

Learning to Price in Feature-Based Pool Markets under Censored Feedback

Blinded for review

Abstract. We study dynamic pricing in a market with a fixed population of persistent customers who revisit intermittently, purchase at most once, and leave only after a transaction. Customer valuations of the product depend on customer features through an unknown parameter, but the seller observes an individual customer’s features only after purchase. Unsuccessful visits and rejected offers remain unobserved. As a result, the seller learns from censored transaction data generated by an endogenously depleted customer pool rather than from a sequence of independent, one-shot arrivals.

We consider markdown policies over a finite price ladder. When the valuation parameter is known, a valuation-segment representation reduces the population markdown problem to a finite-dimensional concave program. When the parameter is unknown, we derive a censored likelihood that captures the joint effects of latent valuations, intermittent attention, prior price exposure, and pool depletion. Building on this likelihood, we develop a markdown-preserving learn-then-optimize policy that screens inactive prices, collects transactions at the first active price, estimates the valuation parameter, and optimizes the remaining markdown schedule for the residual pool.

After m transactions, the resulting estimator attains a mean-square error $O(1/m)$. Strong concavity of the residual markdown schedule problem then converts this estimation error into an $O(n/m)$ revenue loss, while the data-collection phase incurs an $O(m)$ loss. Balancing these two terms yields $O(\sqrt{n})$ expected regret relative to both the population markdown benchmark and the hindsight-optimal markdown schedule for the realized customer pool. We further identify a free-learning regime in which population regret improves to $O(\log n)$, and we construct a local $\Omega(n^{1/3})$ lower bound near the boundary at which a high price changes from being skipped to being used.

Keywords: Dynamic pricing · Online learning · Censored feedback · Persistent customers · Markdown pricing

1 Introduction

Dynamic pricing with learning is commonly studied in a stream-based market: customers arrive sequentially, observe the current price, make a one-time purchase decision, and then leave. This framework has generated a broad literature on demand learning, exploration-exploitation tradeoffs, and regret minimization [4–7, 29, 30, 38]. More recent work incorporates customer covariates and studies how observable characteristics shape demand or willingness to pay [2, 8, 18, 25, 33]. A common feature of these models is that each pricing decision is followed by an observable response from a newly arriving customer.

Many online and retail markets, however, do not follow this interaction pattern. A seller may instead face a fixed population of potential buyers who remain in the market, check prices intermittently, and exit only after purchasing. Examples include ticket sales, durable goods, subscription upgrades, and markdown markets for apparel and electronics. A customer who does not purchase today may return and buy later. Conversely, a customer who eventually purchases at a low price may have been willing to pay more but may simply have missed the earlier offer.

This persistence changes both the pricing problem and the learning problem. Early transactions remove customers from the market and thereby alter the composition of future demand. A high-valuation customer may purchase and leave soon after the market opens, whereas another customer with the same valuation may remain in the pool only because she has not yet checked the price. Prices therefore affect not only current revenue but also the latent population from which all subsequent observations are generated. Pool-based pricing models formalize these depletion and price-order effects and show that they can lead to pricing structures that do not arise in standard sequential arrival models [21].

We introduce feature-based valuations and statistical learning into this persistent-market setting. Each customer has a fixed valuation consisting of a feature-based component and an idiosyncratic shock. The seller knows the population distribution of customer features but observes an individual feature vector only after the corresponding customer purchases. The unknown parameter linking features to valuations must therefore be learned from a selected sample of completed transactions.

Features play a different role here than in standard personalized-pricing models. The seller posts a common price, and customer features are unavailable when the pricing decision is made. Instead, the feature distribution among observed buyers reveals how the collecting price selects customers from the underlying population. This selected sample identifies the valuation parameter and allows the seller to infer latent demand at lower prices that have not yet been posted, or have been posted only briefly. Aggregate transaction counts at a single price generally reveal only the total mass of customers currently willing to buy. Buyer features provide the cross-sectional variation needed to recover how willingness to pay changes across prices.

The observation process also creates a form of censoring that is specific to persistent markets. A period without a transaction may mean that customers did not check the price, that they checked but found the price too high, or that the most likely buyers have already purchased and exited. Likewise, the transaction price does not reveal the buyer’s valuation segment: a customer who would have accepted a higher price may purchase at a lower price simply because she did not observe the earlier offer. Any valid learning procedure must therefore account jointly for latent valuations, intermittent attention, prior price exposure, and endogenous pool depletion.

We study markdown policies over a finite price ladder. Under a weakly decreasing price path, each customer becomes willing to purchase at a well-defined point in the ladder and remains willing to purchase at all subsequent prices. This observation partitions the population into valuation segments and yields tractable representations of expected revenue, the censored likelihood, and the composition of the residual customer pool.

Our contributions are fourfold. First, when the valuation parameter is known, we derive a segment-based revenue representation and show that the population markdown problem is a finite-dimensional concave program over price durations. We also compare this deterministic benchmark

with the hindsight-optimal markdown schedule for the realized valuation pool. The expected optimal values of the two problems differ by at most $O(\sqrt{n})$, so replacing the realized-pool benchmark with the population benchmark does not weaken the order of the regret guarantee.

Second, when the valuation parameter is unknown, we construct a censored likelihood from purchase times, transaction prices, buyer features, and aggregate nonpurchase information. The likelihood accounts for the fact that different valuation segments accumulate different amounts of active exposure under a markdown path. It also incorporates the survival probability of customers who remain in the pool, which is essential because the observed buyer sample is selected by both willingness to pay and prior price exposure.

Third, we propose a markdown-preserving learn-then-optimize policy. The policy screens prices from high to low, collects transactions at the first active price, estimates the valuation parameter, and then optimizes the remaining markdown schedule for the residual pool. After m transactions, the estimator achieves mean-square error $O(1/m)$. Strong concavity of the residual markdown schedule problem converts this estimation error into an $O(n/m)$ revenue loss, while the data-collection phase incurs an $O(m)$ loss. Balancing the two terms yields $O(\sqrt{n})$ expected regret relative to both the population markdown benchmark and the hindsight-optimal markdown schedule for the realized pool.

Fourth, we refine the worst-case guarantee in two directions. When the optimal markdown schedule itself spends a nonvanishing amount of time at the collecting price, the policy can learn along an economically optimal path, and population regret improves to $O(\log n)$. At the other extreme, we construct a two-price parametric submodel and establish a local $\Omega(n^{1/3})$ lower bound near the boundary at which assigning positive duration to the high price first becomes optimal.

The remainder of the paper is organized as follows. [Section 2](#) relates our model to the literature on dynamic pricing, censored learning, and persistent customers. [Section 3](#) introduces the market and the valuation-segment representation. [Section 4](#) develops the population and hindsight markdown benchmarks. [Section 5](#) derives the censored likelihood and presents the learning policy. [Section 6](#) establishes the statistical and optimization results underlying the regret analysis and develops the free-learning and lower-bound results.

2 Related Literature

Our work connects four strands of research: dynamic pricing with learning, feature-based pricing, learning from censored observations, and pricing with persistent customers. The main distinction is the interaction structure. We repeatedly price a fixed population, observe customer-level features only after purchase, and learn from data whose distribution changes endogenously as customers transact and leave the market.

Dynamic and feature-based pricing. A large stream of literature studies dynamic pricing when demand is initially unknown and must be learned while prices are chosen [4, 6, 7, 29, 30, 38]; see [5] for a survey. These models typically assume a sequence of fresh arrivals, so each price generates an observable response from a new customer. Feature-based and contextual-pricing models enrich this framework by allowing demand or valuations to depend on observed covariates [2, 18, 25, 33]. Related work also considers high-dimensional and multi-product environments [26], separable contextual demand [8], fairness constraints [15], and offline data [10].

Most of this literature retains the contextual-bandit interaction: a customer arrives with observed features, receives a price, and immediately generates an observable purchase decision. Our setting differs in each of these respects. The seller does not observe the features of customers currently in the pool, rejected offers are unobserved, and the same customer may face several prices before purchasing. Moreover, earlier transactions change the composition of the population that remains available for future sales. The observed buyer sample is therefore selected jointly by valuations,

intermittent attention, the full price history, and prior departures. The resulting statistical problem is not a standard contextual-pricing problem with missing covariates.

Censored demand and survival analysis. The likelihood we use is related to work on learning from censored demand. In lost-sales inventory systems, the decision maker observes sales but not demand beyond the available inventory. Nonparametric learning under this form of censoring is studied by [14, 24]. Offline pricing with censored observations is considered by [9], while [39] studies dynamic pricing under adversarial censoring.

The censoring mechanism in our model is different. It arises from unobserved customer attention and from the fact that a customer’s purchase hazard becomes active only after the posted price falls below her latent valuation. The absence of a transaction does not reveal whether a customer checked the price, and a completed transaction does not reveal the first price that the customer would have accepted. Our transaction-time formulation is also related to survival analysis with covariates [16, 20]. In a standard proportional-hazards model, however, the risk process is typically observed or specified directly. Here, the relevant exposure period is determined jointly by the posted price path and an unobserved valuation segment. Our likelihood therefore integrates over latent segments and explicitly accounts for the survival probability of customers who remain unpurchased.

Persistent customers and intertemporal pricing. A separate literature studies pricing when customers remain in the market and how the timing of prices affects purchase behavior. Classical models of durable-goods monopoly and intertemporal price discrimination examine strategic waiting by forward-looking consumers [3, 17, 19, 35–37]. Related work in revenue management studies strategic purchasing, quick response, inventory display, and preannounced markdowns [1, 11, 12, 22, 31, 40].

Customers in our model do not strategically delay their purchases. Instead, persistence arises because customers observe prices only intermittently. A customer may therefore purchase late even though she would have accepted an earlier price, simply because she did not see the earlier offer. We build most directly on the pool-based framework of [21], which studies pricing a fixed population of unit-demand customers who monitor prices over time and exit after purchasing. We extend that framework by introducing feature-linked valuations, transaction-only feedback, censored parameter estimation, and regret analysis. The central learning difficulty is that the seller must infer demand from a pool whose composition is itself altered by the prices used to collect information.

Markdown pricing and limited price changes. The restriction to weakly decreasing price paths connects our model to the literature on clearance and markdown optimization [13, 23, 34] and to learning problems with monotonicity constraints or a limited number of price changes [27, 28, 32]. In much of the literature, monotonicity is imposed as an operational constraint or motivated by strategic customer behavior. In a persistent customer pool, markdown structure also serves an economic screening role: high prices initially target customers with high willingness to pay, and completed transactions remove those customers from the residual market. Our learning policy preserves this structure throughout. In particular, it never raises the price to revisit an earlier experiment.

3 Market Model and Valuation Segments

3.1 Market and observation structure

We consider a monopolist who sells a single product over a finite horizon $[0, T]$ to a fixed pool of customers $\mathcal{N} = \{1, \dots, n\}$. Each customer demands at most one unit and exits the market immediately after purchasing. Customer j is characterized by a feature vector $X_j \in \mathcal{X} \subset \mathbb{R}^d$. The feature vectors are drawn independently from a known distribution F . We assume the seller knows

the population distribution F but does not observe an individual customer's realized feature vector before that customer purchases.

Customer willingness to pay follows a feature-based valuation model with idiosyncratic noise. Specifically, customer j has a fixed valuation

$$V_j = X_j^\top \theta^* + \varepsilon_j, \quad (1)$$

where $\theta^* \in \Theta \subset \mathbb{R}^d$ is an unknown parameter. The idiosyncratic shocks ε_j are drawn independently from a known distribution G with density g , independently of the feature vectors, and are realized at the beginning of the horizon. At each time t , the seller posts a common market price $p(t)$ chosen from the finite price ladder

$$\mathcal{P} = \{p_1 > p_2 > \dots > p_k > 0\}. \quad (2)$$

Throughout this paper, we use $[k] = \{1, \dots, k\}$.

We assume customers do not continuously monitor the market price. Instead, each customer checks the posted price according to an independent Poisson process with a known rate $\lambda > 0$. Customer j purchases at a check time t if and only if $V_j \geq p(t)$. Upon a transaction, the seller observes the purchase time, the transaction price, and the buyer's feature vector. Checks that do not result in a purchase remain unobserved. Because the initial pool size is known, the seller also knows at every point in time how many customers remain unpurchased. This observation structure is central to the learning problem. A customer may remain in the pool either because she has not yet checked the price or because her valuation is below the prices she has observed. Moreover, once a customer purchases, she leaves the pool and no longer contributes to future demand. The data available to the seller are therefore both censored and shaped by the endogenous depletion of the customer pool.

3.2 Markdown policies and valuation segments

In this paper, we focus on markdown schedules over the price ladder $\mathcal{P}$. Such a schedule can be represented by switching times

$$0 = S_0 \leq S_1 \leq \dots \leq S_k = T,$$

where price p_i is posted on $(S_{i-1}, S_i]$. Let $\Delta_i = S_i - S_{i-1}$ denote the duration assigned to price p_i , and define the feasible duration set

$$\mathcal{D} = \left\{ \Delta \in \mathbb{R}_+^k : \sum_{i=1}^k \Delta_i = T \right\}. \quad (3)$$

Since zero durations are allowed, a markdown schedule may skip one or more prices.

Because customer valuations are fixed over the selling horizon and the posted price is weakly decreasing, each customer becomes willing to purchase at a unique point in the price ladder and remains willing to purchase at every lower price. This observation induces a natural partition of the valuation space. Let $p_0 = +\infty$ and define

$$\mathcal{V}_i = \{v : p_i \leq v < p_{i-1}\}, \quad i \in [k]. \quad (4)$$

A customer in segment i first becomes willing to buy when the price reaches p_i and remains willing to buy at all subsequent prices.

For a parameter θ and feature vector x , the probability that the customer's valuation lies in segment i is

$$\pi_i(x; \theta) = G(p_{i-1} - x^\top \theta) - G(p_i - x^\top \theta), \quad i \in [k], \quad (5)$$

$$\pi_{k+1}(x; \theta) = G(p_k - x^\top \theta). \quad (6)$$

The additional segment $k+1$ contains customers whose valuations fall below the lowest available price, and who therefore never purchase under any admissible markdown schedule. Integrating over the feature distribution gives the population mass of each segment:

$$\pi_i(\theta) = \int_{x \in \mathcal{X}} \pi_i(x; \theta) dF(x), \quad i = 1, \dots, k+1. \quad (7)$$

We also define

$$q_i(x; \theta) = 1 - G(p_i - x^\top \theta), \quad A_i(\theta) = \mathbb{E}[q_i(X; \theta)] = \sum_{\ell=1}^i \pi_\ell(\theta), \quad (8)$$

where $q_i(x; \theta)$ is the probability that a customer with feature vector x is willing to purchase at price p_i , and $A_i(\theta)$ is the corresponding population fraction. The valuation-segment representation will allow us to express both expected revenue and the censored likelihood in terms of the exposure accumulated by each segment under a markdown path.

