \sum_{v=L_T+1}^{t-1} \frac{1}{(T-v)^2}.$$

Summing over t gives

$$\sum_{t=L_T+1}^{T-h_T} \mathbb{E}[\|\zeta_{t,k}\|_2^2 \mid \mathcal{G}_{L_T}] \le C \sum_{r=h_T}^{N_T} \frac{1}{r} \le C\eta_T.$$

Put

$$s_t := \sum_{v=L_T+1}^{t-1} \frac{1}{T-v}, \qquad \kappa_t := 1 - s_t.$$

When $M_k = 0$, $K_{t,k} = \kappa_t I$. The product expansion gives

$$\|K_{t,k} - \kappa_t I\|_{\mathrm{op}} \le C\varepsilon_{k,T}(s_t + s_t^2).$$

Since $s_t \le \eta_T$, the resulting quadratic-kernel remainder satisfies

$$\sum_{t=L_T+1}^{T-h_T} K_{t,k}^\top Q_0 K_{t,k} = N_T Q_0 + R_{Q,k},$$

with

$$\|R_{Q,k}\|_{\mathrm{op}} \le C\left\{(1+\log T)^3 + N_T \varepsilon_{k,T} \eta_T^3\right\}.$$

Indeed,

$$\sum_{t=L_T+1}^{T-h_T} s_t^3 \le N_T \eta_T^3,$$

and the deterministic harmonic-sum errors are $O(\log^3 T)$.

The quadratic term involving the price-drift component is given by

$$e_{k,T}^\top \mathsf{G}^\top \left(\sum_{t=L_T+1}^{T-h_T} K_{t,k}^\top Q_0 K_{t,k} \right) \mathsf{G} e_{k,T}.$$

The sensitivity component contributes

$$(N_T - h_T) e_{k,T}^\top \mathsf{S}^\top Q_0 \mathsf{S} e_{k,T}.$$

Replacing $N_T - h_T$ by N_T costs at most

$$h_T \mathbb{E}\|e_{k,T}\|_2^2 \le \frac{C h_T}{L_T}.$$

Using

$$\mathbb{E}\|e_{k,T}\|_2^q \le C L_T^{-q/2}, \qquad q = 2,3,4,$$

the $R_{Q,k}$ contribution is bounded by

$$C\left\{ \frac{(1+\log T)^3}{L_T} + N_T L_T^{-3/2} \eta_T^3 \right\}.$$

The cross term between the sensitivity and dynamic components is bounded by

$$C\left\{ \frac{(1+\log T)^2}{L_T} + N_T L_T^{-3/2} \eta_T^2 + N_T \overline{\Delta}_T L_T^{-1/2} \eta_T^3 \right\}.$$

The capacity square and its cross term with the estimator terms are bounded by

$$C\left\{ N_T \overline{\Delta}_T^2 \eta_T^3 + N_T \overline{\Delta}_T L_T^{-1/2} \eta_T^3 \right\}.$$

All terms involving $(I + M_k)\zeta_{t,k}$ are handled by conditional centering. Their total quadratic contribution is at most

$$C\eta_T,$$

and their cross terms with the deterministic components vanish after conditioning on $\mathcal{G}_{L_T}$.

The fourth-order remainder terms involving $\rho_{t,k}$ satisfy

$$\mathbb{E}\sum_t \|\rho_{t,k}\|_2^2 \le C N_T \left[L_T^{-2} \eta_T^2 + \overline{\Delta}_T^2 L_T^{-1} \eta_T^2 + \overline{\Delta}_T^4 \right],$$

and their cross terms with the first-order components are bounded by

$$C\left\{N_T L_T^{-3/2}\eta_T^3 + N_T\overline{\Delta}_T L_T^{-1/2}\eta_T^3 + N_T\overline{\Delta}_T^2\eta_T^3\right\}.$$

For the clipped process in the statement, the preceding calculation applies to $\widetilde{q}^u$. On $\mathcal{B}_{k,T}$, the clipped and unstopped paths differ only on $\mathcal{E}_{\mathrm{ex},k}$. By the variation-of-constants formula and the product bound in the gate, $\max_{t\le T-h_T+1}\|\widetilde{q}^u_{t,k}-x^*\|_2 \le C\eta_T$ on $\mathcal{B}_{k,T}$. $\|D_t\|_2 \le \sqrt{n}D_{\max}$ and $\|\xi_t\|_2 \le \sqrt{n}D_{\max} + \sup_{p\in\mathcal{P}}\|f(p)\|_2$ pathwise, and the fitted inverse is bounded on the learning class. Hence the pathwise difference between the corresponding sums of the linear and quadratic loss terms is at most

$$CN_T\eta_T^2\mathbf{1}_{\mathcal{E}_{\mathrm{ex},k}}.$$

Using the conditional exit bound in Lemma 17,

$$\mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}}N_T\eta_T^2\mathbf{1}_{\mathcal{E}_{\mathrm{ex},k}}\big] \le CN_T^2\eta_T^2 e^{-c_{\mathrm{ex}}h_T} = O\Big(T^{2-c_{\mathrm{ex}}c_h}\log^2 T\Big) = o(1).$$

This is the explicit post-exit charge in (EC.12).

Finally, Lemma 18 gives

$$L_T\mathbb{E}[e_{k,T}e_{k,T}^\top] = V_k^0 + \mathcal{R}_{k,T}^{\mathrm{cov,mat}},$$

where

$$\|\mathcal{R}_{k,T}^{\mathrm{cov,mat}}\|_{\mathrm{op}} \le C\left((\eta_{\Gamma,T}^2 + L_T\eta_{m,T}^2)\mathbf{1}_{\{k=\mathrm{cv}\}} + T^{-2} + \Pr(A_k)^{1/2}\right).$$

Multiplying by N_T/L_T gives the first part of $\mathcal{R}_{k,T}^{\mathrm{cov}}$. The contribution of $\mathcal{B}_{k,T}^c$ is bounded by fourth-moment Cauchy–Schwarz:

$$\frac{N_T}{L_T}\left\|L_T\mathbb{E}[\mathbf{1}_{\mathcal{B}_{k,T}^c}e_{k,T}e_{k,T}^\top]\right\|_{\mathrm{op}} \le CN_TL_T^{-1}\Pr(\mathcal{B}_{k,T}^c)^{1/2}.$$

