

Minimax-Optimal Semiparametric Contextual Dynamic Pricing with Multimodal Revenue

Xueping Gong¹, Zhuoluo Zhang², Zhaowei Miao³, and Jiheng Zhang⁴

¹School of Management, Xiamen University, xgongah@xmu.edu.cn

²School of Management, Xiamen University, zhangzhuoluo@xmu.edu.cn

³School of Management, Xiamen University, miaozhaowei@xmu.edu.cn

⁴Department of Industrial Engineering and Decision Analytics, The Hong Kong University of Science and Technology, jiheng@ust.hk

Abstract

We study contextual dynamic pricing with arbitrary covariate sequences and bounded, possibly nonbinary purchase quantities. Demand follows a semiparametric surplus-index model with an unknown linear valuation parameter and an unknown Hölder-smooth response. We impose neither concavity nor strong unimodality on revenue and allow nonunique optimal prices. We develop a pilot-corrected layered decision-partitioning policy that combines directional pilot estimation, local polynomial learning, predictable data assignment, and global action elimination. Pilot correction removes the first-order effect of valuation-parameter error, while permanent labels enable concentration under adaptive sampling. The policy attains the minimax smoothness-dependent horizon rate up to logarithmic factors; a matching lower bound already holds for a constant-context binary-demand subclass.

Keywords: contextual dynamic pricing; semiparametric demand; bounded quantity feedback; shape-free revenue; minimax regret.

1 Introduction

Dynamic pricing is a central problem in revenue management. A seller must learn how demand responds to prices while simultaneously using the accumulated information to generate revenue. This learning problem becomes more challenging when customer, product, or market heterogeneity is observed through covariates. Although such information enables personalized pricing, every posted price also determines what the seller subsequently learns. This feedback creates the exploration–exploitation trade-off at the heart of contextual dynamic pricing; see Den Boer (2015) for a broad review and Saharan et al. (2020) for applications.

A widely studied framework models a customer’s latent valuation as a finite-dimensional function of the observed covariates plus an additive market shock. In semiparametric formulations, the contextual component is typically linear, whereas the distribution of the market shock is left unspecified (Javanmard and Nazerzadeh, 2019; Luo et al., 2022, 2024; Fan et al., 2024). This framework combines an interpretable representation of customer heterogeneity with a flexible demand curve. Binary purchase feedback dominates this single-index literature. Shape-free regret

guarantees are available for Lipschitz links, whereas the higher-order-smooth literature typically couples smoothness with distributional coverage and/or stable revenue geometry.

At Lipschitz smoothness, Tullii et al. (2024) allow arbitrary adaptive contexts and Gong et al. (2025) allow general valuation classes under independent and identically distributed contexts; both obtain the optimal shape-free two-thirds horizon exponent with binary feedback. For smoother links, recent improvements rely on nondegenerate or feature-diverse stochastic contexts together with stable revenue geometry (Han et al., 2026; Chai et al., 2026), or on a smooth oracle-price map induced by strong unimodality (Fan et al., 2026).

A particularly consequential restriction is strong unimodality (Wang and Chen, 2025; Han et al., 2026; Fan et al., 2026). In the form used by recent smooth-pricing analyses, this condition requires every contextual revenue function to have a unique interior maximizer and its revenue loss to be uniformly comparable to the squared distance from that maximizer. It is therefore considerably stronger than smoothness or a local second-order condition. Strong unimodality rules out separated revenue modes, flat optimal-price regions, and many asymmetric or boundary-optimal revenue landscapes. At the same time, it provides several powerful analytical advantages: the policy can safely localize around one estimated optimizer, price-estimation errors translate into quadratic revenue losses, and the context-dependent optimal price can often be represented by a regular oracle price map. Indeed, these properties are central to the algorithms and improved regret guarantees in Wang and Chen (2025), Han et al. (2026), and Fan et al. (2026).

This paper studies contextual pricing without these structural conveniences. In each period, the seller observes a covariate vector, posts a price, and observes a bounded purchase quantity, which may be binary, discrete, or continuous. Latent valuation is linear in the observed covariates and subject to an unknown market shock. After integrating out the shock, expected demand is an unknown function of the difference between price and a linear contextual valuation index; both the index parameter and the response function are unknown. The usual binary-purchase model is obtained as a special case, but the formulation also accommodates bounded unit sales and purchase volume. We allow the covariates to arrive arbitrarily and impose only Hölder smoothness on the observable response function. The induced revenue function may consequently be multimodal, may contain a flat optimal region, and need not have a unique maximizer. The smoothness assumption also has natural primitive foundations: it follows, for example, when the market-shock distribution is smooth and the quantity-response function has bounded variation, or when the quantity-response function is itself sufficiently smooth.