4 Markdown Benchmark

We first characterize the optimal markdown schedule when the valuation parameter is known. Fix a duration vector $\Delta \in \mathcal{D}$ and consider a customer in valuation segment i . Such a customer becomes willing to purchase when the price reaches p_i and remains willing to purchase at every lower price. To purchase at price p_j , where $j \geq i$, the customer must have no checks during her active exposure at prices $p_i, \dots, p_{j-1}$, which occurs with probability $\exp\left(-\lambda \sum_{\ell=i}^{j-1} \Delta_\ell\right)$, and at least one check while p_j is posted, which occurs with probability $(1 - e^{-\lambda \Delta_j})$. Her expected revenue contribution is therefore

$$r_i(\Delta) = \sum_{j=i}^k p_j \exp\left(-\lambda \sum_{\ell=i}^{j-1} \Delta_\ell\right) (1 - e^{-\lambda \Delta_j}). \quad (9)$$

Averaging over valuation segments gives the expected revenue from the entire customer population:

$$R(\Delta, \theta) = n \sum_{i=1}^k \pi_i(\theta) r_i(\Delta). \quad (10)$$

We define the corresponding population markdown benchmark by

$$\text{OPT}(\theta) = \max_{\Delta \in \mathcal{D}} R(\Delta, \theta). \quad (11)$$

The population objective depends only on the parameter θ and the known feature distribution F , which makes it analytically convenient. It is also close to a stronger realized-pool benchmark. As shown in Appendix B, the expected gain from choosing the markdown schedule after observing the realized segment counts is at most $O(\sqrt{n})$. Consequently, replacing the population benchmark with the hindsight-optimal markdown schedule for the realized pool does not change the order of the regret guarantees established below. Both benchmarks optimize over precommitted markdown schedules; the hindsight benchmark is stronger only because it observes the realized composition of the customer pool before selecting the price durations.

We impose the following regularity condition for the analysis.

Assumption 1 (Regularity). *The sets $\mathcal{X}$ and Θ are compact, and $\theta^* \in \text{int}(\Theta)$. The distribution G has a twice continuously differentiable density g . The functions g , g' , and g'' are uniformly bounded on a neighborhood of*

$$\{p_i - x^\top \theta : x \in \mathcal{X}, \theta \in \Theta, i \in [k]\}.$$

Assumption 1 ensures the smoothness of the segment probabilities while allowing considerable flexibility in the feature distribution. In particular, F may be discrete or supported on a lower-dimensional subset of $\mathbb{R}^d$, so the model accommodates a conventional intercept specification such as $X = (1, Z)$. Full-rank feature variation is required later, where it is needed for statistical identification.

In the following theorem, we formally establish the tractability of the population benchmark.

Theorem 1 (Concavity of the population markdown objective). *Under Assumption 1, the function $R(\Delta, \theta)$ is three-times continuously differentiable in (Δ, θ) and concave in Δ on $\mathcal{D}$*

for every $\theta \in \Theta$. Consequently, (11) is a finite-dimensional concave maximization problem over a compact simplex.

The concavity follows from a useful telescoping representation. For each segment i ,

$$r_i(\Delta) = p_i - \sum_{j=i}^{k-1} (p_j - p_{j+1}) \exp\left(-\lambda \sum_{\ell=i}^j \Delta_\ell\right) - p_k \exp\left(-\lambda \sum_{\ell=i}^k \Delta_\ell\right). \quad (12)$$

Each nonconstant term is the negative of a positive multiple of the exponential of an affine function of Δ . Since such exponentials are convex, $r_i(\Delta)$ is concave, and the population objective inherits concavity as a nonnegative weighted sum of the segment revenues. The complete proof, together with the comparison between the population and realized-pool benchmarks, is provided in Appendix B.

5 Learning from Censored Transactions

5.1 Censored likelihood

We begin by deriving the likelihood induced by a realized markdown path $p(\cdot)$. Let $\mathcal{H}_t$ denote the seller's observed history up to time t , including the posted prices, purchase times, transaction prices, and buyer features. For each valuation segment i , define its cumulative active exposure by

$$H_i(t) = \int_0^t \mathbf{1}\{p(s) \leq p_i\} ds. \quad (13)$$

Thus, $H_i(t)$ is the total amount of time by t during which a segment- i customer has been willing to purchase.

Suppose customer j purchases at time τ_j and has observed feature vector $X_j = x_j$. Conditional on belonging to segment i , the customer must survive without a purchase throughout her active exposure before τ_j and then check the price at time τ_j . Her conditional purchase density is therefore $\lambda e^{-\lambda H_i(\tau_j)}$, when the posted price is acceptable and zero otherwise. Integrating over the customer's latent valuation segment gives the buyer contribution to the likelihood

$$L^{\text{buy}}(\theta \mid x_j, \tau_j) = \lambda \sum_{i=1}^k \pi_i(x_j; \theta) \mathbf{1}\{p(\tau_j) \leq p_i\} e^{-\lambda H_i(\tau_j)}. \quad (14)$$

A customer may also remain in the pool without ever completing a transaction by time t . This occurs either because her valuation lies below the lowest available price or because she has not checked the price during any period in which she was willing to buy. Accordingly, the probability that a customer remains unpurchased through time t is

$$Q_t(\theta) = \pi_{k+1}(\theta) + \sum_{i=1}^k \pi_i(\theta) e^{-\lambda H_i(t)}. \quad (15)$$

Let B_t denote the set of customers who have purchased by time t , and let $m_t = |B_t|$. Up to factors that do not depend on θ , the resulting log-likelihood is

$$\ell_t(\theta) = \sum_{j \in B_t} \log L^{\text{buy}}(\theta \mid X_j, \tau_j) + (n - m_t) \log Q_t(\theta). \quad (16)$$

We define the censored maximum-likelihood estimator as any global maximizer

$$\hat{\theta}_t \in \arg \max_{\theta \in \Theta} \ell_t(\theta). \quad (17)$$

The survival term in (16) is essential. The observed buyers are selected not only by their willingness to pay, but also by the amount of active exposure they have accumulated and by the customers who have already purchased and left the market. Treating buyer features as an uncensored sample from F , or conditioning only on completed transactions, would ignore the information contained in the number of customers who remain in the pool and would generally lead to a misspecified likelihood.

5.2 Markdown-preserving learning policy

We now use the censored likelihood to construct a learning policy that preserves the markdown constraint throughout. The policy begins at the highest price p_1 and screens each price for

$$h_n = \frac{c_h \log n}{n}, \quad m_0 = \lceil c_m \sqrt{n} \rceil, \quad (18)$$

where m_0 is the target number of transactions used for estimation.

If no transaction occurs during a screening interval, the policy marks down to the next price. Once a transaction is observed, the current price is declared active, and the policy continues posting that price until a total of m_0 transactions has been collected. Let τ_0 denote the time of the m_0 th transaction, and suppose the collecting price is p_i . At time τ_0 , the seller computes $\hat{\theta}_{\tau_0}$ and then commits to an estimated optimal suffix schedule over the remaining prices $\{p_i, \dots, p_k\}$ and the remaining horizon $[\tau_0, T]$.

The continuation problem is formulated for the residual customer pool. Conditional on $\mathcal{H}_t$ and on a customer remaining unpurchased at time t , the segment probabilities under parameter θ are

$$\tilde{\pi}_h(\theta \mid \mathcal{H}_t) = \frac{\pi_h(\theta) e^{-\lambda H_h(t)}}{Q_t(\theta)}, \quad h \leq k, \quad \tilde{\pi}_{k+1}(\theta \mid \mathcal{H}_t) = \frac{\pi_{k+1}(\theta)}{Q_t(\theta)}. \quad (19)$$

These probabilities reweight each valuation segment by its probability of surviving without a purchase through time t . Let $N_t = n - m_t$, $T_t = T - t$, and define the feasible suffix-duration set

$$\mathcal{D}_i(T_t) = \left\{ (\Delta_i, \dots, \Delta_k) \in \mathbb{R}_+^{k-i+1} : \sum_{j=i}^k \Delta_j = T_t \right\}. \quad (20)$$

Consider a remaining customer who originally belongs to segment h . If the continuation schedule begins at price p_i , the first price she is willing to accept is indexed by

$$a_i(h) = \max\{i, h\}.$$

The expected continuation revenue from her is therefore

$$r_{h,i}^{\text{suf}}(\Delta) = \sum_{j=a_i(h)}^k p_j \exp\left(-\lambda \sum_{\ell=a_i(h)}^{j-1} \Delta_\ell\right) (1 - e^{-\lambda \Delta_j}). \quad (21)$$

Averaging over the survival-adjusted segment distribution gives the conditional suffix objective

$$R_i^{\text{suf}}(\Delta, \theta \mid \mathcal{H}_t) = N_t \sum_{h=1}^k \tilde{\pi}_h(\theta \mid \mathcal{H}_t) r_{h,i}^{\text{suf}}(\Delta). \quad (22)$$

At time τ_0 , the policy solves (22) with $\theta = \hat{\theta}_{\tau_0}$ over $\mathcal{D}_i(T - \tau_0)$ and commits to the resulting markdown schedule. This continuation problem accounts simultaneously for the shorter remaining horizon and for the survival-weighted composition of the residual pool. The latter adjustment is particularly important because the data-collection phase disproportionately removes customers who are most likely to purchase at the collecting price. The complete pseudocode of the algorithm is provided in Appendix A.

6 Regret Analysis

We now analyze the performance of the learning policy. Fix the true parameter θ^* and define the first active price index by

$$i_0 = \min\{i : A_i(\theta^*) > 0\}, \quad Q^* = A_{i_0}(\theta^*) = \pi_{i_0}(\theta^*) > 0. \quad (23)$$

Thus, every price above p_{i_0} generates no transactions, whereas a positive fraction Q^* of the population is willing to purchase at p_{i_0} . Throughout the analysis, we hold T , k , d , λ , and the price ladder fixed as $n \rightarrow \infty$.

The exact inactivity of prices above p_{i_0} is a support-separation condition. If the valuation shock has full support, then $i_0 = 1$ and the screening stage is vacuous. If valuations have bounded support,

screening allows the policy to move past prices that cannot generate a transaction. Extending the analysis to prices with small but nonzero active mass would require a count-based screening test and is left for future work.

6.1 Identification and estimation

We begin with the statistical ingredients needed to learn θ^* from transactions at the first active price. The following conditions ensure that the observed buyer features contain sufficient local variation and that the buyer likelihood remains well behaved near the truth.

Assumption 2 (Informative feature variation). *There exist $c_q, c_g, \kappa > 0$ such that*

$$\mathbb{E}[XX^\top \mathbf{1}\{q_{i_0}(X; \theta^*) \geq c_q, g(p_{i_0} - X^\top \theta^*) \geq c_g\}] \succeq \kappa I_d. \quad (24)$$

Assumption 3 (Local overlap). *There exist a neighborhood $U_0 \subset \text{int}(\Theta)$ of θ^* and a constant $\underline{q} > 0$ such that*

$$q_{i_0}(x; \theta) \geq \underline{q}$$

for every $\theta \in U_0$ and every x in the support of the feature distribution conditional on willingness to purchase at p_{i_0} under θ^ .*

[Assumption 2](#) requires full-rank feature variation on a subset of customers whose willingness to purchase is both nonnegligible and locally sensitive to the parameter. [Assumption 3](#) ensures that the logarithmic buyer-likelihood terms remain uniformly controlled in a neighborhood of θ^* .

In the following proposition, we summarize the screening, collection, and estimation guarantees.

Proposition 1 (Screening, collection, and estimation). *Suppose [Assumptions 1 to 3](#) hold. There exist constants $C_0, C_\theta, C_\tau, C, c, c' > 0$ and $\rho \in (0, 1)$ such that, whenever*

$$C_0(\log n)^2 \leq m \leq \rho n Q^*,$$

the following statements hold.

- (i) *The screening procedure stops at p_{i_0} with probability at least $1 - n^{-4}$.*
- (ii) *If τ_m denotes the time of the m th transaction and t_{i_0} is the time at which the policy first posts p_{i_0} , then*

$$\mathbb{P}\left(\tau_m - t_{i_0} > C_\tau \frac{m}{n}\right) \leq e^{-cm} + n^{-4}.$$

- (iii) *For every $1 \leq u \leq c'm$, every global censored MLE based on the observations through the m th transaction satisfies*

$$\mathbb{P}\left(\|\hat{\theta}_m - \theta^*\| > C_\theta \sqrt{\frac{u}{m}}\right) \leq Ce^{-cu} + Ce^{-cm} + n^{-4}. \quad (25)$$

Consequently,

$$\mathbb{E}\|\hat{\theta}_m - \theta^*\|^2 \leq \frac{C}{m} + Ce^{-cm} + Cn^{-4}. \quad (26)$$

The proof proceeds in two stages. First, the stopped fixed-price experiment is represented as a marked order-statistic model, and a square-root likelihood-ratio bound establishes a uniform Hellinger separation away from θ^* . A finite covering argument, together with likelihood envelopes, extends this separation to the entire parameter space even when candidate purchase probabilities approach zero. Once every global maximizer has been localized to a neighborhood of θ^* , concentration of the score and uniform local curvature yield [\(25\)](#). The complete proof is given in [Appendix C](#).

6.2 Stability of the residual continuation problem

The estimation rate alone is not sufficient for a sharp regret bound. A direct Lipschitz comparison of continuation values would convert an $m^{-1/2}$ parameter error into an $O(nm^{-1/2})$ revenue loss, which would be too large. The optimization structure of the continuation problem removes this first-order effect.

Proposition 2 (Strong concavity and second-order loss). *Suppose data collection ends at time $\tau < T$ while the policy posts p_{i_0} , and suppose the residual pool satisfies*

$$N_\tau = n - m_\tau \geq \gamma n$$

for some constant $\gamma > 0$. There exist a neighborhood U_1 of θ^ and constants $\mu, C_\Delta, C > 0$, independent of n , such that, for every $\theta, \theta' \in U_1$:*

- (i) *the suffix objective $R_{i_0}^{\text{suf}}(\cdot, \theta \mid \mathcal{H}_\tau)$ is μn -strongly concave on the affine hull of $\mathcal{D}_{i_0}(T - \tau)$;*
- (ii) *the optimal suffix markdown solution $\Delta_{i_0}^*(\theta \mid \mathcal{H}_\tau)$ is unique and satisfies*

$$\|\Delta_{i_0}^*(\theta \mid \mathcal{H}_\tau) - \Delta_{i_0}^*(\theta' \mid \mathcal{H}_\tau)\| \leq C_\Delta \|\theta - \theta'\|; \quad (27)$$

- (iii) *optimizing under θ but evaluating under the true parameter θ^* incurs the second-order loss*

$$R_{i_0}^{\text{suf}}(\Delta_{i_0}^*(\theta^* \mid \mathcal{H}_\tau), \theta^* \mid \mathcal{H}_\tau) - R_{i_0}^{\text{suf}}(\Delta_{i_0}^*(\theta \mid \mathcal{H}_\tau), \theta^* \mid \mathcal{H}_\tau) \leq Cn \|\theta - \theta^*\|^2. \quad (28)$$

If $i_0 = k$, the feasible suffix set is a singleton and the conclusions hold trivially.

The residual-pool condition is stated directly because it is the source of uniform curvature. Under the learning policy, $m_\tau = m_0 = O(\sqrt{n})$, so $N_\tau \geq n/2$ for all sufficiently large n . Strong concavity is generated by the positive residual mass of segment i_0 , whose customers are active from the beginning of the suffix problem. First-order variational inequalities then yield both optimizer stability and the second-order loss bound, without requiring the optimizer to lie in the interior of the duration simplex. Appendix D contains the complete proof.