This is the second part of $\mathcal{R}_{k,T}^{\mathrm{cov}}$. Since $L_T \asymp \sqrt{T}$, $h_T = O(\log T)$, $\overline{\Delta}_T = O(L_T/T)$, $\eta_{\Gamma,T}^2 + L_T\eta_{m,T}^2 \le C/(1+\log T)$ by Lemma 15, and $\Pr(\mathcal{B}_{k,T}^c) = o(1)$, every displayed remainder is $o(\sqrt{T})$. □

Proof of Theorem 7

Proof of Theorem 7. Fix $k \in \{\mathrm{raw}, \mathrm{cv}\}$. Let R_T^{exp} denote the common exploration revenue and define

$$C_{\mathrm{exp},T} := L_T\ell(x^*) - \mathbb{E}[R_T^{\mathrm{exp}}].$$

Let $R_{k,T}^{\mathrm{term}}$ denote the terminal-window revenue of policy k and define

$$T_{k,T} := h_T\ell(x^*) - \mathbb{E}[R_{k,T}^{\mathrm{term}}].$$

Because the terminal capacities can differ across policies, $T_{\mathrm{raw},T}$ and $T_{\mathrm{cv},T}$ can differ. Bounded prices and bounded-potential requests imply

$$|T_{k,T}| \le Ch_T, \qquad |T_{\mathrm{raw},T} - T_{\mathrm{cv},T}| \le Ch_T.$$

For $q \in \mathbb{R}^n$, the affine relation

$$g_k(q) = \alpha + Bp_k(q)$$

implies

$$p_k(q)^\top g_k(q) = g_k(q)^\top B^{-1}(g_k(q) - \alpha) = \ell(g_k(q)).$$

Because

$$\ell(y) = \ell(x^*) + \nabla\ell(x^*)^\top(y - x^*) - (y - x^*)^\top Q_0(y - x^*),$$

and

$$\nabla\ell(x^*) = A_J^\top \lambda_J^*,$$

we have the exact identity

$$\ell(x^*) - p_k(q)^\top g_k(q) = \lambda_J^{*\top} A_J(x^* - g_k(q)) + (g_k(q) - x^*)^\top Q_0(g_k(q) - x^*). \tag{EC.13}$$

The global unstopped virtual recursion satisfies, for every $L_T + 1 \le t \le T - h_T$,

$$(T - t + 1)\widetilde{q}^u_{t,k} - (T - t)\widetilde{q}^u_{t+1,k} = g_k(\widetilde{q}^u_{t,k}) + \xi_t.$$

Therefore,

$$\sum_{t=L_T+1}^{T-h_T} \widetilde{D}^{u,\mathrm{virt}}_{t,k} = N_T x_k - h_T \widetilde{q}^u_{T-h_T+1,k}, \tag{EC.14}$$

where

$$\widetilde{D}^{u,\mathrm{virt}}_{t,k} := g_k(\widetilde{q}^u_{t,k}) + \xi_t.$$

This identity keeps $N_T - h_T$ and N_T distinct. On $\mathcal{B}_{k,T}$, the clipped and unstopped paths differ only on $\mathcal{E}_{\mathrm{ex},k}$, and the linear loss difference is bounded by

$$\left| \sum_{t=L_T+1}^{T-h_T} \lambda_J^{*\top} A_J \{ g_k(\widetilde{q}_{t,k}) - g_k(\widetilde{q}^u_{t,k}) \} \right| \le C N_T \eta_T^2 \mathbf{1}_{\mathcal{E}_{\mathrm{ex},k}}.$$

Its expectation is at most

$$C N_T^2 \eta_T^2 e^{-c_{\mathrm{ex}} h_T} = o(1).$$

The event $\mathcal{B}_{k,T}$ is $\mathcal{G}_{L_T}$-measurable. For $t > L_T$, $p_k(\widetilde{q}_{t,k})$ is predictable with respect to $\mathcal{G}_{t-1}$. Therefore, by (EC.3),

$$\mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}} p_k(\widetilde{q}_{t,k})^\top \xi_t\big] = \mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}} p_k(\widetilde{q}_{t,k})^\top \mathbb{E}[\xi_t \mid \mathcal{G}_{t-1}]\big] = 0. \tag{EC.15}$$

This is the zero-mean noise-revenue step used below.

On $\mathcal{B}_{k,T}$, the active constraints satisfy

$$A_J x_k = (b_T)_J, \qquad A_J x^* = \bar{c}_J.$$

Taking the conditional expectation of the telescoping identity gives

$$\begin{aligned} &\mathbb{E}\left[\mathbf{1}_{\mathcal{B}_{k,T}} \sum_{t=L_T+1}^{T-h_T} \lambda_J^{*\top} A_J(x^* - g_k(\widetilde{q}_{t,k}))\right] \\ &\quad = \mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}} \Big\{ N_T \lambda_J^{*\top}(\bar{c}_J - (b_T)_J) + h_T \lambda_J^{*\top} \Big(A_J \mathbb{E}[\widetilde{q}^u_{T-h_T+1,k} \mid \mathcal{G}_{L_T}] - \bar{c}_J \Big) \Big\} \Big] + \mathfrak{r}^{\mathrm{lin,exit}}_{k,T} \end{aligned} \tag{EC.16}$$

where

$$|\mathfrak{r}^{\mathrm{lin,exit}}_{k,T}| \le C N_T^2 \eta_T^2 e^{-c_{\mathrm{ex}} h_T} = o(1).$$

On $\mathcal{B}_{k,T}$, the noise part of $\widetilde{q}^u_{T-h_T+1,k}$ has conditional mean zero given $\mathcal{G}_{L_T}$, so $\mathbb{E}[\widetilde{q}^u_{T-h_T+1,k} \mid \mathcal{G}_{L_T}]$ lies within $r_0/4 + r_0/8$ of x^* by the gate and the drift bound. Hence the endpoint term is $O(h_T)$ on all of $\mathcal{B}_{k,T}$. The clipped–unstopped linear charge is bounded by $C N_T^2 \eta_T^2 e^{-c_{\mathrm{ex}} h_T} = o(1)$ using Lemma 17 and Lemma 19.