Removing strong unimodality fundamentally changes the learning problem. Hölder smoothness controls only local approximation and provides no information about the global geometry of revenue. Learning accurately near a provisional optimizer is therefore insufficient: another, statistically similar region may contain a better and well-separated revenue mode. A policy that localizes prematurely may never collect enough information to discover that region. Consequently, the algorithm must preserve and compare candidates across the entire price domain rather than organize its exploration around a single estimated optimum or oracle price map.

A second difficulty arises from the interaction between the unknown valuation parameter and the nonparametric response function. Directly inserting a pilot parameter estimate into a nonparametric regression transfers the pilot error to the fitted response curve at first order. This error can dominate the higher-order approximation accuracy that smoothness would otherwise provide. A natural remedy, used in classical single-index estimation and recent semiparametric pricing analyses, is to profile a nonparametric fit over candidate index parameters and refine the parameter through constrained least squares (Härdle et al., 1993; Ichimura, 1993; Horowitz and Härdle, 1996; Wang and

Chen, 2025; Han et al., 2026). This remedy is statistically and computationally demanding. Because the local fits and their underlying sample assignments vary with the candidate parameter, the resulting objective is generally nonconvex, even in the twice-smooth case. Standard local optimization methods thus do not provide a global-solution guarantee. Moreover, the same observations are used both to construct the nonparametric fit and to evaluate the least-squares criterion, creating a complex dependence structure in online settings and making finite-sample confidence analysis delicate (Han et al., 2026). These issues motivate an approach that exploits higher-order smoothness without repeatedly solving a nonconvex joint-estimation problem.

Adaptive data collection introduces a third challenge. Posted prices, residual bins, and subsequent exploration decisions all depend on previous outcomes. If past observations are reassigned whenever the valuation estimate changes, membership in a local dataset is determined retrospectively using information that was unavailable when the observation was collected. This destroys the predictable sampling structure needed for martingale concentration and makes standard offline nonparametric guarantees inapplicable.

We address these challenges through a pilot-corrected layered decision-partitioning (LDP) policy. The pilot module uses uniform-price exploration to estimate the linear valuation component and evaluates its uncertainty only along the currently observed covariate direction. A well-covered direction enters the main pricing stage immediately, whereas an insufficiently covered direction triggers additional pilot exploration. This directional mechanism avoids distributional assumptions on the covariate sequence.

During the main pricing stage, prices are represented relative to the estimated valuation index, and demand is learned locally in the resulting residual coordinate. An augmented local-polynomial representation absorbs the leading pilot perturbation into the regression coefficients, leaving only a second-order pilot error in addition to the usual nonparametric approximation error. Thus, the policy benefits from higher-order smoothness without solving the nonconvex constrained least-squares problems used in alternative joint estimation procedures. Each observation is also assigned a permanent layer-bin label before its demand is observed. These labels are never recomputed after the pilot estimate changes, preserving predictable sampling and enabling self-normalized concentration within every local dataset.

Finally, the layered policy maintains candidate actions over the entire residual domain. An action is eliminated only when its optimistic revenue is statistically separated from the best surviving benchmark. This design protects distant modes and flat near-optimal regions and therefore performs global rather than local revenue learning.

Balancing pilot exploration, local approximation, and statistical uncertainty yields the minimax smoothness-dependent horizon rate, up to logarithmic factors, for fixed problem primitives. We establish a matching expected-regret lower bound over an admissible constant-context, binary-demand subclass. The construction begins with a smooth instance whose revenue is flat over a nondegenerate interval and introduces statistically indistinguishable perturbations in separated price regions. Together, the upper and lower bounds characterize the minimax dependence on the horizon for the general revenue class considered here.

2 Contributions and Related Literature

Contributions. Our contributions are twofold.