6.3 Regret against population and hindsight benchmarks

We next combine the statistical and optimization results. Let $J_n(\Pi, \theta^*)$ denote the expected revenue earned by the learning policy. For the realized customer pool, let $N_i(\theta^*)$ be the number of customers in valuation segment i , and define

$$R_n(\Delta, \theta^*) = \sum_{i=1}^k N_i(\theta^*) r_i(\Delta), \quad \text{OPT}_n^H(\theta^*) = \mathbb{E} \left[\max_{\Delta \in \mathcal{D}} R_n(\Delta, \theta^*) \right]. \quad (29)$$

The corresponding population and hindsight regrets are

$$\begin{aligned} \text{Regret}_n^{\text{pop}}(\Pi, \theta^*) &= \text{OPT}(\theta^*) - J_n(\Pi, \theta^*), \\ \text{Regret}_n^H(\Pi, \theta^*) &= \text{OPT}_n^H(\theta^*) - J_n(\Pi, \theta^*). \end{aligned} \quad (30)$$

Theorem 2 (Regret bound). *Suppose Assumptions 1 to 3 hold and $A_{i_0}(\theta^*) \geq a_0 > 0$. With*

$$h_n = \frac{c_h \log n}{n}, \quad m_0 = \lceil c_m \sqrt{n} \rceil,$$

for sufficiently large constants c_h and c_m ,

$$\text{Regret}_n^{\text{pop}}(\Pi, \theta^*) = O(\sqrt{n}), \quad \text{Regret}_n^H(\Pi, \theta^*) = O(\sqrt{n}). \quad (31)$$

The proof separates the regret incurred before and after estimation. Let

$$s_n = kh_n + C_\tau \frac{m_0}{n}.$$

With high probability, the policy reaches the first active price and completes data collection by time s_n . Coupling the policy with a population-optimal schedule that assigns zero duration to inactive prices gives

$$\begin{aligned} \mathbb{E}[\text{Loss}_{\text{pre}}] &\leq np_1(1 - e^{-\lambda s_n}) + o(1) = O(k \log n + m_0), \\ \mathbb{E}[\text{Loss}_{\text{post}}] &\leq Cn \mathbb{E}[\|\hat{\theta}_{\tau_0} - \theta^*\|^2] + o(1) = O(n/m_0). \end{aligned}$$

The first term captures the cost of screening and data collection, while the second follows from the second-order continuation loss in [Proposition 2](#). Choosing m_0 of order $\sqrt{n}$ balances these two contributions and yields $O(\sqrt{n})$ regret against the population benchmark. The uniform population-pool comparison then gives the same order relative to the hindsight-optimal schedule for the realized pool. The full proof appears in [Appendix E](#).

6.4 Instance-dependent free learning

The preceding result treats data collection as potentially costly. In some instances, however, the population-optimal schedule itself spends a nonvanishing amount of time at the collecting price. The policy can then gather information while following an economically optimal path.

For $\theta \in \Theta$, let

$$i_0(\theta) = \min\{i : A_i(\theta) > 0\}.$$

For constants $a_0, \underline{\Delta} > 0$, let Θ_{free} denote the set of parameters for which

$$A_{i_0(\theta)}(\theta) \geq a_0$$

and at least one population-optimal schedule assigns duration at least $\underline{\Delta}$ to $p_{i_0(\theta)}$. This definition is intrinsic and does not depend on how an optimizer is selected when several optimal schedules exist.

Theorem 3 (Instance-dependent logarithmic regret). *Suppose [Assumptions 1 to 3](#) hold uniformly on Θ_{free} . Choose*

$$m_0 = \lceil \gamma n \rceil$$

with

$$0 < \gamma < \min \left\{ \frac{a_0}{2} (1 - e^{-\lambda \underline{\Delta}/2}), \rho a_0 \right\}.$$

Then, for all sufficiently large n such that

$$kh_n \leq \underline{\Delta}/2,$$

the learning policy achieves regret $O(\log n)$ uniformly over Θ_{free} ,

$$\text{Regret}_n^{\text{pop}}(\Pi, \theta) = O(\log n).$$

If $i_0(\theta) = 1$, the bound improves to $O(1)$.

In this regime, the benchmark itself generates $\Theta(n)$ transactions at the collecting price. Data collection therefore creates no first-order opportunity cost. The only first-order loss is the $O(kh_n)$ delay incurred while screening inactive prices, while estimation error and continuation optimization each contribute only $O(1)$. [Appendix F](#) provides the proof.

6.5 A local exploration-estimation tradeoff

The free-learning result does not extend uniformly to parameters near the boundary at which the high price changes from being optimally skipped to receiving positive duration. To illustrate the difficulty, consider a one-dimensional submodel with $X \equiv 1$, a smooth bounded-support distribution G , and two prices $p_1 > p_2$ such that every customer is willing to purchase at p_2 . Define

$$\alpha(\theta) = 1 - G(p_1 - \theta).$$

Locally, $\alpha(\theta)$ is one-to-one, and the entire observation law depends on θ only through the high-valuation mass α .

A policy posts p_1 until a possibly randomized, history-dependent stopping time τ and then permanently marks down to p_2 . For a deterministic high-price duration Δ , the expected revenue per customer is

$$\begin{aligned} r(\Delta, \alpha) &= \alpha \left[p_1 (1 - e^{-\lambda \Delta}) + p_2 e^{-\lambda \Delta} (1 - e^{-\lambda(T-\Delta)}) \right] \\ &\quad + (1 - \alpha) p_2 (1 - e^{-\lambda(T-\Delta)}). \end{aligned} \tag{32}$$

The function $r(\cdot, \alpha)$ is strictly concave. Define

$$\alpha_0 = \frac{p_2 e^{-\lambda T}}{(p_1 - p_2) + p_2 e^{-\lambda T}}, \quad (33)$$

so that

$$\partial_{\Delta} r(0, \alpha_0) = 0.$$

For

$$\alpha_{\pm} = \alpha_0 \pm \delta,$$

the optimal high-price duration is zero under α_- and of order $\Theta(\delta)$ under α_+ . The gain from using the high price under α_+ is therefore of order $\Theta(n\delta^2)$.

Let M_t denote the number of customers remaining in the pool while p_1 is posted, and define the expected high-price exposure under α_- by

$$\mathcal{A}_-(\Pi) = \mathbb{E}_- \left[\int_0^{\tau} M_t dt \right].$$

Let

$$\text{Regret}_n(\Pi, \alpha) = n \max_{\Delta \in [0, T]} r(\Delta, \alpha) - J_n(\Pi, \alpha).$$

The same exposure quantity governs both the economic and statistical sides of the learning problem:

$$\text{Regret}_n(\Pi, \alpha_-) \geq c\delta \mathcal{A}_-(\Pi), \quad \text{KL}(\mathbb{P}_-^{\Pi}, \mathbb{P}_+^{\Pi}) \leq C\delta^2 \mathcal{A}_-(\Pi).$$

Thus, accumulating enough exposure to distinguish the two instances is costly under α_- , while insufficient exposure leaves the instances statistically close and leads to decision loss under α_+ .

Proposition 3 (Local regret lower bound). *Fix T , λ , and $p_1 > p_2 > 0$. There exist constants $c, \delta_0 > 0$ and $n_0 < \infty$ such that, for every $n \geq n_0$ and every $\delta \in (0, \delta_0]$ satisfying $n\delta^2 \geq 1$, every randomized, history-dependent markdown policy satisfies*

$$\sup_{\alpha \in \{\alpha_0 - \delta, \alpha_0 + \delta\}} \text{Regret}_n(\Pi, \alpha) \geq c \min\{\delta^{-1}, n\delta^2\}. \quad (34)$$

Consequently, for $\delta_n = n^{-1/3}$,

$$\inf_{\Pi} \sup_{\alpha \in \{\alpha_0 - \delta_n, \alpha_0 + \delta_n\}} \text{Regret}_n(\Pi, \alpha) \geq cn^{1/3}.$$

The lower bound is local and is not intended to match the uniform $O(\sqrt{n})$ upper bound in [Theorem 2](#). Rather, it distinguishes the free-learning region from the skip-use boundary. Near that boundary, every policy must incur a polynomial exploration cost in order to determine whether the high price should receive positive duration. [Appendix G](#) provides the complete stopped-process proof.

7 Conclusion

We study feature-based pricing in a persistent customer pool under transaction-only feedback. In this setting, customer features are observed only after purchase, unsuccessful price checks remain hidden, and each transaction changes the composition of the residual market. To address these challenges, we combine a censored maximum-likelihood estimator with a continuation problem that explicitly accounts for both prior exposure and endogenous depletion.

The resulting markdown-preserving policy achieves $O(\sqrt{n})$ expected regret relative to both the deterministic population benchmark and the hindsight-optimal markdown schedule for the realized valuation pool. When the benchmark itself spends sufficient time at an informative price, the policy can learn while following an economically optimal path, reducing regret to $O(\log n)$. By contrast, a two-price parametric submodel exhibits a local $\Omega(n^{1/3})$ lower bound near the boundary at which the high price changes from being optimally skipped to receiving positive duration. This contrast identifies the central source of difficulty: learning is inexpensive when informative

exposure is naturally available, but becomes costly when the seller must experiment to determine the appropriate markdown structure.

More broadly, the analysis shows that customer features can support demand extrapolation even when they are unavailable at the time of pricing and are observed only for completed transactions. Several extensions remain open, including screening when high prices have small but nonzero demand, jointly learning customer attention rates, allowing richer or continuous price sets, and developing policies that update continually as the customer pool evolves.

Bibliography

- [1] Aviv, Y., Pazgal, A.: Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management* **10**(3), 339–359 (2008) [3](#)
- [2] Ban, G.Y., Keskin, N.B.: Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science* **67**(9), 5549–5568 (2021) [1](#), [2](#)
- [3] Besanko, D., Winston, W.L.: Optimal price skimming by a monopolist facing rational consumers. *Management Science* **36**(5), 555–567 (1990) [3](#)
- [4] Besbes, O., Zeevi, A.: Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research* **57**(6), 1407–1420 (2009) [1](#), [2](#)
- [5] den Boer, A.V.: Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* **20**(1), 1–18 (2015) [2](#)
- [6] den Boer, A.V., Zwart, B.: Simultaneously learning and optimizing using controlled variance pricing. *Management Science* **60**(3), 770–783 (2014) [2](#)
- [7] Broder, J., Rusmevichientong, P.: Dynamic pricing under a general parametric choice model. *Operations Research* **60**(4), 965–980 (2012) [1](#), [2](#)
- [8] Bu, J., Simchi-Levi, D., Wang, C.: Context-based dynamic pricing with separable demand models. *Management Science* (2025), forthcoming [1](#), [2](#)
- [9] Bu, J., Simchi-Levi, D., Wang, L.: Offline pricing and demand learning with censored data. *Management Science* **69**(2), 885–903 (2023) [3](#)
- [10] Bu, J., Simchi-Levi, D., Xu, Y., Zhai, C.: Feature-based dynamic pricing with online learning and offline data (2025), available at SSRN 5261068 [2](#)
- [11] Cachon, G.P., Swinney, R.: Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science* **55**(3), 497–511 (2009) [3](#)
- [12] Cachon, G.P., Swinney, R.: The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science* **57**(4), 778–795 (2011) [3](#)
- [13] Caro, F., Gallien, J.: Clearance pricing optimization for a fast-fashion retailer. *Operations Research* **60**(6), 1404–1422 (2012) [3](#)
- [14] Chen, B., Chao, X., Shi, C.: Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Mathematics of Operations Research* **46**(2), 726–756 (2021) [3](#)
- [15] Chen, X., Simchi-Levi, D., Wang, Y.: Utility fairness in contextual dynamic pricing with demand learning. *Management Science* **72**(3), 2619–2633 (2026) [2](#)
- [16] Choi, Y.G., Kim, G.S., Choi, Y., Cho, W., Paik, M.C., Oh, M.: Semi-parametric contextual pricing algorithm using cox proportional hazards model. In: *Proceedings of the International Conference on Machine Learning*. pp. 5771–5786 (2023) [3](#)
- [17] Coase, R.H.: Durability and monopoly. *Journal of Law and Economics* **15**(1), 143–149 (1972) [3](#)
- [18] Cohen, M.C., Lobel, I., Paes Leme, R.: Feature-based dynamic pricing. *Management Science* **66**(11), 4921–4943 (2020) [1](#), [2](#)
- [19] Conlisk, J., Gerstner, E., Sobel, J.: Cyclic pricing by a durable goods monopolist. *Quarterly Journal of Economics* **99**(3), 489–505 (1984) [3](#)
- [20] Cox, D.R.: Regression models and life-tables. *Journal of the Royal Statistical Society: Series B* **34**(2), 187–202 (1972) [3](#)
- [21] Cui, T., Jia, S., Lavastida, T.: From stream to pool: Pricing under the law of diminishing marginal utility (2023), arXiv:2310.19220 [1](#), [3](#)
- [22] Elmaghraby, W., Gülcü, A., Keskinocak, P.: Designing optimal preannounced markdowns in the presence of rational customers with multiunit demands. *Manufacturing & Service Operations Management* **10**(1), 126–148 (2008) [3](#)

- [23] Heching, A., Gallego, G., van Ryzin, G.: Mark-down pricing: An empirical analysis of policies and revenue potential at one apparel retailer. *Journal of Revenue and Pricing Management* **1**(2), 139–160 (2002) [3](#)
- [24] Huh, W.T., Rusmevichientong, P.: A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1), 103–123 (2009) [3](#)
- [25] Javanmard, A., Nazerzadeh, H.: Dynamic pricing in high-dimensions. *Journal of Machine Learning Research* **20**(9), 1–49 (2019) [1](#), [2](#)
- [26] Javanmard, A., Nazerzadeh, H., Shao, S.: Multi-product dynamic pricing in high-dimensions with heterogeneous price sensitivity. In: *IEEE International Symposium on Information Theory*. pp. 2652–2657 (2020) [2](#)
- [27] Jia, S., Li, A., Ravi, R.: Markdown pricing under unknown demand (2021), available at SSRN 3861379 [3](#)
- [28] Jia, S., Li, A., Ravi, R.: Dynamic pricing with monotonicity constraint under unknown parametric demand model. In: *Advances in Neural Information Processing Systems*. vol. 35, pp. 19179–19188 (2022) [3](#)
- [29] Keskin, N.B., Zeevi, A.: Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research* **62**(5), 1142–1167 (2014) [1](#), [2](#)
- [30] Kleinberg, R., Leighton, T.: The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In: *44th Annual IEEE Symposium on Foundations of Computer Science*. pp. 594–605 (2003) [1](#), [2](#)
- [31] Mersereau, A.J., Zhang, D.: Markdown pricing with unknown fraction of strategic customers. *Manufacturing & Service Operations Management* **14**(3), 355–370 (2012) [3](#)
- [32] Perakis, G., Singhvi, D.: Dynamic pricing with unknown nonparametric demand and limited price changes. *Operations Research* **72**(6), 2726–2744 (2024) [3](#)
- [33] Qiang, S., Bayati, M.: Dynamic pricing with demand covariates (2016), arXiv:1604.07463 [1](#), [2](#)
- [34] Smith, S.A., Achabal, D.D.: Clearance pricing and inventory policies for retail chains. *Management Science* **44**(3), 285–300 (1998) [3](#)
- [35] Sobel, J.: Durable goods monopoly with entry of new consumers. *Econometrica* **59**(5), 1455–1485 (1991) [3](#)
- [36] Stokey, N.L.: Intertemporal price discrimination. *Quarterly Journal of Economics* **93**(3), 355–371 (1979)
- [37] Su, X.: Intertemporal pricing with strategic customer behavior. *Management Science* **53**(5), 726–741 (2007) [3](#)
- [38] Wang, Z., Deng, S., Ye, Y.: Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research* **62**(2), 318–331 (2014) [1](#), [2](#)
- [39] Xu, J., Wang, Y., Chen, X., Wang, Y.X.: Dynamic pricing with adversarially-censored demands (2025), arXiv:2502.06168 [3](#)
- [40] Yin, R., Aviv, Y., Pazgal, A., Tang, C.S.: Optimal markdown pricing: Implications of inventory display formats in the presence of strategic customers. *Management Science* **55**(8), 1391–1408 (2009) [3](#)

A Complete Learning Policy

This appendix summarizes the complete markdown-preserving learn-then-optimize policy. The policy proceeds in three stages. It first screens prices from high to low until it identifies an active price. It then remains at that price until the target number of transactions has been collected. Finally, it estimates the valuation parameter, updates the composition of the residual pool, and commits to an optimized suffix schedule for the remaining horizon.