Applying Lemma 20 to the quadratic term in (EC.13) gives

$$\begin{aligned}\mathbb{E}\Bigg[\mathbf{1}_{\mathcal{B}_{k,T}} \sum_{t=L_T+1}^{T-h_T} (g_k(\widetilde{q}_{t,k}) - x^*)^\top Q_0 (g_k(\widetilde{q}_{t,k}) - x^*)\Bigg] \\ = \frac{N_T}{L_T} \operatorname{tr}(Q_{\mathrm{dyn}} V_k^0) + \mathfrak{r}_{k,T}^{(2)},\end{aligned} \tag{EC.17}$$

where $|\mathfrak{r}_{k,T}^{(2)}|$ is at most the right side of (EC.12), which is $o(\sqrt{T})$.

The virtual regret is random, so we take expectations when combining the terms. By the definition of $\mathrm{Reg}_{T,k}^{\mathrm{virt}}$,

$$\begin{aligned}\mathrm{Reg}_T^{\mathrm{fl}}(\pi_T^k) = C_{\mathrm{exp},T} + T_{k,T} + (N_T - h_T)\ell(x^*)\Pr(\mathcal{B}_{k,T}^c) + \mathbb{E}[\mathrm{Reg}_{T,k}^{\mathrm{virt}}] \\ + \mathbb{E}\Big[\mathbf{1}_{\mathcal{B}_{k,T}}\Big(R_{k,T}^{\mathrm{virt}} - R_{k,T}^{\mathrm{tr,virt}}\Big)\Big] + \mathbb{E}[R_{k,T}^{\mathrm{tr,virt}} - R_{k,T}^{\mathrm{act}}].\end{aligned} \tag{EC.18}$$

Indeed, on $\mathcal{B}_{k,T}$,

$$\mathrm{Reg}_{T,k}^{\mathrm{virt}} = (N_T - h_T)\ell(x^*) - \mathbf{1}_{\mathcal{B}_{k,T}} R_{k,T}^{\mathrm{virt}},$$

while on $\mathcal{B}_{k,T}^c$ the virtual regret is zero. Adding the term

$$(N_T - h_T)\ell(x^*)\Pr(\mathcal{B}_{k,T}^c)$$

restores the benchmark on the complement of the good event. Thus (EC.18) is an equality.

Using (EC.15), (EC.13), (EC.16), and (EC.17), we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}_{T,k}^{\mathrm{virt}}] = \mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}} N_T \lambda_J^{*\top}(\bar{c}_J - (b_T)_J)\big] \\ + T_{k,T}^{\mathrm{end}} + \frac{N_T}{L_T}\operatorname{tr}(Q_{\mathrm{dyn}} V_k^0) + \mathfrak{r}_{k,T}^{(2)} + \mathfrak{r}_{k,T}^{\mathrm{exit}} + o(\sqrt{T}),\end{aligned} \tag{EC.19}$$

where $T_{k,T}^{\mathrm{end}} = O(h_T)$ is the endpoint correction from (EC.16), and

$$|\mathfrak{r}_{k,T}^{\mathrm{exit}}| \le C N_T^2 \eta_T^2 e^{-c_{\mathrm{ex}} h_T} = o(1)$$

is the explicit clipped–unstopped exit charge.

The capacity term is common in law. Since b_T is determined by the common exploration path,

$$\mathbb{E}[b_T]$$

is the same under both policies. Moreover,

$$\|b_T - \bar{c}\|_\infty \le \overline{\Delta}_T = C_{\mathrm{load}} \frac{L_T}{T},$$

so

$$N_T \left|\lambda_J^{*\top}(\bar{c}_J - (b_T)_J)\right| \le C N_T \overline{\Delta}_T \le C L_T.$$

Therefore

$$\begin{aligned}\left|\mathbb{E}\big[\mathbf{1}_{\mathcal{B}_{k,T}} N_T \lambda_J^{*\top}(\bar{c}_J - (b_T)_J)\big] - N_T \lambda_J^{*\top}\{\bar{c}_J - \mathbb{E}[(b_T)_J]\}\right| \\ \le C L_T \Pr(\mathcal{B}_{k,T}^c) = o(\sqrt{T}).\end{aligned} \tag{EC.20}$$

The complement charge in (EC.18) satisfies

$$(N_T - h_T)\ell(x^*)\Pr(\mathcal{B}_{k,T}^c) = o(\sqrt{T}).$$

The virtual-to-truncated-virtual charge and the actual-to-truncated-virtual charge satisfy, by Lemma 19,

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{B}_{k,T}}\left|R^{\mathrm{virt}}_{k,T}-R^{\mathrm{tr,virt}}_{k,T}\right|\right]+\mathbb{E}\left[\left|R^{\mathrm{tr,virt}}_{k,T}-R^{\mathrm{act}}_{k,T}\right|\right]=o(\sqrt{T}).$$

Collecting the terms in (EC.18) gives, for each $k\in\{\mathrm{raw},\mathrm{cv}\}$,

$$\mathrm{Reg}^{\mathrm{fl}}_T(\pi^k_T)=C_{\mathrm{exp},T}+N_T\lambda^{*\top}_J\{\bar{c}_J-\mathbb{E}[(b_T)_J]\}+T_{k,T}+\frac{N_T}{L_T}\,\mathrm{tr}(Q_{\mathrm{dyn}}V^0_k)+o(\sqrt{T}).$$

The exploration term and expected capacity term are common because the policies share the exploration design, exploration requests and capacity schedule. The terminal terms satisfy

$$|T_{\mathrm{raw},T}-T_{\mathrm{cv},T}|\le Ch_T=o(\sqrt{T}).$$

Subtracting the two expansions yields

$$\mathrm{Reg}^{\mathrm{fl}}_T(\pi^{\mathrm{raw}}_T)-\mathrm{Reg}^{\mathrm{fl}}_T(\pi^{\mathrm{cv}}_T)=\frac{N_T}{L_T}\,\mathrm{tr}\left[Q_{\mathrm{dyn}}\left(V^0_{\mathrm{raw}}-V^0_{\mathrm{cv}}\right)\right]+o(\sqrt{T}).$$

By definition,

$$\gamma_Q=\mathrm{tr}\left[Q_{\mathrm{dyn}}\left(V^0_{\mathrm{raw}}-V^0_{\mathrm{cv}}\right)\right]\ge\underline{\gamma}_Q>0.$$

Hence

$$\mathrm{Reg}^{\mathrm{fl}}_T(\pi^{\mathrm{raw}}_T)-\mathrm{Reg}^{\mathrm{fl}}_T(\pi^{\mathrm{cv}}_T)=\frac{N_T}{L_T}\gamma_Q+o(\sqrt{T})>0$$

for all sufficiently large T.