Algorithm 1 Markdown-Preserving Learn-Then-Optimize Policy

Require: Price ladder $\mathcal{P}$, horizon T , screening length h_n , transaction target m_0

```

1: Initialize  $t \leftarrow 0$ ,  $i \leftarrow 1$ ,  $B \leftarrow \emptyset$ , and  $m \leftarrow 0$ 
2: while  $t < T$  and  $i \leq k$  do
3:   Set  $\delta \leftarrow \min\{h_n, T - t\}$ 
4:   Post price  $p_i$  on  $(t, t + \delta]$  and record all transactions
5:   Update the buyer set  $B$ , set  $m \leftarrow |B|$ , and set  $t \leftarrow t + \delta$ 
6:   if no transaction occurs during the screening interval and  $i < k$  then
7:     Set  $i \leftarrow i + 1$  and continue
8:   end if
9:   while  $m < m_0$  and  $t < T$  do
10:    Continue posting  $p_i$  until the next transaction or until time  $T$ 
11:    Update the buyer set  $B$ , set  $m \leftarrow |B|$ , and update  $t$ 
12:   end while
13:   if  $m < m_0$  then
14:     terminate
15:   end if
16:   Compute  $\hat{\theta}_t \in \arg \max_{\theta \in \Theta} \ell_t(\theta)$ 
17:   Compute the residual segment probabilities using (19)
18:   Solve (22) over  $\mathcal{D}_i(T - t)$  with residual pool size  $N_t = n - m$ 
19:   Commit to the resulting suffix schedule and terminate
20: end while

```

The policy never raises the posted price. Screening therefore preserves the markdown structure, while the collection phase keeps the price fixed so that the resulting transaction data admit the censored-likelihood analysis developed in Section 5. Once estimation is complete, the suffix problem is solved using both the remaining horizon and the survival-adjusted composition of the residual customer pool.

B Benchmark Details

We first establish the smoothness of the valuation-segment probabilities used in the population objective.

Lemma 1 (Smooth segment probabilities). *Under Assumption 1, each function $\pi_i(x; \theta)$ and $\pi_i(\theta)$ is three-times continuously differentiable in θ , with derivatives through order three uniformly bounded over their respective domains.*

Proof. Consider an interior segment $i \in \{2, \dots, k\}$. Differentiating (5) gives

$$\nabla_{\theta} \pi_i(x; \theta) = [g(p_i - x^{\top} \theta) - g(p_{i-1} - x^{\top} \theta)]x,$$

and

$$\nabla_{\theta\theta}^2 \pi_i(x; \theta) = [g'(p_{i-1} - x^{\top} \theta) - g'(p_i - x^{\top} \theta)]xx^{\top}.$$

The third derivative is obtained analogously as the difference of the corresponding g'' terms, multiplied by $x^{\otimes 3}$. For the endpoint segments $i = 1$ and $i = k + 1$, the same calculations apply with only one finite boundary term.

Because $\mathcal{X}$ and Θ are compact, and because g , g' , and g'' are uniformly bounded on the relevant neighborhood by [Assumption 1](#), all derivatives up to order three are uniformly bounded. Dominated convergence then permits differentiation under the integral in (7), establishing the same smoothness and boundedness properties for $\pi_i(\theta)$. $\square$

Proof (Proof of [Theorem 1](#)). Expanding the segment-revenue expression in (9) and collecting adjacent price terms yields the telescoping representation (12). Each exponential term in that representation is convex in Δ because its exponent is affine in the duration vector. It follows that $r_i(\Delta)$ is a constant minus a nonnegative weighted sum of convex functions and is therefore concave.

The population objective in (10) is a nonnegative weighted sum of the segment-revenue functions r_i , so it is also concave in Δ . Its smoothness in (Δ, θ) follows from the exponential representation of r_i and the smoothness of the segment probabilities established in [Lemma 1](#). $\square$

We next compare the deterministic population benchmark with the corresponding benchmark for a realized customer pool. For a fixed parameter θ , let

$$N_i(\theta) = \sum_{j=1}^n \mathbf{1} \{p_i \leq X_j^\top \theta + \varepsilon_j < p_{i-1}\}, \quad R_n(\Delta, \theta) = \sum_{i=1}^k N_i(\theta) r_i(\Delta).$$

Thus, $N_i(\theta)$ is the realized number of customers in valuation segment i , and $R_n(\Delta, \theta)$ is the corresponding realized-pool revenue.

Proposition 4 (Uniform population-pool comparison). *Fix $\theta \in \Theta$. For every $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\sup_{\Delta \in \mathcal{D}} |R(\Delta, \theta) - R_n(\Delta, \theta)| \leq kp_1 \sqrt{\frac{n \log(2k/\delta)}{2}}. \quad (35)$$

Consequently,

$$0 \leq \text{OPT}_n^H(\theta) - \text{OPT}(\theta) \leq \frac{kp_1}{2} \sqrt{n}. \quad (36)$$

Proof. For each segment i , the normalized count $N_i(\theta)/n$ is the empirical mean of n i.i.d. Bernoulli random variables with mean $\pi_i(\theta)$. Hoeffding's inequality and a union bound over $i \in [k]$ therefore imply that, with probability at least $1 - \delta$,

$$\max_{i \leq k} \left| \frac{N_i(\theta)}{n} - \pi_i(\theta) \right| \leq \sqrt{\frac{\log(2k/\delta)}{2n}}.$$

Moreover, the expected revenue contribution of any customer is bounded by the highest available price, so

$$0 \leq r_i(\Delta) \leq p_i \leq p_1$$

uniformly over $\Delta \in \mathcal{D}$. Hence,

$$\begin{aligned} |R_n(\Delta, \theta) - R(\Delta, \theta)| &= \left| \sum_{i=1}^k (N_i(\theta) - n\pi_i(\theta)) r_i(\Delta) \right| \\ &\leq kp_1 \max_{i \leq k} |N_i(\theta) - n\pi_i(\theta)|. \end{aligned}$$

Taking the supremum over $\Delta \in \mathcal{D}$ proves (35).

For the expected comparison, the inequality

$$\max_{\Delta \in \mathcal{D}} R_n(\Delta, \theta) - \max_{\Delta \in \mathcal{D}} R(\Delta, \theta) \leq \sup_{\Delta \in \mathcal{D}} |R_n(\Delta, \theta) - R(\Delta, \theta)|$$

gives

$$\text{OPT}_n^H(\theta) - \text{OPT}(\theta) \leq p_1 \sum_{i=1}^k \mathbb{E} |N_i(\theta) - n\pi_i(\theta)|.$$

Since

$$\mathbb{E} |N_i(\theta) - n\pi_i(\theta)| \leq \sqrt{n\pi_i(\theta)(1 - \pi_i(\theta))} \leq \frac{\sqrt{n}}{2},$$

we obtain the upper bound in (36). Finally, nonnegativity follows from Jensen's inequality, because the pointwise maximum is a convex function of the realized segment counts:

$$\mathbb{E} \left[\max_{\Delta \in \mathcal{D}} R_n(\Delta, \theta) \right] \geq \max_{\Delta \in \mathcal{D}} \mathbb{E} [R_n(\Delta, \theta)] = \text{OPT}(\theta).$$

□

C Statistical Learning Guarantees

In this section, we establish the screening, collection-time, and estimation guarantees summarized in Proposition 1. Throughout, all constants may depend on the fixed model primitives $(T, k, d, \lambda, \mathcal{P}, \Theta, \mathcal{X}, F, G)$ and on the true parameter θ^* , but not on n , m , or the deviation parameter u .

Let t_{i_0} denote the time at which the policy first posts the first active price p_{i_0} , and let $\mathcal{E}_{\text{scr}}$ be the event that the screening procedure stops at this price.

C.1 Screening and collection

Lemma 2 (Screening). *With $h_n = c_h \log n/n$, we have*

$$\mathbb{P}(\mathcal{E}_{\text{scr}}^c) \leq \exp\{-nQ^*(1 - e^{-\lambda h_n})\}.$$

For a sufficiently large constant c_h , this probability is at most n^{-4} .

Proof. By the definition of i_0 , every price strictly above p_{i_0} has zero active mass under θ^* . Those screening intervals therefore produce no transactions almost surely, and the policy reaches p_{i_0} .

At p_{i_0} , a generic customer is willing to purchase with probability Q^* . Conditional on being willing to purchase, the probability of at least one check during an interval of length h_n is $1 - e^{-\lambda h_n}$. These events are independent across customers, so the probability that the screening interval at p_{i_0} produces no transaction is

$$[1 - Q^*(1 - e^{-\lambda h_n})]^n.$$

Using $1 - x \leq e^{-x}$ gives the first claim.

Since $h_n \rightarrow 0$, for all sufficiently large n ,

$$1 - e^{-\lambda h_n} \geq \frac{\lambda h_n}{2}.$$

Hence

$$\mathbb{P}(\mathcal{E}_{\text{scr}}^c) \leq \exp\left\{-\frac{\lambda Q^* c_h}{2} \log n\right\}.$$

Choosing c_h so that $\lambda Q^* c_h/2 \geq 4$ yields the bound n^{-4} . □

Once the policy reaches p_{i_0} and observes a transaction, it keeps that price fixed until the collection target is reached. The next lemma controls the corresponding collection time.

Lemma 3 (Collection time). *Suppose*

$$C \log n \leq m \leq \rho n Q^*$$

for a sufficiently small constant $\rho > 0$. There exist constants $C_\tau, c_\tau > 0$ such that

$$\mathbb{P}\left(\tau_m - t_{i_0} > C_\tau \frac{m}{n}\right) \leq e^{-c_\tau m}$$

under the experiment that posts p_{i_0} continuously after t_{i_0} . On $\mathcal{E}_{\text{scr}}$, the same bound applies to the actual learning policy.

Proof. Set

$$s = C_\tau \frac{m}{n}.$$

If p_{i_0} is posted continuously on $(t_{i_0}, t_{i_0} + s]$, then each customer purchases during this interval with probability

$$Q^*(1 - e^{-\lambda s}).$$

Consequently, the number of purchases during the interval satisfies

$$N(s) \sim \text{Binomial}(n, Q^*(1 - e^{-\lambda s})).$$

Choose ρ small enough that

$$s \leq C_\tau \rho Q^*$$

lies in a fixed neighborhood of zero. Then

$$1 - e^{-\lambda s} \geq \frac{\lambda s}{2}.$$

It follows that

$$\mathbb{E}[N(s)] = nQ^*(1 - e^{-\lambda s}) \geq \frac{\lambda C_\tau Q^*}{2} m.$$

Choose C_τ so that the right-hand side is at least $2m$. A multiplicative Chernoff bound then gives

$$\mathbb{P}(N(s) < m) \leq e^{-c_\tau m}.$$

The event

$$\{\tau_m - t_{i_0} > s\}$$

is precisely the event that fewer than m purchases occur in the first s units of time at p_{i_0} , which proves the claim.

On $\mathcal{E}_{\text{scr}}$, the actual policy posts p_{i_0} continuously from t_{i_0} until the m th transaction, so the two experiments coincide up to τ_m . $\square$

C.2 A stopped fixed-price experiment

For the remainder of the section, let

$$p = p_{i_0}, \quad q_\theta(x) = 1 - G(p - x^\top \theta), \quad Q_\theta = \mathbb{E}[q_\theta(X)],$$

and let

$$q^* = q_{\theta^*}, \quad Q^* = Q_{\theta^*}.$$

Conditional on being willing to purchase at p , a customer's feature vector has a distribution

$$dF^B(x) = \frac{q^*(x)}{Q^*} dF(x). \quad (37)$$

We first record two properties of the transformed stopping time. For each customer, let E_j be the first checking time after t_{i_0} . Define

$$Y_j = 1 - e^{-\lambda E_j}$$

if the customer is willing to purchase at p , and set $Y_j = +\infty$ otherwise. Under the true parameter,

$$\mathbb{P}_{\theta^*}(Y_j \leq w \mid X_j = x) = wq^*(x), \quad 0 \leq w \leq 1. \quad (38)$$

Let $Y_{(1)} < \dots < Y_{(m)}$ denote the first m finite marks, whenever they exist, and define

$$W_m = Y_{(m)} = 1 - e^{-\lambda(\tau_m - t_{i_0})}.$$

If fewer than m finite marks exist, set $W_m = 1$. This convention affects only an exponentially unlikely event under the range of m considered below.

Lemma 4 (Concentration of the transformed stopping time). *Let*

$$\alpha = \frac{m}{n}, \quad w_\alpha = \frac{\alpha}{Q^*}.$$

There exist constants $C, c, c' > 0$ and $\rho \in (0, 1)$ such that, whenever $m \leq \rho n Q^$, for every $1 \leq u \leq c'm$,*

$$\mathbb{P}\left(|W_m - w_\alpha| > Cw_\alpha \sqrt{\frac{u}{m}}\right) \leq 2e^{-cu}. \quad (39)$$

In particular, for every sufficiently small fixed $\eta \in (0, 1/2)$,

$$\mathbb{P}(\mathcal{E}_W^c) \leq Ce^{-cm}, \quad \mathcal{E}_W := \{|W_m - w_\alpha| \leq \eta w_\alpha\}. \quad (40)$$

After reducing ρ if necessary, there exists $\bar{\rho} < 1$ such that $W_m \leq \bar{\rho}$ on $\mathcal{E}_W$.

Moreover, with

$$\beta_m = \frac{(n-m)W_m}{m(1-W_mQ^*)}, \quad (41)$$

we have, for $1 \leq u \leq c'm$,

$$\mathbb{P}\left(\left|\beta_m - \frac{1}{Q^*}\right| > C\sqrt{\frac{u}{m}}\right) \leq 2e^{-cu}. \quad (42)$$

Proof. For a fixed $w \in [0, 1)$, the number of finite marks not exceeding w is

$$N(w) \sim \text{Binomial}(n, wQ^*).$$

Moreover,

$$\{W_m > w\} = \{N(w) < m\}, \quad \{W_m < w\} = \{N(w) \geq m\}.$$

Let

$$\zeta = K\sqrt{\frac{u}{m}},$$

where $K > 0$ is a sufficiently large constant. By choosing $c' > 0$ small enough, we may ensure that $\zeta \leq 1/2$ whenever $u \leq c'm$. Set

$$w_+ = (1 + \zeta)w_\alpha, \quad w_- = (1 - \zeta)w_\alpha.$$

After reducing ρ if necessary, both points lie in $(0, 1)$.

At w_+ ,

$$\mathbb{E}[N(w_+)] = n(1 + \zeta)w_\alpha Q^* = (1 + \zeta)m.$$

Thus,

$$\mathbb{P}(W_m > w_+) = \mathbb{P}(N(w_+) < m) \leq e^{-c\zeta^2 m} \leq e^{-cu}$$

by the lower-tail Chernoff inequality. Similarly,

$$\mathbb{E}[N(w_-)] = (1 - \zeta)m,$$

and the upper-tail Chernoff inequality gives

$$\mathbb{P}(W_m < w_-) = \mathbb{P}(N(w_-) \geq m) \leq e^{-c\zeta^2 m} \leq e^{-cu}.$$

Combining the two bounds proves (39). Taking u proportional to m gives (40).

Because

$$w_\alpha = \frac{\alpha}{Q^*} \leq \rho,$$

on $\mathcal{E}_W$,

$$W_m \leq (1 + \eta)\rho.$$

Choosing ρ so that $(1 + \eta)\rho < 1$ gives the claimed $\bar{\rho}$.

Finally, write

$$b_\alpha(w) = \frac{(1 - \alpha)w}{\alpha(1 - wQ^*)}.$$

Then

$$b_\alpha(w_\alpha) = \frac{1}{Q^*},$$

and, for $w \leq \bar{\rho}$,

$$|b'_\alpha(w)| = \frac{1 - \alpha}{\alpha(1 - wQ^*)^2} \leq \frac{C}{\alpha}.$$

Since

$$|W_m - w_\alpha| \leq C\frac{\alpha}{Q^*}\sqrt{\frac{u}{m}}$$

on the event in (39), the mean-value theorem yields

$$\left| \beta_m - \frac{1}{Q^*} \right| \leq C \sqrt{\frac{u}{m}}.$$

This proves (42). $\square$

The marked representation also yields a useful conditional factorization.

Lemma 5 (Marked order-statistic factorization). *On the event $W_m < 1$, the joint density of $(Y_{(1)}, \dots, Y_{(m)}, X_1^B, \dots, X_m^B)$*

with respect to

$$dy_1 \cdots dy_m dF(x_1) \cdots dF(x_m)$$

is, up to the parameter-independent factor $n!/(n-m)!$,

$$\mathbf{1}\{0 < y_1 < \cdots < y_m < 1\} \prod_{j=1}^m q^*(x_j) (1 - y_m Q^*)^{n-m}. \quad (43)$$

Consequently, conditional on $W_m = w$, the buyer features $X_1^B, \dots, X_m^B$ are i.i.d. from F^B and are independent of the ordered transformed purchase times.

Proof. Under (38), a customer with feature x contributes a finite transformed mark in dy with probability density $q^*(x) dy$, while the probability that the customer's mark exceeds w is

$$1 - wQ^*.$$

To obtain the first m ordered marks and their associated features, choose and order m of the n customers, which contributes the factor $n!/(n-m)!$, multiply their mark-feature densities, and require the remaining $n-m$ marks to exceed y_m . This gives (43). The dependence on the features in (43) factorizes as

$$\prod_{j=1}^m q^*(x_j) dF(x_j) = (Q^*)^m \prod_{j=1}^m dF^B(x_j).$$

The remaining terms depend only on the ordered marks. The asserted conditional independence follows. $\square$

C.3 Censored MLE concentration

Under a candidate parameter θ , the fixed-price likelihood through the m th purchase is, up to parameter-independent factors,

$$L_m^0(\theta) = \prod_{j=1}^m q_\theta(X_j^B) (1 - W_m Q_\theta)^{n-m}. \quad (44)$$

Let

$$\ell_m^0(\theta) = \log L_m^0(\theta), \quad \bar{\ell}_m^0(\theta) = \frac{1}{m} \ell_m^0(\theta).$$

The actual likelihood also accounts for the short screening exposures at prices above p_{i_0} . For $\ell < i_0$, write

$$d_\ell = H_\ell(t_{i_0}) \leq kh_n.$$

After removing parameter-independent factors, the buyer and survival contributions of the complete likelihood are

$$b_j(\theta) = q_\theta(X_j^B) - \sum_{\ell < i_0} \pi_\ell(X_j^B; \theta) (1 - e^{-\lambda d_\ell}), \quad (45)$$

$$S_m(\theta) = 1 - W_m Q_\theta - e^{-\lambda(\tau_m - t_{i_0})} \sum_{\ell < i_0} \pi_\ell(\theta) (1 - e^{-\lambda d_\ell}). \quad (46)$$

Thus,

$$L_m(\theta) = \prod_{j=1}^m b_j(\theta) S_m(\theta)^{n-m}.$$

Let

$$\ell_m(\theta) = \log L_m(\theta), \quad \bar{\ell}_m(\theta) = \frac{1}{m} \ell_m(\theta).$$

Lemma 6 (Censored MLE concentration). *Under [Assumptions 1 to 3](#), there exist constants $C_0, C_\theta, C, c, c' > 0$ and $\rho \in (0, 1)$ such that, whenever*

$$C_0(\log n)^2 \leq m \leq \rho n Q^*,$$

every global maximizer $\hat{\theta}_m$ of the likelihood observed through the m th transaction satisfies, for every $1 \leq u \leq c'm$,

$$\mathbb{P}\left(\|\hat{\theta}_m - \theta^*\| > C_\theta \sqrt{\frac{u}{m}}\right) \leq C e^{-cu} + C e^{-cm}, \quad (47)$$

apart from the screening-failure probability. Consequently,

$$\mathbb{E}\|\hat{\theta}_m - \theta^*\|^2 \leq \frac{C}{m} + C e^{-cm}, \quad (48)$$

again apart from the screening-failure contribution.

Proof. The proof has two stages. We first establish global localization: with probability at least $1 - C e^{-cm}$, every global maximizer lies in a fixed neighborhood of θ^* . We then prove concentration of the score and uniform strong concavity of the normalized log-likelihood on that neighborhood.

All arguments are carried out under the unconditional experiment that posts p continuously after t_{i_0} . On $\mathcal{E}_{\text{scr}}$, this experiment agrees with the actual policy through the m th transaction.

Step 1: Likelihood domination. The subtracted terms in (45) and (46) are nonnegative. At the true parameter, all valuation-segment masses above p_{i_0} vanish:

$$\pi_\ell(x; \theta^*) = 0 \quad \text{for } F\text{-almost every } x \text{ and every } \ell < i_0.$$

Indeed,

$$0 = A_{i_0-1}(\theta^*) = \sum_{\ell < i_0} \pi_\ell(\theta^*),$$

and each summand is nonnegative. Hence, almost surely,

$$b_j(\theta^*) = q^*(X_j^B), \quad S_m(\theta^*) = 1 - W_m Q^*.$$

It follows that

$$\frac{L_m(\theta)}{L_m(\theta^*)} \leq \frac{L_m^0(\theta)}{L_m^0(\theta^*)}, \quad \theta \in \Theta. \quad (49)$$

This inequality allows us to localize the complete MLE through the simpler fixed-price likelihood, without imposing a global lower bound on q_θ .

Step 2: Pointwise Hellinger separation. Conditional on $W_m = w$, [Lemma 5](#) implies that the buyer features are i.i.d. from F^B . Therefore,

$$\mathbb{E}_{\theta^*} \left[\left(\frac{L_m^0(\theta)}{L_m^0(\theta^*)} \right)^{1/2} \middle| W_m = w \right] = \Psi_{\alpha, w}(\theta)^m, \quad (50)$$

where

$$\Psi_{\alpha, w}(\theta) = a(\theta) \left(\frac{1 - w Q_\theta}{1 - w Q^*} \right)^{(1-\alpha)/(2\alpha)}, \quad (51)$$

$$a(\theta) = \frac{1}{Q^*} \int \sqrt{q_\theta(x) q^*(x)} dF(x). \quad (52)$$

At the central value

$$w = w_\alpha = \frac{\alpha}{Q^*},$$

set

$$z = \frac{Q_\theta}{Q^*}.$$

Cauchy-Schwarz inequality gives

$$a(\theta)^2 = \frac{1}{(Q^*)^2} \left(\int \sqrt{q_\theta q^*} dF \right)^2 \leq \frac{Q_\theta}{Q^*} = z.$$

Consequently,

$$\log \Psi_{\alpha, w_\alpha}(\theta) \leq \phi_\alpha(z),$$

where

$$\phi_\alpha(z) = \frac{1}{2} \log z + \frac{1-\alpha}{2\alpha} \log \left(\frac{1-\alpha z}{1-\alpha} \right).$$

For $\alpha = 0$, define the continuous extension

$$\phi_0(z) = \frac{1}{2}(\log z - z + 1).$$

A direct calculation gives

$$\phi'_\alpha(z) = \frac{1-z}{2z(1-\alpha z)}.$$

Thus ϕ_α is increasing on $(0, 1)$ and decreasing on $(1, \infty)$, and

$$\phi_\alpha(z) \leq 0$$

with equality if and only if $z = 1$.

Equality in the preceding Hellinger bound therefore requires $Q_\theta = Q^*$ and equality in Cauchy-Schwarz. The latter implies that q_θ is proportional to q^* F -almost surely. Since their integrals agree, the proportionality constant is one:

$$q_\theta(x) = q^*(x) \quad \text{for } F\text{-almost every } x. \quad (53)$$

We now show that (53) identifies θ^* . Define

$$\mathcal{S} = \{x : q^*(x) \geq c_q, \ g(p - x^\top \theta^*) \geq c_g\}.$$

If (53) holds, then for F -almost every $x \in \mathcal{S}$,

$$G(p - x^\top \theta) = G(p - x^\top \theta^*).$$

Because g is continuous and strictly positive at $p - x^\top \theta^*$, the function G is strictly increasing on a neighborhood of that point. Monotonicity of G then implies

$$x^\top (\theta - \theta^*) = 0$$

for F -almost every $x \in \mathcal{S}$. Hence

$$\begin{aligned} 0 &= (\theta - \theta^*)^\top \mathbb{E}[X X^\top \mathbf{1}\{X \in \mathcal{S}\}] (\theta - \theta^*) \\ &\geq \kappa \|\theta - \theta^*\|^2 \end{aligned}$$

by Assumption 2. Therefore $\theta = \theta^*$.

Choose $r_0 > 0$ sufficiently small that

$$\overline{B}(\theta^*, r_0) \subset U_0 \cap \text{int}(\Theta),$$

and define

$$\Theta_{r_0} = \{\theta \in \Theta : \|\theta - \theta^*\| \geq r_0\}.$$

The map

$$(\alpha, \theta) \mapsto \Psi_{\alpha, w_\alpha}(\theta)$$

has the continuous extension

$$\Psi_{0,0}(\theta) = a(\theta) \exp \left\{ \frac{1}{2} \left(1 - \frac{Q_\theta}{Q^*} \right) \right\}$$

at $\alpha = 0$. It is therefore continuous on the compact set

$$[0, \rho Q^*] \times \Theta_{r_0}.$$

By the strict-identification argument above, there exists $c_0 > 0$ such that

$$\sup_{0 \leq \alpha \leq \rho Q^*} \sup_{\theta \in \Theta_{r_0}} \Psi_{\alpha, w_\alpha}(\theta) \leq e^{-4c_0}. \quad (54)$$

For $w \leq \bar{\rho}$,

$$\partial_w \log \Psi_{\alpha, w}(\theta) = \frac{1 - \alpha}{2\alpha} \left[-\frac{Q_\theta}{1 - wQ_\theta} + \frac{Q^*}{1 - wQ^*} \right],$$

and therefore

$$|\partial_w \log \Psi_{\alpha, w}(\theta)| \leq \frac{C}{\alpha}.$$

On $\mathcal{E}_W$,

$$|W_m - w_\alpha| \leq \eta w_\alpha = \frac{\eta \alpha}{Q^*}.$$

Choosing η sufficiently small and applying the mean-value theorem to $\log \Psi_{\alpha, w}$ gives

$$\sup_{\theta \in \Theta_{r_0}} \Psi_{\alpha, W_m}(\theta) \leq e^{-3c_0} \quad \text{on } \mathcal{E}_W. \quad (55)$$

Step 3: Uniform localization by finite likelihood envelopes. The preceding bound is pointwise in θ . We now extend it uniformly over the uncountable set Θ_{r_0} .

By [Assumption 1](#), compactness of $\mathcal{X}$ and Θ , and the boundedness of g , there exists $L < \infty$ such that

$$|q_\theta(x) - q_\vartheta(x)| \leq L\|\theta - \vartheta\|, \quad |Q_\theta - Q_\vartheta| \leq L\|\theta - \vartheta\|. \quad (56)$$

Fix $\epsilon > 0$ and cover Θ_{r_0} by finitely many balls

$$B(\theta_s, \epsilon), \quad s = 1, \dots, N_\epsilon,$$

with centers $\theta_s \in \Theta_{r_0}$. The covering number N_ϵ depends only on the fixed compact set Θ and the fixed radius ϵ .

If $\theta \in B(\theta_s, \epsilon)$, then

$$q_\theta(X_j^B) \leq q_{\theta_s}(X_j^B) + L\epsilon$$

and

$$1 - W_m Q_\theta \leq 1 - W_m Q_{\theta_s} + W_m L\epsilon.$$

Thus

$$L_m^0(\theta) \leq \bar{L}_{m,s},$$

where

$$\bar{L}_{m,s} = \prod_{j=1}^m (q_{\theta_s}(X_j^B) + L\epsilon) (1 - W_m Q_{\theta_s} + W_m L\epsilon)^{n-m}.$$

Conditional on $W_m = w$,

$$\mathbb{E}_{\theta^*} \left[\left(\frac{\bar{L}_{m,s}}{L_m^0(\theta^*)} \right)^{1/2} \middle| W_m = w \right] = \bar{\Psi}_{\alpha, w, s}^m,$$

where

$$\bar{\Psi}_{\alpha, w, s} = \bar{a}_s \left(\frac{1 - wQ_{\theta_s} + wL\epsilon}{1 - wQ^*} \right)^{(1-\alpha)/(2\alpha)}$$

and

$$\bar{a}_s = \frac{1}{Q^*} \int \sqrt{(q_{\theta_s}(x) + L\epsilon)q^*(x)} dF(x).$$

Using

$$\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$$

and $\int q^* dF = Q^*$, we obtain

$$\bar{a}_s \leq a(\theta_s) + C\sqrt{\epsilon}.$$

Furthermore, on $\mathcal{E}_W$,

$$w = O(\alpha)$$

uniformly, and

$$1 - wQ_{\theta_s} \geq 1 - \bar{\rho} > 0.$$

Therefore,

$$\left(1 + \frac{wL\epsilon}{1 - wQ_{\theta_s}}\right)^{(1-\alpha)/(2\alpha)} \leq e^{C\epsilon}.$$

The unperturbed survival factor

$$\left(\frac{1 - wQ_{\theta_s}}{1 - wQ^*}\right)^{(1-\alpha)/(2\alpha)}$$

is uniformly bounded on the compact set under consideration. Combining these observations with (55) yields

$$\bar{\Psi}_{\alpha, W_m, s} \leq e^{C\epsilon} (e^{-3c_0} + C\sqrt{\epsilon}) \quad \text{on } \mathcal{E}_W.$$

Choose the fixed radius ϵ sufficiently small that

$$\max_{s \leq N_\epsilon} \bar{\Psi}_{\alpha, W_m, s} \leq e^{-2c_0} \quad \text{on } \mathcal{E}_W. \quad (57)$$

Conditional Markov's inequality now gives

$$\begin{aligned} \mathbb{P}(\bar{L}_{m,s} \geq L_m^0(\theta^*), \mathcal{E}_W) &= \mathbb{E} \left[\mathbf{1}_{\mathcal{E}_W} \mathbb{P} \left(\frac{\bar{L}_{m,s}}{L_m^0(\theta^*)} \geq 1 \mid W_m \right) \right] \\ &\leq \mathbb{E} \left[\mathbf{1}_{\mathcal{E}_W} \mathbb{E} \left[\left(\frac{\bar{L}_{m,s}}{L_m^0(\theta^*)} \right)^{1/2} \mid W_m \right] \right] \\ &\leq e^{-2c_0 m}. \end{aligned}$$

A union bound over the fixed cover therefore yields

$$\mathbb{P} \left(\sup_{\theta \in \Theta_{r_0}} L_m^0(\theta) \geq L_m^0(\theta^*), \mathcal{E}_W \right) \leq N_\epsilon e^{-2c_0 m} \leq C e^{-cm}.$$

Combining the preceding union bound with (40) and (49) gives

$$\mathbb{P}(\text{some global maximizer of } L_m \text{ lies outside } B(\theta^*, r_0)) \leq C e^{-cm}. \quad (58)$$

The conclusion holds simultaneously for every global maximizer.