A sufficient condition for strict positivity is $\Sigma_d\succ 0$, $\rho_\star>0$, and $Q_{\mathrm{dyn}}\neq 0$, together with

$$\Sigma_{\mathrm{eff}}\preceq(1-\rho^2_\star)\Sigma_d.$$

Then $\Sigma_d-\Sigma_{\mathrm{eff}}\succeq\rho^2_\star\Sigma_d\succ 0$, so $V^0_{\mathrm{raw}}-V^0_{\mathrm{cv}}\succ 0$, and since $Q_{\mathrm{dyn}}\succeq 0$ and $Q_{\mathrm{dyn}}\neq 0$, $\gamma_Q=\mathrm{tr}[Q_{\mathrm{dyn}}(V^0_{\mathrm{raw}}-V^0_{\mathrm{cv}})]>0$. □

EC.2 Paired-Signal Generation and Empirical Estimators

Samples and signal laws. Write $\mu_{ji}=f_i(p_j)$ for product i at design price j. The five historical prices are $p_0=L+0.1(U-L)$ and $p_j=p_0+0.8(U_j-L_j)e_j$. All uniforms below are independent across observations, products, prices and sample banks, except for the shared innovation in each request–score pair. Paired policies use the same in-season uniforms, each at its own prices. History is independent of in-season noise. Online requests are $D_{ti}=f_i(p_t)+a_i(p_t)U_{ti}$, where $a_i(p)=\min\{\sqrt{3}\sigma,0.95f_i(p)\}$ and $U_{ti}\sim\mathrm{Unif}[-1,1]$.

The supported-precision and guarded-deployment studies use $\sigma=0.25$, $a=\sqrt{3}\sigma$, and $D_{j\ell i}=\mu_{ji}+aU_{j\ell i}$, $\quad S_{j\ell i}=s_i\{\mu_{ji}+1+a[\rho_{ji}U_{j\ell i}+\sqrt{1-\rho^2_{ji}}V_{j\ell i}]\}$. Centers are $S^c=s_i\{\mu_{ji}+1+b+\kappa a[\rho_{ji}U^c+\sqrt{1-\rho^2_{ji}}V^c]\}$. Defaults are $s_i=\kappa=1,b=0$. The stresses use $b=0.1$, $\kappa\in\{0.5,1.5\}$, $s=(0.5,1,2,3)$, product correlations $(0.9,0.5,0,-0.9)$, or price correlations $(0.9,0.5,0,-0.5,-0.9)$. Scalar correlations are $0,0.25,0.5,0.7,0.9,-0.9$. Deterministic scores suppress both noise terms. These corrections use historical scores only. Each price has $M=20$ labels, $q=1$ online request and $C=20,200$ or 2000 centers. The guard reserves 200 of the latter bank.

The earlier signal-value, symmetric-fit, horizon and cross-price studies (Section EC.2) instead use $D_{j\ell i}=\mu_{ji}+d+a_i(p_j)U_{j\ell i}$ and $S_{j\ell i}=a_i(p_j)[\rho U_{j\ell i}+\sqrt{1-\rho^2}V_{j\ell i}]$. Their centers are $S^c_{j\ell i}=b+a_i(p_j)W_{j\ell i}$.

Here $d = b = 0$ normally. Separate stresses change only historical label means by $d = \pm 0.1$, or only centers by $b = \pm 0.1$.

The triangular-center stress replaces W by $(W_1 + W_2)/\sqrt{2}$. Noise levels, coupling and sample counts are given with each study in Section EC.4 or in the replication materials. Score means are zero and hidden from the policies. Centers match the mean and variance of the paired score but generally not its full marginal. Linear correction needs only matching means. Nonlinear centering needs matching transformed means, which full marginal matching ensures. The midrange variance formula also needs uniform centers. At $\rho = 0.9$ the paired score is not uniform, so the midrange procedure requires a separate channel of uniform centers with matched means.

Linear correction and precision weights. Let $k = M + q$, $X = [\mathbf{1}, P]$, $K = (X^\top X)^{-1} X^\top$ and $w_j = \|K_{:j}\|^2$. Five interlaced folds use indices ℓ mod 5. Each coefficient and screen exclude the entire held-out fold at every price. Outside-fold sample moments v_D, v_S, c use divisor $M - m_f - 1$. Blend c, v_S with their within-product price averages with weight ω. Denote these by $\widetilde{c}, \widetilde{v}_S$. With supplied allowance b_0, $g = \mathrm{clip}_{[-2,2]} \frac{\widetilde{c}}{(1+M/C)\widetilde{v}_S + M b_0^2}$, $\quad Q(g) = \frac{k v_D - 2Mgc + Mg^2 v_S}{k^2} + \frac{M^2 g^2}{k^2}(v_S/C + b_0^2)$. A fold passes when $r_f = \sum_{ji} w_j Q(g) / \sum_{ji} w_j v_D / k \le 1 - \tau$. For a failing fold, we set the coefficients to zero and keep all raw labels. A coefficient denominator $\le 10^{-12}$ gives a zero coefficient, and an aggregate raw risk $\le 10^{-12}$ fails the fold. Supported precision and signal-value use $(\omega, \tau) = (0, 0.3)$. Symmetric fitting and horizon budgets use $(0.5, 0.2)$: $\widetilde{c}_{ji} = 0.5 c_{ji} + 0.5 J^{-1} \sum_h c_{hi}$, likewise for v_S. Their risk screen still uses w_j. The pricewise extension fully pools moments (weight one), using its separate formula and 10% screen retained in the replication source for the earlier study (Section EC.2). Its centers use the sample midrange. Corrected and matched raw response rows use sample means. Other linear fits use sample means. History-only studies purchase centers only if a coefficient is nonzero.