Step 4: Effect of the preliminary screening intervals. We next compare the derivatives of the complete and fixed-price log-likelihoods on the localization ball. By [Assumption 3](#),

$$q_\theta(x) \geq \underline{q}$$

for every $\theta \in B(\theta^*, r_0)$ and every x in the support of F^B . Moreover,

$$1 - W_m Q_\theta \geq 1 - \bar{\rho}$$

on $\mathcal{E}_W$.

By [Lemma 1](#),

$$\sup_{\theta, x} \left\| \nabla_\theta^r \sum_{\ell < i_0} \pi_\ell(x; \theta) (1 - e^{-\lambda d_\ell}) \right\| \leq C h_n, \quad r = 0, 1, 2, 3,$$

because $d_\ell \leq k h_n$ and $1 - e^{-\lambda d_\ell} \leq \lambda d_\ell$. The analogous population bound also holds:

$$\sup_{\theta} \left\| \nabla_\theta^r \sum_{\ell < i_0} \pi_\ell(\theta) (1 - e^{-\lambda d_\ell}) \right\| \leq C h_n.$$

For all sufficiently large n , these bounds imply

$$b_j(\theta) \geq \frac{\underline{q}}{2}, \quad S_m(\theta) \geq \frac{1 - \bar{\rho}}{2}$$

uniformly over the localization ball.

Repeated differentiation of $\log b_j(\theta)$ and $\log S_m(\theta)$, together with the preceding denominator bounds, therefore gives, for $r = 1, 2, 3$,

$$\sup_{\theta \in B(\theta^*, r_0)} \left\| \nabla_\theta^r [\log b_j(\theta) - \log q_\theta(X_j^B)] \right\| \leq C h_n$$

and

$$\sup_{\theta \in B(\theta^*, r_0)} \|\nabla_{\theta}^r [\log S_m(\theta) - \log(1 - W_m Q_{\theta})]\| \leq C h_n.$$

After summing and dividing by m ,

$$\sup_{\theta \in B(\theta^*, r_0)} \|\nabla_{\theta}^r \bar{\ell}_m(\theta) - \nabla_{\theta}^r \bar{\ell}_m^0(\theta)\| \leq C \left(h_n + \frac{n-m}{m} h_n \right) \leq C \frac{\log n}{m}, \quad (59)$$

for $r = 1, 2, 3$. Here we used $h_n = c_h \log n/n$.

Step 5: Local score concentration. Let

$$\dot{q}(x) = \nabla q_{\theta}(x)|_{\theta=\theta^*}, \quad \dot{Q} = \nabla Q_{\theta}|_{\theta=\theta^*}.$$

Differentiating the fixed-price likelihood gives

$$\nabla \bar{\ell}_m^0(\theta^*) = \frac{1}{m} \sum_{j=1}^m \frac{\dot{q}(X_j^B)}{q^*(X_j^B)} - \beta_m \dot{Q},$$

where β_m is defined in (41). Combining this identity with (59) yields

$$\nabla \bar{\ell}_m(\theta^*) = \frac{1}{m} \sum_{j=1}^m \frac{\dot{q}(X_j^B)}{q^*(X_j^B)} - \beta_m \dot{Q} + r_{1m}, \quad (60)$$

with

$$\|r_{1m}\| \leq C \frac{\log n}{m}.$$

Under F^B ,

$$\mathbb{E}_B \left[\frac{\dot{q}(X)}{q^*(X)} \right] = \frac{1}{Q^*} \int \dot{q}(x) dF(x) = \frac{\dot{Q}}{Q^*}.$$

Moreover, Assumption 3, compactness of $\mathcal{X}$, and boundedness of g imply that

$$\left\| \frac{\dot{q}(X)}{q^*(X)} \right\| \leq C$$

almost surely under F^B .

By Lemma 5, the buyer features are i.i.d. from F^B and independent of W_m . Coordinatewise Bernstein inequalities therefore imply, for $1 \leq u \leq c'm$,

$$\mathbb{P} \left(\left\| \frac{1}{m} \sum_{j=1}^m \frac{\dot{q}(X_j^B)}{q^*(X_j^B)} - \frac{\dot{Q}}{Q^*} \right\| > C \sqrt{\frac{u}{m}} \right) \leq C e^{-cu}.$$

Together with (42),

$$\mathbb{P} \left(\left\| \left(\beta_m - \frac{1}{Q^*} \right) \dot{Q} \right\| > C \sqrt{\frac{u}{m}} \right) \leq C e^{-cu}.$$

Finally,

$$m \geq C_0 (\log n)^2$$

implies

$$\frac{\log n}{m} \leq \frac{C}{\sqrt{m}} \leq C \sqrt{\frac{u}{m}} \quad \text{for } u \geq 1.$$

Substituting these bounds into (60) gives

$$\mathbb{P} \left(\|\nabla \bar{\ell}_m(\theta^*)\| > C \sqrt{\frac{u}{m}} \right) \leq C e^{-cu} + C e^{-cm}. \quad (61)$$

The e^{-cm} term accounts for $\mathcal{E}_W^c$.

Step 6: Uniform local curvature. We first study the conditional expectation of the negative Hessian of the fixed-price log-likelihood. For $\alpha > 0$, write

$$v = \frac{W_m}{\alpha}.$$

Conditional on $W_m = \alpha v$, define

$$\mathcal{I}(\alpha, v, \theta) = \mathbb{E}_{\theta^*} [-\nabla^2 \bar{\ell}_m^0(\theta) \mid W_m = \alpha v].$$

Using the i.i.d. buyer representation,

$$\mathcal{I}(\alpha, v, \theta) = -\mathbb{E}_B[\nabla^2 \log q_\theta(X)] + \frac{(1-\alpha)v}{1-\alpha v Q_\theta} \nabla^2 Q_\theta + \frac{\alpha(1-\alpha)v^2}{(1-\alpha v Q_\theta)^2} \nabla Q_\theta \nabla Q_\theta^\top. \quad (62)$$

This expression has a continuous extension to $\alpha = 0$.

At $\theta = \theta^*$ and $v = 1/Q^*$,

$$-\mathbb{E}_B[\nabla^2 \log q^*(X)] = \frac{1}{Q^*} \int \left[\frac{\dot{q}(x)\dot{q}(x)^\top}{q^*(x)} - \nabla^2 q_{\theta^*}(x) \right] dF(x).$$

The coefficient of $\nabla^2 Q_{\theta^*}$ in the second term of (62) is

$$\frac{(1-\alpha)/Q^*}{1-\alpha} = \frac{1}{Q^*}.$$

Thus the $\nabla^2 Q_{\theta^*}$ terms cancel, leaving

$$\mathcal{I}\left(\alpha, \frac{1}{Q^*}, \theta^*\right) = \frac{1}{Q^*} \int \frac{\dot{q}(x)\dot{q}(x)^\top}{q^*(x)} dF(x) + \frac{\alpha}{(1-\alpha)(Q^*)^2} \dot{Q} \dot{Q}^\top. \quad (63)$$

The second term is positive semidefinite. On the set

$$\mathcal{S} = \{x : q^*(x) \geq c_q, \ g(p - x^\top \theta^*) \geq c_g\},$$

we have

$$\dot{q}(x) = g(p - x^\top \theta^*)x.$$

Since $q^*(x) \leq 1$,

$$\begin{aligned} \frac{1}{Q^*} \int \frac{\dot{q}(x)\dot{q}(x)^\top}{q^*(x)} dF(x) &\succeq \frac{c_g^2}{Q^*} \mathbb{E}[XX^\top \mathbf{1}\{X \in \mathcal{S}\}] \\ &\succeq \frac{c_g^2 \kappa}{Q^*} I_d \end{aligned}$$

by [Assumption 2](#).

The map in (62) is continuous in (α, v, θ) on any compact set on which $1 - \alpha v Q_\theta$ is bounded away from zero. Therefore, by choosing $\eta > 0$, $\rho > 0$, and $r_0 > 0$ sufficiently small, there exists $c_I > 0$ such that

$$\mathcal{I}(\alpha, v, \theta) \succeq 2c_I I_d \quad (64)$$

for every

$$0 \leq \alpha \leq \rho Q^*, \quad \frac{1-\eta}{Q^*} \leq v \leq \frac{1+\eta}{Q^*}, \quad \theta \in B(\theta^*, r_0).$$

On $\mathcal{E}_W$, the realized value $v = W_m/\alpha$ belongs to this range.

It remains to control deviations of the empirical Hessian from its conditional expectation. Conditional on W_m , the only random part of the fixed-price Hessian is

$$-\frac{1}{m} \sum_{j=1}^m \nabla^2 \log q_\theta(X_j^B).$$

On $B(\theta^*, r_0)$, [Assumption 3](#) and [Assumption 1](#) imply

$$\sup_{\theta \in B(\theta^*, r_0)} \|\nabla^2 \log q_\theta(X)\|_{\text{op}} \leq C$$

almost surely under F^B . The third derivatives are also uniformly bounded:

$$\sup_{\theta \in B(\theta^*, r_0)} \|\nabla^3 \log q_\theta(X)\| \leq C.$$

Choose a finite ε_H -net $\{\vartheta_1, \dots, \vartheta_M\}$ of $B(\theta^*, r_0)$, where $\varepsilon_H > 0$ is fixed and sufficiently small. At each net point, conditional matrix Bernstein gives

$$\mathbb{P}\left(\left\|-\nabla^2 \bar{\ell}_m^0(\vartheta_s) - \mathcal{I}(\alpha, v, \vartheta_s)\right\|_{\text{op}} > \frac{c_I}{2} \mid W_m\right) \leq 2d e^{-cm}.$$

The net cardinality M is independent of m and n . A union bound therefore gives simultaneous control at every net point with conditional probability at least $1 - Ce^{-cm}$.

The uniform third-derivative bound implies

$$\|\nabla^2 \bar{\ell}_m^0(\theta) - \nabla^2 \bar{\ell}_m^0(\vartheta_s)\|_{\text{op}} \leq C\|\theta - \vartheta_s\|$$

and the same Lipschitz bound holds for $\mathcal{I}(\alpha, v, \theta)$. Choosing ε_H small enough extends the net bound to the entire localization ball. Combining this extension with (64) yields

$$-\nabla^2 \bar{\ell}_m^0(\theta) \succeq c_I I_d, \quad \theta \in B(\theta^*, r_0),$$

with probability at least $1 - Ce^{-cm}$ on $\mathcal{E}_W$.

Finally, the screening perturbation bound (59) with $r = 2$ gives

$$\sup_{\theta \in B(\theta^*, r_0)} \|\nabla^2 \bar{\ell}_m(\theta) - \nabla^2 \bar{\ell}_m^0(\theta)\|_{\text{op}} \leq C \frac{\log n}{m}.$$

Since $m \geq C_0(\log n)^2$, the right-hand side is smaller than $c_I/2$ for all sufficiently large C_0 and n . Therefore, with probability at least $1 - Ce^{-cm}$,

$$-\nabla^2 \bar{\ell}_m(\theta) \succeq c_H I_d, \quad \theta \in B(\theta^*, r_0), \quad (65)$$

where $c_H = c_I/2$.

Step 7: Parameter concentration. On the intersection of the localization event (58) and the curvature event (65), every global maximizer lies in $B(\theta^*, r_0) \subset \text{int}(\Theta)$. Hence

$$\nabla \bar{\ell}_m(\hat{\theta}_m) = 0.$$

By the fundamental theorem of calculus,

$$0 = \nabla \bar{\ell}_m(\theta^*) + \left[\int_0^1 \nabla^2 \bar{\ell}_m(\theta^* + s(\hat{\theta}_m - \theta^*)) ds \right] (\hat{\theta}_m - \theta^*).$$

The segment joining θ^* and $\hat{\theta}_m$ lies inside the localization ball. By (65), the negative of the integrated Hessian is positive definite with minimum eigenvalue at least c_H . Therefore,

$$c_H \|\hat{\theta}_m - \theta^*\| \leq \|\nabla \bar{\ell}_m(\theta^*)\|.$$

Combining this inequality with (61) and adding the localization and curvature failure probabilities proves (47).

Step 8: Mean-square error. Let

$$D_\Theta = \sup_{\theta, \vartheta \in \Theta} \|\theta - \vartheta\| < \infty.$$

The tail bound implies that, for $1 \leq u \leq c'm$,

$$\mathbb{P}\left(m\|\hat{\theta}_m - \theta^*\|^2 > C_\theta^2 u\right) \leq Ce^{-cu} + Ce^{-cm}.$$

Using the tail-integral identity,

$$\mathbb{E}\|\hat{\theta}_m - \theta^*\|^2 = \int_0^{D_\Theta^2} \mathbb{P}\left(\|\hat{\theta}_m - \theta^*\|^2 > s\right) ds,$$

we split the integral at C_θ^2/m and at $C_\theta^2 c'$. On the intermediate interval, substitute $u = ms/C_\theta^2$ and integrate the exponential tail. On the remaining interval, use the exponentially small localization and curvature failure probability. This gives

$$\mathbb{E}\|\hat{\theta}_m - \theta^*\|^2 \leq \frac{C}{m} + Ce^{-cm},$$

which proves (48). $\square$

D Continuation Stability

In this section, we establish the stability properties of the residual continuation problem used in Proposition 2. For notational convenience, let

$$F_\tau(\Delta, \theta) = R_{i_0}^{\text{sup}}(\Delta, \theta \mid \mathcal{H}_\tau),$$

and let

$$\Delta^*(\theta) \in \arg \max_{\Delta \in \mathcal{D}_{i_0}(T-\tau)} F_\tau(\Delta, \theta)$$

denote an optimal suffix-duration vector. If $i_0 = k$, the feasible set contains a single point, so all conclusions follow immediately. We therefore focus on the nontrivial case $i_0 < k$.

Because

$$\pi_{i_0}(\theta^*) = Q^* > 0,$$

continuity of the segment probabilities implies that there exists a neighborhood U_1 of θ^* such that

$$\pi_{i_0}(\theta) \geq \frac{Q^*}{2}, \quad \theta \in U_1.$$

Moreover,

$$H_{i_0}(\tau) \leq T \quad \text{and} \quad Q_\tau(\theta) \leq 1.$$

If the residual pool satisfies $N_\tau \geq \gamma n$, then

$$\begin{aligned} N_\tau \tilde{\pi}_{i_0}(\theta \mid \mathcal{H}_\tau) &= N_\tau \frac{\pi_{i_0}(\theta) e^{-\lambda H_{i_0}(\tau)}}{Q_\tau(\theta)} \\ &\geq \frac{\gamma Q^* e^{-\lambda T}}{2} n =: c_{\text{res}} n. \end{aligned} \tag{66}$$

Thus, uniformly over $\theta \in U_1$, a positive fraction of the residual pool belongs to segment i_0 .