Let H be the corrected historical mean, $\bar{g} = \sum_f (m_f/M) g_f$, and $\widehat{v}_D, \widehat{v}_S$ the full-history sample variances. For $q > 0$, $\widehat{R}_H = \widehat{\mathrm{Var}}(\widetilde{D})/M + \bar{g}^2(\widehat{v}_S/C + b_0^2)$, $\quad a = 1 - \eta + \eta\, \mathrm{clip}_{[a_{\min},1]} \frac{k\widehat{R}_H}{\widehat{v}_D + q\widehat{R}_H}$, $\quad \Omega = qa/k$. Here $\eta, a_{\min} \in [0, 1]$. The selected empirical values are $(1, 0.25)$. Each online fit row is $H + a(D^{\mathrm{on}} - H)$. Pooling all $M + q$ rows gives $(1 - \Omega)H + \Omega \bar{D}^{\mathrm{on}}$. Zero coefficients at a coordinate force $a = 1$ there. A denominator $\le 10^{-12}$ also gives $a = 1$. Unless every linear fold passes with $r_f \le 0.35$ and every coordinate has $M\widehat{R}_H/\widehat{v}_D \le 0.5$, all factors revert to one and only the historical correction is used. The later protocols use thresholds $0.75, 0.85$ if $q > 0$. Their history-only branch has $q = 0$, uses H alone and bypasses precision weighting. Nonfinite support ratios or raw variances $\le 10^{-12}$ fail the precision support gate.

Guarded nonlinear increment. Outside each held-out fold, standardize the score by pricewise training means and a productwise scale, fit price-centered requests on $(z, \mathrm{clip}_{[-1,1]} z)$ with ridge $0.02 \operatorname{diag} G$, and clip slopes at twice the product's root mean request variance. Here G is the Gram matrix of the product's two features in the training folds. Eligibility requires $C \ge 1000$, $b_0 = 0$, a passing linear risk ratio ≤ 0.3, positive variances $> 10^{-12}$, every coordinate's adjusted risk ratio ≤ 0.45, and pilot/training score-variance ratio in $[0.6, 1.6]$. Reserve 200 independent centers and use this untouched tail to center only the nonlinear increments that pass. The linear base uses all centers. If the increment fails, the policy is exactly the supported-precision policy. The precision weights keep both center blocks and use the same support gates. The replication materials give the full regression, weighting and tail-centering formulas. Folds overlap and share centers and learned weights, so they are dependent.

Operational records. Table EC.1 separates the data used for calibration from the information that the simulator imposes. Allowances for instability and bias are fixed in advance. The coefficient is trained

without the pairs it corrects, and the nonlinear calibration sets aside a pilot sample for the centers. In-season requests determine stock and revenue, including after fallback, and (19) charges the incremental costs.

Table EC.1: Information inputs in the retail illustration.

Input	Availability and timing	Cost convention
Transactions	Observed sales/prices calibrate four curves, without identifying uncensored demand.	Historical calibration
Historical requests	Imposed potential requests at five covered prices before deployment; stable campaign means assumed.	One unit per vector
Paired scores	Imposed contemporaneous pairs; conditional moments require joint observations.	Extra score fees
Independent centers	Imposed independent pricewise samples; uniform centers match score means/variances, generally not full marginals.	0.25 units per vector
Campaign state	Capacity, incidence and innovations imposed; full unserved requests assumed observable.	Revenue/stock through fallback
Screening and fitting	Use available pairs before center purchases. Implementation requires access to these pairs.	Extra scores/computation

Earlier studies. The network results of Section 5 replace the scalar informed policies of the conference version. The replication materials contain the earlier signal-value, robustness, symmetric-fitting, horizon-budget, cross-price transfer and pricewise-center studies, each with its original policies, controls, seeds and test family. The current pooled comparison, which includes the incumbent, is in Section EC.6.

EC.3 Residual-Control Experiment Design

The first benchmark has four products sharing one resource, prices in $[1,4]^4$, and mean demand $f(\boldsymbol{p}) = 6\mathbf{1} - \boldsymbol{p}$. Each unit consumes one resource unit. Capacity is 85% of unconstrained optimal load, giving binding fluid demand 2.55 per product. Requests add independent centered uniform shocks with standard deviation 0.25, stay nonnegative, and are served proportionally as in Section 2.

With known demand, the fallback is the fluid price and $h_T = \lceil 40(1+\log T)\rceil$, which meets the terminal condition of Section 3.

For unknown demand, Algorithm 2 uses five interior-simplex prices, $L_T = 5\lceil 2\sqrt{T}/5\rceil$, $h_T = 2L_T$, one least-squares estimate and one fitted fluid solve. The projection keeps intercepts bounded, slopes symmetric negative definite and demand nonnegative on the box. The true slope $-I$ is in this class. The estimate is then frozen, and cumulative request residuals update the target.

We use horizons $T \in \{800, 1600, 3200, 6400, 12800\}$ and 200 independent replications at each horizon. Figure 2 uses 1.96 standard errors. Fitted-rate intervals use 1,000 bootstrap resamples.

EC.3.1 Estimation and implementation diagnostics

Figure 2(a) gives mean regret under known demand 19.20 ± 4.32 at $T = 800$ and 35.71 ± 18.56 at $T = 12{,}800$. The coefficient of a fitted affine function of $\log T$ is 6.32, with bootstrap interval $[0.61, 12.06]$. The standard deviations of regret, 31.16–133.94, exceed the means. Paths with negative regret are kept. Section EC.3.2 reduces this noise with a mean-zero adjustment chosen on an independent pilot.

For unknown demand, the fitted log–log slope is 0.495 with bootstrap interval $[0.488, 0.501]$. Regret/$\sqrt{T}$ is 44.17–46.17 and the parameter-error slope is -0.244. All 2,000 paths include fallback and service losses, and none rations requests before the last periods. Targets leave the demand image on no known-demand path and on 1%–2.5% of unknown-demand paths. Eligible targets are implemented to within 3×10^{-15}.

EC.3.2 Full-Information Monte Carlo Precision

On the full-information paths, let $Y_T = \sum_t p_t^\top \widehat{D}_t$, $Z_T = V_T^{\text{Fluid}} - Y_T$ and $M_T = \sum_t p_t^\top \xi_t$. Predictability and conditional centering give $\mathbb{E}M_T = 0$, including terminal, exit and stockout periods. We have

$$Z_T + M_T = V_T^{\text{Fluid}} - \sum_t p_t^\top f(p_t) + \sum_t p_t^\top (D_t - \widehat{D}_t). \tag{EC.21}$$

Adding a multiple of M_T changes the estimator's variance but not expected regret. The regret also accounts for the lost sales. A pilot with 32 paths per horizon selects one coefficient from $\{-1, -0.5, 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5\}$ by the mean variance ratio across horizons (floor 10^{-12}; ties go to the smaller absolute value, then the smaller coefficient). The chosen value 0.5 is then applied to 2,048 fresh paths per horizon, with seeds disjoint from the pilot and from earlier experiments. All 4,255 negative raw

Table EC.2: All five precision-validation horizons, 2,048 fresh paths each. Means [pointwise normal 95% intervals]; variance ratios compare pilot-adjusted with raw regret.