Lemma 7 (Strong concavity). *Under the conditions of Proposition 2, there exists $\mu > 0$ such that, for every $\theta \in U_1$, every $\Delta \in \mathcal{D}_{i_0}(T - \tau)$, and every feasible direction u satisfying*

$$\sum_{j=i_0}^k u_j = 0,$$

we have

$$u^\top \nabla_{\Delta\Delta}^2 F_\tau(\Delta, \theta) u \leq -\mu n \|u\|^2.$$

Proof. A customer in original segment i_0 is willing to purchase from the beginning of the suffix problem. Define

$$S_j(\Delta) = \sum_{\ell=i_0}^j \Delta_\ell, \quad C_j(u) = \sum_{\ell=i_0}^j u_\ell.$$

The telescoping representation of the segment revenue gives

$$r_{i_0, i_0}^{\text{sup}}(\Delta) = p_{i_0} - \sum_{j=i_0}^{k-1} (p_j - p_{j+1}) e^{-\lambda S_j(\Delta)} - p_k e^{-\lambda S_k(\Delta)}.$$

On the affine hull of $\mathcal{D}_{i_0}(T - \tau)$,

$$S_k(\Delta) = T - \tau$$

is fixed. Consequently, every feasible direction satisfies $C_k(u) = 0$, and

$$u^\top \nabla_{\Delta\Delta}^2 r_{i_0, i_0}^{\text{sup}}(\Delta) u = -\lambda^2 \sum_{j=i_0}^{k-1} (p_j - p_{j+1}) e^{-\lambda S_j(\Delta)} C_j(u)^2.$$

The linear map

$$u \mapsto (C_{i_0}(u), \dots, C_{k-1}(u))$$

is a bijection from the tangent space

$$\left\{ u \in \mathbb{R}^{k-i_0+1} : \sum_{j=i_0}^k u_j = 0 \right\}$$

to $\mathbb{R}^{k-i_0}$. Because the space is finite dimensional, norm equivalence implies that there exists $c_C > 0$ such that

$$\sum_{j=i_0}^{k-1} C_j(u)^2 \geq c_C \|u\|^2$$

for every feasible direction u . In addition,

$$S_j(\Delta) \leq T$$

and

$$p_j - p_{j+1} > 0$$

for every $j < k$. Therefore, there exists $c_r > 0$, independent of Δ , such that

$$u^\top \nabla_{\Delta\Delta}^2 r_{i_0, i_0}^{\text{sup}}(\Delta) u \leq -c_r \|u\|^2.$$

The segment- i_0 term appears in F_τ with weight

$$N_\tau \tilde{\pi}_{i_0}(\theta \mid \mathcal{H}_\tau) \geq c_{\text{res}} n$$

by (66). Every other segment-revenue contribution is concave in Δ and therefore contributes a negative-semidefinite Hessian. Hence

$$u^\top \nabla_{\Delta\Delta}^2 F_\tau(\Delta, \theta) u \leq -c_{\text{res}} c_r n \|u\|^2.$$

The result follows with

$$\mu = c_{\text{res}} c_r.$$

□

Strong concavity ensures uniqueness of the continuation markdown schedule. The next lemma shows that this optimizer changes smoothly with the estimated parameter.

Lemma 8 (Optimal solution stability). *Under the conditions of Lemma 7, the optimal solution $\Delta^*(\theta)$ is unique for every $\theta \in U_1$. Moreover, there exists $C_\Delta > 0$ such that*

$$\|\Delta^*(\theta) - \Delta^*(\theta')\| \leq C_\Delta \|\theta - \theta'\|, \quad \theta, \theta' \in U_1.$$

Proof. Uniqueness follows immediately from strong concavity on the affine hull of the feasible simplex. Let

$$x = \Delta^*(\theta), \quad y = \Delta^*(\theta').$$

Because the feasible set is convex, the first-order variational inequalities for the two maximization problems give

$$\langle \nabla_\Delta F_\tau(x, \theta), y - x \rangle \leq 0$$

and

$$\langle \nabla_\Delta F_\tau(y, \theta'), x - y \rangle \leq 0.$$

Adding the two inequalities and inserting $\nabla_\Delta F_\tau(y, \theta)$ yields

$$\langle \nabla_\Delta F_\tau(x, \theta) - \nabla_\Delta F_\tau(y, \theta), y - x \rangle \leq \langle \nabla_\Delta F_\tau(y, \theta') - \nabla_\Delta F_\tau(y, \theta), y - x \rangle.$$

Strong concavity from Lemma 7 implies that the left-hand side is at least

$$\mu n \|x - y\|^2.$$

Therefore,

$$\mu n \|x - y\|^2 \leq \|\nabla_\Delta F_\tau(y, \theta) - \nabla_\Delta F_\tau(y, \theta')\| \|x - y\|.$$

It remains to bound the variation of the gradient with respect to θ . The residual survival probability is uniformly bounded away from zero. Indeed,

$$Q_\tau(\theta) = \pi_{k+1}(\theta) + \sum_{h=1}^k \pi_h(\theta) e^{-\lambda H_h(\tau)} \geq e^{-\lambda T}.$$

Together with Assumption 1, this implies that, after shrinking U_1 if necessary,

$$\sup_{\substack{\Delta \in \mathcal{D}_{i_0}(T-\tau) \\ \vartheta \in U_1}} \|\nabla_{\Delta\theta}^2 F_\tau(\Delta, \vartheta)\| \leq Ln$$

for some constant $L > 0$. The mean-value theorem therefore gives

$$\|\nabla_\Delta F_\tau(y, \theta) - \nabla_\Delta F_\tau(y, \theta')\| \leq Ln \|\theta - \theta'\|.$$

If $x \neq y$, dividing by $\mu n \|x - y\|$ yields

$$\|x - y\| \leq \frac{L}{\mu} \|\theta - \theta'\|.$$

The same inequality is immediate when $x = y$. Thus the claim holds with

$$C_\Delta = \frac{L}{\mu}.$$

□

The Lipschitz continuity of the optimal solution, together with first-order optimality, converts parameter error into a second-order revenue loss.

Lemma 9 (Second-order continuation loss). *There exists $C > 0$ such that, for every $\theta \in U_1$,*

$$F_\tau(\Delta^*(\theta^*), \theta^*) - F_\tau(\Delta^*(\theta), \theta^*) \leq Cn\|\theta - \theta^*\|^2.$$

Proof. Let

$$x = \Delta^*(\theta^*), \quad y = \Delta^*(\theta).$$

By concavity of $F_\tau(\cdot, \theta^*)$,

$$F_\tau(x, \theta^*) - F_\tau(y, \theta^*) \leq \langle \nabla_\Delta F_\tau(y, \theta^*), x - y \rangle.$$

Because y maximizes $F_\tau(\cdot, \theta)$ over the same convex feasible set,

$$\langle \nabla_\Delta F_\tau(y, \theta), x - y \rangle \leq 0.$$

Consequently,

$$\begin{aligned} F_\tau(x, \theta^*) - F_\tau(y, \theta^*) &\leq \langle \nabla_\Delta F_\tau(y, \theta^*) - \nabla_\Delta F_\tau(y, \theta), x - y \rangle \\ &\leq Ln\|\theta - \theta^*\|\|x - y\|. \end{aligned}$$

Applying the optimizer-stability bound from [Lemma 8](#) gives

$$F_\tau(x, \theta^*) - F_\tau(y, \theta^*) \leq LC_\Delta n\|\theta - \theta^*\|^2.$$

The argument relies only on the variational inequality and therefore does not require either optimizer to lie in the interior of the simplex. □

Proof (Proof of [Proposition 2](#)). Part (i) follows from [Lemma 7](#). Part (ii) follows from [Lemma 8](#), and part (iii) follows from [Lemma 9](#). When $i_0 = k$, the feasible suffix set is a singleton, so all three conclusions hold trivially. □

E Full Regret Proof

Proof (Proof of [Theorem 2](#)). We decompose regret into the loss incurred before estimation is completed and the loss arising from optimizing the continuation schedule under the estimated parameter.

Let $\mathcal{E}$ denote the event that the screening procedure stops at p_{i_0} , the policy collects m_0 transactions by time

$$s_n = kh_n + C_\tau \frac{m_0}{n},$$

and the estimator $\hat{\theta}_{\tau_0}$ belongs to the neighborhood U_1 in [Proposition 2](#). By [Proposition 1](#), after choosing the constants in the policy sufficiently large,

$$\mathbb{P}(\mathcal{E}^c) \leq n^{-4} + Ce^{-cm_0}.$$

Fix a population-optimal duration vector

$$\Delta^* \in \arg \max_{\Delta \in \mathcal{D}} R(\Delta, \theta^*)$$

that assigns zero duration to every inactive price above p_{i_0} . Such a choice is without loss of optimality because those prices cannot generate a transaction under θ^* .

To isolate the cost of screening and data collection, we construct an auxiliary comparator as follows. The comparator follows the learning policy through time τ_0 and, from that time onward, follows the calendar-time tail of the population-optimal schedule Δ^* . Couple the learning policy, the

comparator, and the population benchmark using the same customer valuations and price-checking processes.

On the event $\mathcal{E}$, we have $\tau_0 \leq s_n$. Any customer who does not check the price during $[0, s_n]$ remains unpurchased in both the population-benchmark and comparator systems and is therefore in the same state at time s_n . Because the two systems follow the same calendar-time price path thereafter, only customers who check during $[0, s_n]$ can contribute to the revenue difference accumulated before the continuation problem begins. Each customer contributes at most p_1 , so

$$\begin{aligned}\mathbb{E}[\text{Loss}_{\text{pre}} \mathbf{1}_{\mathcal{E}}] &\leq np_1(1 - e^{-\lambda s_n}) \\ &\leq np_1 \lambda s_n = O(k \log n + m_0).\end{aligned}\tag{67}$$

On the complement $\mathcal{E}^c$, the total revenue difference is bounded by np_1 . Hence

$$\mathbb{E}[\text{Loss}_{\text{pre}} \mathbf{1}_{\mathcal{E}^c}] \leq np_1 \mathbb{P}(\mathcal{E}^c) = o(1).$$

We next control the continuation loss. Conditional on $\mathcal{H}_{\tau_0}$ and $\mathcal{E}$, let

$$\Delta_{i_0}^*(\theta^* \mid \mathcal{H}_{\tau_0})$$

denote the true optimal suffix schedule for the residual pool. Its value is at least that of the comparator's continuation schedule, since the latter is feasible for the same residual problem. The learning policy, in contrast, implements

$$\Delta_{i_0}^*(\hat{\theta}_{\tau_0} \mid \mathcal{H}_{\tau_0}).$$

Therefore, [Proposition 2](#) gives

$$\begin{aligned}\text{Loss}_{\text{post}} &\leq R_{i_0}^{\text{sup}}(\Delta_{i_0}^*(\theta^* \mid \mathcal{H}_{\tau_0}), \theta^* \mid \mathcal{H}_{\tau_0}) - R_{i_0}^{\text{sup}}(\Delta_{i_0}^*(\hat{\theta}_{\tau_0} \mid \mathcal{H}_{\tau_0}), \theta^* \mid \mathcal{H}_{\tau_0}) \\ &\leq Cn \|\hat{\theta}_{\tau_0} - \theta^*\|^2\end{aligned}\tag{68}$$

on $\mathcal{E}$. Taking expectations and applying [\(26\)](#) yields

$$\mathbb{E}[\text{Loss}_{\text{post}} \mathbf{1}_{\mathcal{E}}] = O\left(\frac{n}{m_0}\right).$$

The contribution from $\mathcal{E}^c$ is again bounded by

$$np_1 \mathbb{P}(\mathcal{E}^c) = o(1).$$

Combining the pre-estimation and post-estimation losses, we obtain

$$\text{Regret}_n^{\text{pop}}(II, \theta^*) \leq C_1(k \log n + m_0) + C_2 \frac{n}{m_0} + O(1).$$

With

$$m_0 = \lceil c_m \sqrt{n} \rceil,$$

the two leading terms are both $O(\sqrt{n})$, while $k \log n = o(\sqrt{n})$. Therefore,

$$\text{Regret}_n^{\text{pop}}(II, \theta^*) = O(\sqrt{n}).$$

Finally, the population-pool comparison in [\(36\)](#) gives

$$\begin{aligned}\text{Regret}_n^{\text{H}}(II, \theta^*) &= \text{OPT}_n^{\text{H}}(\theta^*) - J_n(II, \theta^*) \\ &\leq \text{OPT}(\theta^*) - J_n(II, \theta^*) + \frac{kp_1}{2} \sqrt{n} \\ &= \text{Regret}_n^{\text{pop}}(II, \theta^*) + \frac{kp_1}{2} \sqrt{n}.\end{aligned}$$

Hence,

$$\text{Regret}_n^{\text{H}}(II, \theta^*) = O(\sqrt{n}),$$

which completes the proof. $\square$

F Instance-Dependent Free Learning

Proof (Proof of [Theorem 3](#)). Fix $\theta \in \Theta_{\text{free}}$, and choose a population-optimal schedule that assigns at least $\underline{\Delta}$ units of time to the first active price $p_{i_0(\theta)}$. For brevity, let $i_0 = i_0(\theta)$.

By Lemma 2, the policy reaches p_{i_0} and correctly identifies it as the first active price with probability at least $1 - O(n^{-4})$. The total time spent screening inactive prices is at most

$$i_0 h_n \leq k h_n \leq \frac{\underline{\Delta}}{2}.$$

Once the policy reaches p_{i_0} , consider the transactions generated during the next $\underline{\Delta}/2$ units of time. Since

$$A_{i_0}(\theta) \geq a_0,$$

each customer purchases during this interval with probability at least

$$a_0(1 - e^{-\lambda \underline{\Delta}/2}).$$

Accordingly, the number of purchases stochastically dominates

$$\text{Binomial}\left(n, a_0(1 - e^{-\lambda \underline{\Delta}/2})\right).$$

Because

$$m_0 = \lceil \gamma n \rceil$$

and

$$\gamma < \frac{a_0}{2}(1 - e^{-\lambda \underline{\Delta}/2}),$$

a Chernoff bound implies that the policy collects m_0 transactions within this interval except with probability at most e^{-cn} .

Combining the screening and collection bounds, with probability $1 - O(n^{-4}) - e^{-cn}$, estimation is completed no later than

$$k h_n + \frac{\underline{\Delta}}{2} \leq \underline{\Delta}.$$

Thus, the policy completes data collection before the collecting-price interval of the selected optimal schedule ends.

We now compare the learning policy with this benchmark under a common coupling of customer valuations and checking processes. The only first-order deviation before estimation arises from the time spent screening inactive prices. After the policy reaches p_{i_0} , it posts the same price as the benchmark throughout the data-collection phase. Because each customer contributes at most p_1 , the expected screening loss is bounded by

$$n p_1 (1 - e^{-\lambda k h_n}) \leq n p_1 \lambda k h_n = O(n h_n) = O(\log n).$$

Since $m_0 = \Theta(n)$, Proposition 1 gives

$$\mathbb{E} \|\hat{\theta} - \theta^*\|^2 = O\left(\frac{1}{n}\right).$$

Applying the second-order continuation bound in Proposition 2 therefore yields

$$O\left(n \mathbb{E} \|\hat{\theta} - \theta^*\|^2\right) = O(1)$$

post-estimation loss. The screening and collection failure events contribute only

$$O(n)(O(n^{-4}) + e^{-cn}) = o(1).$$

Combining these terms gives

$$\text{Regret}_n^{\text{POP}}(\Pi, \theta) = O(\log n)$$

uniformly over Θ_{free} .

Finally, if $i_0(\theta) = 1$, the policy begins immediately at the collecting price, so there is no screening loss. In that case, both the estimation loss and the continuation loss are $O(1)$, and hence

$$\text{Regret}_n^{\text{POP}}(\Pi, \theta) = O(1).$$

□

G Local Exploration-Estimation Lower Bound

This section proves Proposition 3. We consider a one-dimensional submodel with $d = 1$ and $X \equiv 1$. Choose a smooth distribution G with bounded support and a neighborhood of θ_0 such that every

customer is willing to purchase at p_2 and

$$g(p_1 - \theta) > 0$$

throughout the neighborhood. Then

$$\alpha(\theta) = 1 - G(p_1 - \theta)$$

is continuously differentiable and locally one-to-one. Because the feature is constant and every customer accepts the lower price, the distribution of the observed history depends on θ only through the fraction $\alpha(\theta)$ of customers willing to purchase at p_1 . It is therefore sufficient to index the experiment directly by α .

A markdown policy begins at p_1 , chooses a possibly randomized stopping time $\tau \in [0, T]$ adapted to the high-price transaction history, and posts p_2 permanently after τ . If the high price is posted for a deterministic duration Δ , the expected revenue per customer is given by (32). Differentiating with respect to Δ gives

$$\partial_{\Delta} r(\Delta, \alpha) = \lambda \alpha (p_1 - p_2) e^{-\lambda \Delta} - \lambda (1 - \alpha) p_2 e^{-\lambda(T-\Delta)}, \quad (69)$$

$$\partial_{\Delta} \partial_{\Delta} r(\Delta, \alpha) = -\lambda^2 \alpha (p_1 - p_2) e^{-\lambda \Delta} - \lambda^2 (1 - \alpha) p_2 e^{-\lambda(T-\Delta)} < 0. \quad (70)$$

Hence $r(\cdot, \alpha)$ is strictly concave. The value α_0 defined in (33) satisfies

$$\partial_{\Delta} r(0, \alpha_0) = 0,$$

so it marks the boundary at which the high price first becomes profitable.

To analyze adaptive policies, consider the counterfactual process in which p_1 is posted throughout the entire horizon. Let N_t denote the number of transactions observed by time t under this process, and let

$$M_t = n - N_t$$

be the number of customers who remain in the pool. The actual policy and the counterfactual process coincide up to the switching time τ .

Conditional on remaining in the pool at time t , a customer is a high type with probability

$$h_{\alpha}(t) = \frac{\alpha e^{-\lambda t}}{D_{\alpha}(t)}, \quad D_{\alpha}(t) = 1 - \alpha + \alpha e^{-\lambda t}. \quad (71)$$

Therefore, before the markdown occurs, the predictable intensity of the transaction-counting process is

$$\nu_{\alpha}(t) = \lambda M_t h_{\alpha}(t). \quad (72)$$

If the seller marks down at time t , the expected continuation revenue from one remaining customer is

$$s(t) = p_2 (1 - e^{-\lambda(T-t)}).$$

Define the instantaneous value of continuing at the high price by

$$g_{\alpha}(t) = \lambda h_{\alpha}(t) (p_1 - s(t)) + s'(t). \quad (73)$$

A direct calculation yields the useful identity

$$D_{\alpha}(t) g_{\alpha}(t) = \partial_{\Delta} r(t, \alpha). \quad (74)$$

Thus, the sign of $g_{\alpha}(t)$ coincides with the marginal value of extending the deterministic high-price duration.

Lemma 10 (Revenue identity and deterministic optimality). *For every randomized, history-dependent markdown policy Π ,*

$$J_n(\Pi, \alpha) = ns(0) + \mathbb{E}_{\alpha} \left[\int_0^{\tau} M_t g_{\alpha}(t) dt \right]. \quad (75)$$

Consequently, if

$$\Delta^*(\alpha) \in \arg \max_{\Delta \in [0, T]} r(\Delta, \alpha),$$

then

$$J_n(\Pi, \alpha) \leq nr(\Delta^*(\alpha), \alpha).$$

Proof. Conditional on the stopped high-price history, the expected revenue earned after switching to p_2 is

$$M_\tau s(\tau).$$

Hence

$$J_n(\Pi, \alpha) = \mathbb{E}_\alpha [p_1 N_\tau + M_\tau s(\tau)].$$

Define

$$X_t = p_1 N_t + M_t s(t).$$

Since $M_t = n - N_t$, the semimartingale decomposition of X_t is

$$dX_t = M_{t-} s'(t) dt + (p_1 - s(t)) dN_t.$$

Using the compensator

$$\nu_\alpha(t) dt = \lambda M_{t-} h_\alpha(t) dt$$

from (72), and applying bounded optional sampling at τ , gives

$$\mathbb{E}_\alpha[X_\tau] = ns(0) + \mathbb{E}_\alpha \left[\int_0^\tau M_t \{s'(t) + \lambda h_\alpha(t)(p_1 - s(t))\} dt \right].$$

This is exactly (75).

Because $r(\cdot, \alpha)$ is strictly concave, $\partial_\Delta r(t, \alpha)$ is nonnegative for $t \leq \Delta^*(\alpha)$ and nonpositive for $t \geq \Delta^*(\alpha)$. By (74), the same sign pattern holds for $g_\alpha(t)$. Therefore,

$$J_n(\Pi, \alpha) - ns(0) \leq \mathbb{E}_\alpha \left[\int_0^{\Delta^*(\alpha)} M_t g_\alpha(t) dt \right].$$

Under the counterfactual process,

$$\mathbb{E}_\alpha[M_t] = nD_\alpha(t).$$

Using (74), the right-hand side becomes

$$\begin{aligned} \mathbb{E}_\alpha \left[\int_0^{\Delta^*(\alpha)} M_t g_\alpha(t) dt \right] &= n \int_0^{\Delta^*(\alpha)} D_\alpha(t) g_\alpha(t) dt \\ &= n \int_0^{\Delta^*(\alpha)} \partial_\Delta r(t, \alpha) dt \\ &= n [r(\Delta^*(\alpha), \alpha) - r(0, \alpha)]. \end{aligned}$$

Since $r(0, \alpha) = s(0)$, the desired upper bound follows. $\square$

We now consider two nearby instances,

$$\alpha_- = \alpha_0 - \delta, \quad \alpha_+ = \alpha_0 + \delta.$$

The following lemma describes the local geometry of the deterministic optimization problem around the boundary point α_0 .

Lemma 11 (Local geometry). *There exist constants $c_1, c_2, c_3, c_4, c_5, a, \delta_0 > 0$ such that, for every $\delta \in (0, \delta_0]$,*

$$\Delta^*(\alpha_-) = 0, \quad c_1 \delta \leq \Delta^*(\alpha_+) \leq c_2 \delta, \quad (76)$$

$$r(\Delta^*(\alpha_+), \alpha_+) - r(0, \alpha_+) \geq c_3 \delta^2. \quad (77)$$

Moreover,

$$\begin{aligned} g_{\alpha_-}(t) &\leq -c_4 \delta, & 0 \leq t \leq T, \\ g_{\alpha_+}(t) &\geq c_5 \delta, & 0 \leq t \leq a\delta. \end{aligned} \quad (78)$$

Proof. At the boundary point $(0, \alpha_0)$,

$$\partial_{\Delta\alpha} r(0, \alpha_0) = \lambda [(p_1 - p_2) + p_2 e^{-\lambda T}] > 0,$$

whereas

$$\partial_{\Delta\Delta} r(0, \alpha_0) < 0.$$

The implicit-function theorem therefore implies that the interior optimizer emerging on the α_+ side satisfies

$$\Delta^*(\alpha_+) = \Theta(\delta).$$

This gives the duration bounds in (76). A second-order Taylor expansion of $r(\Delta, \alpha_+)$ around $\Delta = 0$ then yields the value gain (77).

Under α_- ,

$$\partial_{\Delta} r(0, \alpha_-) = -c\delta$$

for some $c > 0$. Since $\partial_{\Delta\Delta} r < 0$, the derivative is decreasing in Δ , and hence

$$\partial_{\Delta} r(t, \alpha_-) \leq -c\delta \quad \text{for all } t \in [0, T].$$

Because $D_{\alpha_-}(t)$ is uniformly bounded above and away from zero, (74) gives the first inequality in (78).

Similarly,

$$\partial_{\Delta} r(0, \alpha_+) = c\delta.$$

The second derivative is uniformly bounded in a neighborhood of $(0, \alpha_0)$, so for a sufficiently small constant $a > 0$,

$$\partial_{\Delta} r(t, \alpha_+) \geq c'\delta \quad \text{for } 0 \leq t \leq a\delta.$$

Applying (74) again proves the second inequality in (78). $\square$

For a policy Π , define its expected at-risk exposure under parameter α by

$$A_{\alpha}(\Pi) = \mathbb{E}_{\alpha} \left[\int_0^{\tau} M_t dt \right]. \quad (79)$$

The next result relates this exposure to regret under the two local instances.

Lemma 12 (Economic cost of exposure). *For every policy Π and every sufficiently small δ ,*

$$\text{Regret}_n(\Pi, \alpha_-) \geq c\delta A_{\alpha_-}(\Pi). \quad (80)$$

Furthermore, with

$$Z_a = \int_0^{\tau \wedge a\delta} M_t dt,$$

we have

$$\text{Regret}_n(\Pi, \alpha_+) \geq cn\delta^2 - C\delta \mathbb{E}_{\alpha_+}[Z_a]. \quad (81)$$

Proof. Under α_- , the optimal high-price duration is zero. Combining the revenue identity (75) with the bound

$$g_{\alpha_-}(t) \leq -c_4\delta$$

from (78) gives

$$\text{Regret}_n(\Pi, \alpha_-) = ns(0) - J_n(\Pi, \alpha_-) \geq c_4\delta \mathbb{E}_{\alpha_-} \left[\int_0^{\tau} M_t dt \right],$$

which proves (80).

We next consider α_+ . Contributions after the deterministic optimizer $\Delta^*(\alpha_+)$ can only increase regret because

$$g_{\alpha_+}(t) \leq 0 \quad \text{for } t \geq \Delta^*(\alpha_+).$$

Choose a sufficiently small that

$$a\delta \leq \Delta^*(\alpha_+).$$

Using (75) and the lower bound on g_{α_+} over $[0, a\delta]$ gives

$$\text{Regret}_n(\Pi, \alpha_+) \geq c_5\delta \left\{ \int_0^{a\delta} \mathbb{E}_{\alpha_+}[M_t] dt - \mathbb{E}_{\alpha_+}[Z_a] \right\}.$$

For α in a fixed neighborhood of α_0 , $D_{\alpha}(t)$ is uniformly bounded away from zero. Therefore,

$$\mathbb{E}_{\alpha_+}[M_t] = nD_{\alpha_+}(t) \geq c_0n.$$

It follows that

$$\int_0^{a\delta} \mathbb{E}_{\alpha_+}[M_t] dt \geq c_0 a n \delta,$$

which proves (81). $\square$

The same at-risk exposure that generates regret under α_- also controls the amount of statistical information available for distinguishing the two instances.

Lemma 13 (Information accumulated before switching). *There exists a constant $C > 0$ such that*

$$\text{KL}(\mathbb{P}_{\alpha_-}^{\Pi}, \mathbb{P}_{\alpha_+}^{\Pi}) \leq C\delta^2 A_{\alpha_-}(\Pi). \quad (82)$$

Proof. Condition on the policy's internal randomization. Before the markdown, the transaction-counting intensities under the two instances are

$$\nu_{\pm}(t) = \lambda M_{t-} h_{\alpha_{\pm}}(t).$$

The relative-entropy formula for counting processes stopped at τ gives

$$\text{KL}(\mathbb{P}_{\alpha_-}^{\Pi}, \mathbb{P}_{\alpha_+}^{\Pi}) = \lambda \mathbb{E}_{\alpha_-} \int_0^{\tau} M_t \left[h_{\alpha_-}(t) \log \frac{h_{\alpha_-}(t)}{h_{\alpha_+}(t)} - h_{\alpha_-}(t) + h_{\alpha_+}(t) \right] dt.$$

For α in a sufficiently small neighborhood of α_0 , the posterior probability $h_{\alpha}(t)$ remains in a compact subset of $(0, 1)$ uniformly over $t \in [0, T]$. Moreover, it is uniformly Lipschitz in α . The inequality

$$x \log \frac{x}{y} - x + y \leq C(x - y)^2$$

for x and y in such a compact subset therefore implies

$$h_{\alpha_-}(t) \log \frac{h_{\alpha_-}(t)}{h_{\alpha_+}(t)} - h_{\alpha_-}(t) + h_{\alpha_+}(t) \leq C\delta^2.$$

Substituting this bound into the entropy formula gives

$$\text{KL}(\mathbb{P}_{\alpha_-}^{\Pi}, \mathbb{P}_{\alpha_+}^{\Pi}) \leq C\delta^2 \mathbb{E}_{\alpha_-} \left[\int_0^{\tau} M_t dt \right],$$

which is (82).

After the seller switches to p_2 , every remaining customer is willing to purchase and checks at the same rate under both parameters. Conditional on the stopped high-price history, the post-switch observations therefore have the same distribution under α_- and α_+ and contribute no additional relative entropy. $\square$

We can now complete the lower-bound argument.

Proof (Proof of Proposition 3). Define

$$H_{n,\delta} = \min\{n\delta, \delta^{-2}\},$$

and fix a sufficiently small constant $b > 0$. We distinguish two cases.

Suppose first that

$$A_{\alpha_-}(\Pi) \geq bH_{n,\delta}.$$

Then (80) gives

$$\begin{aligned} \text{Regret}_n(\Pi, \alpha_-) &\geq cb\delta H_{n,\delta} \\ &= cb \min\{n\delta^2, \delta^{-1}\}. \end{aligned}$$

Now suppose instead that

$$A_{\alpha_-}(\Pi) < bH_{n,\delta}.$$

By (82),

$$\text{KL}(\mathbb{P}_{\alpha_-}^{\Pi}, \mathbb{P}_{\alpha_+}^{\Pi}) \leq Cb.$$

Pinsker's inequality therefore gives

$$\left\| \mathbb{P}_{\alpha_-}^{\Pi} - \mathbb{P}_{\alpha_+}^{\Pi} \right\|_{\text{TV}} \leq C\sqrt{b} =: \eta_b.$$

Because

$$0 \leq Z_a \leq na\delta$$

and Z_a is measurable with respect to the stopped high-price history,

$$\begin{aligned} \mathbb{E}_{\alpha_+}[Z_a] &\leq \mathbb{E}_{\alpha_-}[Z_a] + na\delta \eta_b \\ &\leq A_{\alpha_-}(\Pi) + na\delta \eta_b \\ &\leq bH_{n,\delta} + na\delta \eta_b. \end{aligned}$$

Since

$$H_{n,\delta} \leq n\delta,$$

we obtain

$$\mathbb{E}_{\alpha_+}[Z_a] \leq n\delta(b + a\eta_b).$$

Substituting this estimate into (81) gives

$$\text{Regret}_n(\Pi, \alpha_+) \geq n\delta^2 [c - C(b + a\eta_b)].$$

Choosing b sufficiently small ensures that the bracket is bounded below by a positive constant. Hence

$$\text{Regret}_n(\Pi, \alpha_+) \geq c'n\delta^2 \geq c' \min\{n\delta^2, \delta^{-1}\}.$$

Combining the two cases yields

$$\sup_{\alpha \in \{\alpha_-, \alpha_+\}} \text{Regret}_n(\Pi, \alpha) \geq c \min\{n\delta^2, \delta^{-1}\},$$

which proves (34). Finally, taking

$$\delta_n = n^{-1/3}$$

balances the two terms and gives

$$\inf_{\Pi} \sup_{\alpha \in \{\alpha_0 - \delta_n, \alpha_0 + \delta_n\}} \text{Regret}_n(\Pi, \alpha) \geq cn^{1/3}.$$

□