T	Raw regret	Pilot-adjusted regret	Var. ratio	SE reduction
800	14.79 [13.52,16.06]	15.10 [14.57,15.62]	.1725	58.5%
1600	16.35 [14.47,18.22]	16.93 [16.26,17.59]	.1248	64.7%
3200	19.31 [16.46,22.16]	18.52 [17.58,19.46]	.1092	66.9%
6400	18.19 [14.02,22.36]	19.11 [17.79,20.43]	.1001	68.4%
12800	17.32 [11.40,23.23]	19.28 [17.42,21.14]	.0989	68.6%

losses, 1,964 negative adjusted losses and 5,082 paths with terminal rationing are kept. No validation path exits early or rations service before the last periods. The coefficients on $\log T$ are 0.996 $[-0.870, 2.862]$ raw and 1.521 $[0.924, 2.119]$ adjusted, with pointwise normal intervals. The adjustment reuses the same paths and improves precision.

EC.4 Calibration and Comparisons for the Retained Studies

Each study has its own fixed policies, seeds and comparison family. Section EC.2 defines the signal laws, precision weights and guarded fit. Four summaries were regenerated with the same seeds and configurations. The four-product benchmark is calibrated to the UCI Online Retail data with 85% shared fulfillment capacity.

Costs and simple pricing rules. The tables and Figure EC.1 report all comparisons and acquisition costs.

Estimator scope. The supported-precision screen selects the complete-label correction of Theorem 6, whose guarantee holds under its calibration, stable-history, radius and projection conditions. The guarded nonlinear correction is an empirical method, and its support thresholds and price schedule were fixed before evaluation. Learning the coefficient and centering on an untouched sample are separate steps, and both determine the value of a correlated signal for pricing. All fitted controls use the same labels. Section EC.4.2 reports all comparisons of the fixed linear estimators.

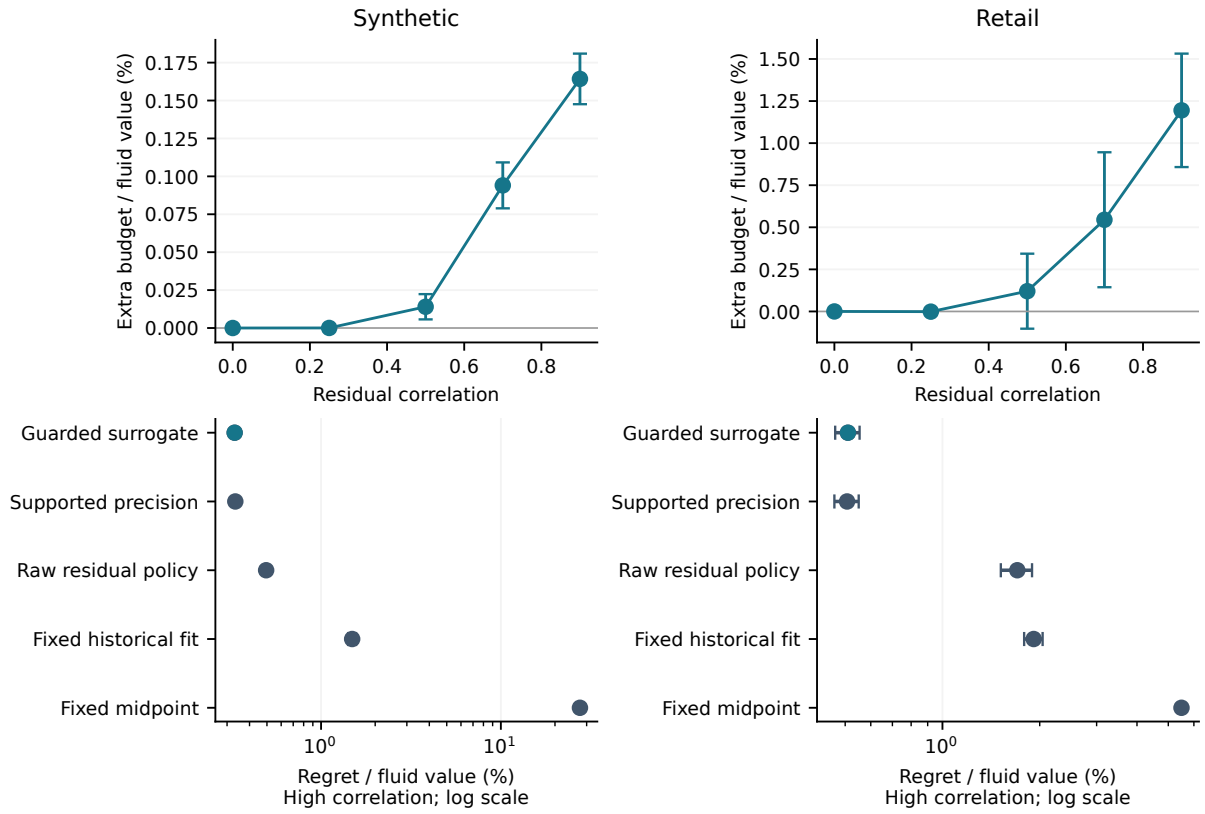


Figure EC.1: Deployment value in the same-seed replay. Top: extra budget for data and computation, relative to raw pricing, as a percentage of initial fluid value, with simultaneous intervals. Bottom: regret under high correlation with pointwise 1.96 standard errors, horizontal axis in log scale. Retail requests and auxiliary signals are simulated.

EC.4.1 Retail calibration and information supplied to the policy

UCI Online Retail contains 541,909 lines (Chen 2015). Removing cancellations, nonpositive quantities/prices, prices at least £50 and codes without five-digit prefixes leaves 527,498. We aggregate by product and week, using total quantity and the unweighted mean line price. Weeks without sales are dropped rather than treated as zero requests. Candidates require at least 20 active weeks, price CV at least 0.05, more than 500 units, positive fitted intercept, negative slope and descriptive elasticity in $[-4.5, -1]$. Before evaluating any policy, we select the four highest-volume products among the 810 candidates, breaking ties by code.

Table EC.3: Retail scenario calibration. Prices are active-week means in GBP. The fitted response describes observed sales, not causal demand.

SKU	Product	a_i	b_i	Observed price
22197	Popcorn holder	3645.8	-2450.6	[0.881, 1.352]
84077	Toy gliders	3966.7	-9180.8	[0.263, 0.421]
85099B	Jumbo storage bag	3070.7	-861.2	[1.945, 3.144]
23084	Rabbit night light	4405.4	-1125.8	[1.932, 4.130]

Common units are $q_u = 440.542$ items and $p_u = 0.421103$ GBP, giving $\alpha_i = a_i/q_u$, $B_{ii} = b_i p_u/q_u$ and zero off-diagonals. The price ranges $[0.3a_i/(-b_i), 0.8a_i/(-b_i)]$ in GBP, divided by p_u, extend beyond the observed prices. Extrapolation, capacity, noise in the requests and paired signals are simulated, because the transactions record neither uncensored requests nor signals. History and the season share the same response law. Fitted policies see only the data. Corrections use the stated bias allowance.

EC.4.2 Supported-precision linear calibration

Table EC.4: Supported-precision comparison across all 30 conditions.

Criterion	Raw	Ordinary cross-fit
Strict improvement (separate families)	19/30	9/30
2% noninferiority, first stage	29/30	29/30
After prescribed raw follow-up	30/30	Not repeated

First-stage paired intervals, all 15 conditions in each scenario, and the separate follow-up remain in replication data. The families are separate. Retail with 200 centers remains unresolved against cross-fit.

EC.4.3 Deployment value against practical controls

Empirical selector comparison. A feedback variant initialized from history was tuned on 16 cells with 16 seeds and evaluated on all 36 cells with 64 fresh paired seeds (two networks, three horizons or sample counts, two noise levels). Against the matched historical inverse with the same recovery rule, the equal-cell revenue and net difference is -0.026 $[-0.046, -0.006]$ pointwise, and no individual contrast is significant in the 1,152-component family. With 16 or 256 historical samples the variant reproduces the historical inverse exactly. Most of its gain over the original rule comes from keeping the history when the gate fails and holding the last price under the no-information algorithm. Against no history, once labels are charged, it has 16 gains, 19 losses and one unresolved case. The comparison leaves this empirical variant and the historical inverse unranked.

Table EC.5: Guarded deployment: all 136 contrasts.

Control	Gain	Loss	Unresolved
Raw residual	23/34	0/34	11/34
Supported precision	4/34	0/34	30/34
Historical fixed	29/34	0/34	5/34
Midpoint	34/34	0/34	0/34

One simultaneous family covers 34 cells and four controls. A gain or loss requires a positive lower or negative upper endpoint. Exact zero differences are unresolved under this criterion. All casewise means and intervals are in replication data.

EC.5 Network Experiments

Figure 3 uses one set of cases for each network. Horizons $400, 1600, 6400, 25600$ are crossed with anchor error 0 and ± 0.04 along a relative-product direction in the null space of A. The grid also includes an exact anchor at $T = 102{,}400$ and two exact but misaligned anchors at $T = 6{,}400$ and $25{,}600$. Each cell has 1,000 paired replications. The curvature bounds are $(\mu, M) = (0.75, 1.25)$ and $(0.8, 1.1)$, with demand radius $r = 0.25$. Over the five horizons with an exactly aligned anchor, direct-learning regret is between 2.064 and 2.520 for two products and between 3.761 and 4.638 for four products.

The 24 cells combine two networks, horizons $1600, 6400, 25600$, aligned and exactly shifted anchors, and two profiles. Each uses 1,000 paired seeds, and all cells form one family of 96 contrasts. All policies except the oracle share the class center and radius. The broader profile uses matrix radius 0.06 and noise amplitude 0.8 (prices $[2, 6.25]$, demand-ball radius 0.15). The policy checks the nominal KKT conditions and the class geometry and then learns the implementing prices. Only the oracle knows the true matrix. Figure 4 reports the four longest shifted cases: narrow-class intervals are $[1234.50, 1235.32]$ and $[1079.40, 1080.18]$, and residuals are $0.008/0.024$ versus 0.435–0.521 for the norm-only plan. Corrected feedback has higher mean revenue than nominal controls in all 24 cells and than original targets in all 12 shifted cells. The oracle, which knows the slope matrix, has the highest mean revenue. All active requests are served, and the accounting error is at most 2.04×10^{-8}. A fixed residual leaves a bias term linear in T.

EC.6 Comparison with 58 Controls: Design and Results

Case grid and seeds. The four-product grid crosses horizons $T \in \{400, 1600\}$, historical sample counts $N \in \{2, 4\}$, shared-fulfillment and product-inventory/shared-fulfillment profiles, cross-price coupling $-0.2, 0$ and 0.2, and noise $\sigma \in \{0.5, 0.65\}$. Uniform-center correlations $\rho = 0, 0.5, 0.9, 0.99$ and a variance-matched triangular-center stress at $\rho = 0.9$ give 240 cases: 96 primary and 48 each null-signal, moderate-signal and center-distribution stress. Each case uses the four regimes `original_shared`, `opposite_product_signs`, `alternating_price_signs`, and `mixed_strengths`. The comparison uses seeds $s_0 + 800000000000 + k$, $k = 0, 1, 2, 3$, where s_0 is the case's first seed. This gives 3,840 units with seeds disjoint from development and from the earlier comparison below. The guarded policy, the least-squares estimate with the same guard, the targets, both tests and the analysis were fixed before these seeds were run, and every control was run on them. In every unit, a second run of the candidate and the incumbent reproduced the first exactly. The one unit in which the guard inherits a failed candidate initialization is kept in the analysis.

Estimand. Each control comparison averages its 16 paired regime–seed observations within each case, then averages the 240 case means equally. The standard error is their sample standard deviation divided by $\sqrt{240}$. For each target, a positive lower endpoint $\bar{x} - 3.7001\,\mathrm{SE}$ passes the Bonferroni-adjusted normal

approximation for $58 \times 4 = 232$ components. Writing $\Delta R = R_{\text{guard}} - R_{\text{control}}$ and $\Delta N = N_{\text{guard}} - N_{\text{control}}$ for revenue and profit-after-fees differences, the targets are positive pooled means of ΔR, ΔN, $\Delta R - 0.01 R_{\text{control}}$ and $\Delta R - 0.05\,\text{Reg}^{\text{fl}}_{\text{control}}$. Percentages divide the pooled revenue difference by pooled control revenue or regret. Regret reduction is the same revenue contrast, measured against the common fluid benchmark. For the comparison with the least-squares estimate, this estimand is used for revenue and profit-after-fees value, with 2.2414 in place of 3.7001 for two targets.

Results by coupling. At cross-price coupling -0.2 and 0, the differences in revenue against the incumbent are +11.4% and +1.6%. At coupling +0.2, the incumbent prices are close to the best: on development seeds the candidate has higher revenue than the incumbent in 15 of 640 units. The guard switches to the incumbent prices in 1,004 of 1,279 comparisons, after posting the candidate for the first 32 periods. Its revenue is 7.1% above that of the unguarded candidate and within 1% of the incumbent's (−0.8%). At coupling -0.2, the guard switches in 133 of 1,280 comparisons, and its revenue is 0.8% below that of the unguarded candidate, which is the cost of the two blocks. At coupling 0, it switches in 655 of 1,280 comparisons, and its revenue is 0.6% above that of the unguarded candidate.

Control names. The factored names below expand to the 58 control identifiers. Controls with identical revenue keep their own data fees. The prefixes `balanced::` and `public::equal_gate::` each take these 21 suffixes: `raw_mean`, `raw_midrange`, `joint_independent`, `joint_all_coordinates`, `joint_signed_shared`, `previous_shared_joint`, `structure_mixture`, `linear_local_covariance`, `linear_local_likelihood`, `linear_mixture_likelihood`, `linear_signed_magnitude_likelihood`, `linear_pooled_covariance`, `linear_pooled_likelihood`, `linear_outside_coordinate_covariance`, `preceding_purchased_covariance`, `preceding_purchased_likelihood`, `preceding_coordinate_covariance`, `preceding_coordinate_likelihood`, `preceding_joint_coordinates`, `preceding_joint_local`, `preceding_pricewise_incumbent`.
The prefix `preceding::candidate__` takes `coordinate_covariance`, `coordinate_likelihood`, `joint_coordinates`, `joint_local`, `joint_shared`, `linear_covariance`, and `linear_likelihood`, each followed by `::theorem4_linear_surrogate`.
The remaining `preceding::` identifiers are `original::` followed by `raw_matched_initialization` or `theorem4_linear_surrogate`; and `raw_mean__joint_local`, `raw_uniform_midrange__incumbent`, `raw_uniform_midrange__joint_local`, each followed by `::raw_matched_initialization`.
Finally, `public::mean_center_` takes `local_covariance`, `pooled_covariance`, `outside_covariance`, and `observed_likelihood`.

Revenue screen. The guard keeps the candidate's fitted controller and pays its full data fee. It posts the candidate's prices for $L = 32$ periods and the incumbent's upper prices for the next L periods. For each block, let $Y_t = p_t^\top D_t$, let $\bar{D}$ be the mean observed request, let c be the initial stock divided by the horizon, and let $s = \min\{1, \min_{j:(A\bar{D})_j > 0} c_j/(A\bar{D})_j\}$. The request-based revenue score of a block is $s\bar{Y}$, with approximate standard error $s\ \text{sd}(Y)/\sqrt{L}$. After both blocks, the guard switches permanently to the upper prices if their score exceeds the candidate's score by more than the root-sum-square of the two standard errors. Otherwise, it returns to the candidate's prices. This one-standard-error rule is separate from the threshold used in the pooled tests. The resource matrix and stock are known. The comparison uses only prices the guard actually posted and the requests it observed, and the candidate's state and the remaining capacity include every probe period.

Raw fit with the same guard. This policy uses a least-squares estimate from the same historical data, subject to the constraints of the candidate, together with the same controller, upper prices, probe length and threshold as the guarded policy. It uses historical request data alone and pays only for the requests it uses, on average 18 less than the guarded policy. The guarded policy has 1,222 more revenue (standard error 106), a pooled revenue gain of 2.77%, and 1,204 more profit after fees (standard error 105). Both targets pass. Against the pricewise incumbent, the least-squares estimate with the same guard has a gain of 0.74% (standardized statistic 2.17). By group, the difference in revenue is 1,464 (standard error 172) for primary, 981 (161) for null-signal, 1,087 (297) for moderate-signal and 1,117 (211) for center-distribution stress. At cross-price coupling −0.2, 0 and +0.2, the guarded policy has 4.9%, 1.8% and 1.8% more revenue than the least-squares estimate with the same guard.

Unguarded candidate. On the comparison seeds, for the unguarded candidate, the revenue, paid-net, revenue-1% and regret-5% targets pass for 56, 56, 49 and 56 of the 58 controls, respectively. Pooled revenue gains range from 1.11% to 3.41% (median 2.65%). With the revenue screen, all 232 components pass, including the two pricewise-incumbent aliases and the observed mean-center likelihood control (Section 7.3).

Earlier comparison. The earlier comparison uses seeds $s_0 + 700000000000 + k$, $k = 0, 1, 2, 3$. The unguarded candidate was compared first. That run was stopped early under a rule fixed independently of outcomes, and its analysis uses all 14,887 completed units. Case weights are equal, and all available units are averaged within each case. Revenue, paid-net, revenue-1% and regret-5% targets pass for 56, 56, 55 and 56 of the 58 controls, with pooled revenue gains from 1.08% to 3.78% (median 2.83%). The guard was then specified: its structure and its only tuning choice, the probe length, were set on development seeds 0–1 and checked on development seeds 2–3, and the guarded policy was run on the same seeds, reusing the control results. All 232 components pass, with pooled revenue gains from 3.6% to 6.3% (median 5.4%) and 3.6% against the incumbent (standardized statistic 8.81). By coupling, the differences in revenue against the incumbent are +11.3%, +1.75% and −0.7%. The comparison above runs the guarded policy and every control on the comparison seeds.