- **A shape-free model and higher-order online method.** We combine arbitrary context

sequences, general Hölder smoothness, and bounded quantity feedback in a translated-residual single-index model without concavity, strong unimodality, or a unique optimal price. Our pilot-corrected LDP policy integrates directional exploration, higher-order local polynomials, predictable permanent labels, and global elimination. Relative to the Lipschitz LDP framework of Gong et al. (2025), this replaces an episodic offline pilot and piecewise-constant learning with an online directional pilot, first-order error correction, and higher-order residual learning.

- **Minimax-optimal regret for shape-free smooth links.** For fixed dimension and fixed problem primitives, the proposed policy achieves $\tilde{\mathcal{O}}\left(T^{\frac{\beta+1}{2\beta+1}}\right)$ regret, matching $\Omega\left(T^{\frac{\beta+1}{2\beta+1}}\right)$ lower bound. Together, the upper and lower bounds characterize the minimax horizon dependence for the general revenue class studied in this paper.

2.1 Related Literature

Semiparametric contextual dynamic pricing. Contextual pricing with a linear valuation component and an additive shock has been studied under known parametric noise distributions (Javanmard and Nazerzadeh, 2019; Ban and Keskin, 2021), unknown nonparametric noise (Luo et al., 2022, 2024; Fan et al., 2024; Choi et al., 2023; Xu and Wang, 2022; Chai et al., 2026), and more general valuation or utility classes (Chen et al., 2024; Gong et al., 2025). These works differ in the feedback available, the distributional assumptions on contexts, and the geometry imposed on revenue. The papers most closely related to ours are summarized in Table 1. The rates in the last column suppress logarithmic factors and polynomial dependence on fixed problem primitives.

Table 1: Representative results for smooth semiparametric contextual pricing. Here β denotes link or tail smoothness; rates suppress logarithmic factors and dependence on fixed problem primitives.

Work	Feedback	Contexts	Smoothness	Revenue geometry	Regret
Tullii et al. (2024)	Binary	Arbitrary	$\beta = 1$	General	$\tilde{\mathcal{O}}(T^{2/3})$
Gong et al. (2025)	Binary	i.i.d.	$\beta = 1$	General	$\tilde{\mathcal{O}}(T^{2/3})$
Wang and Chen (2025)	Binary	i.i.d.	$\beta = 2$	Strongly unimodal	$\tilde{\mathcal{O}}(T^{3/5})$
Han et al. (2026)	Binary	i.i.d.	$\beta \geq 2$	Strongly unimodal	$\tilde{\mathcal{O}}(T^{\frac{\beta+1}{2\beta+1}})$
Fan et al. (2026)	Binary	Arbitrary	$\beta \geq 2$	Strongly unimodal	$\tilde{\mathcal{O}}(T^{\frac{2\beta-1}{4\beta-3}} + \sqrt{T})$
This paper	Bounded quantity	Arbitrary	$\beta \geq 1$	General	$\tilde{\mathcal{O}}(T^{\frac{\beta+1}{2\beta+1}})$

The principal comparison at $\beta = 1$ is Tullii et al. (2024): it already allows arbitrary adaptive contexts and general revenue geometry, and attains the same $T^{2/3}$ horizon exponent, but is restricted to binary feedback and Lipschitz smoothness. Gong et al. (2025) introduce the LDP architecture for a shape-free Lipschitz problem with binary feedback and i.i.d. contexts. Thus, our claim at $\beta = 1$ is not a better horizon exponent. The contribution is to retain arbitrary contexts and shape-free revenue while covering general β and bounded quantity feedback.

For twice-smooth demand, Wang and Chen (2025) establish the sharp $\tilde{\mathcal{O}}(T^{3/5})$ regret rate under strong unimodality and nondegenerate i.i.d. contexts, using contextual successive elimination together with semiparametric estimation. Han et al. (2026) extend this line to general Hölder smoothness by combining local-polynomial estimation with the stationary learning subroutine of Wang and Chen (2025), while retaining similar revenue-geometry and context-coverage conditions. Fan et al. (2026) also impose strong unimodality but allow arbitrary contexts; by directly learning the resulting smooth oracle-price map, they obtain a faster horizon exponent when $\beta > 2$. Our

result addresses the complementary regime in which higher-order smoothness is available but the revenue geometry remains unrestricted. The flat-optimum construction underlying our lower bound shows that this distinction is structural: without strong unimodality, the hard instances need not admit a unique and stable oracle-price map on which localization can be based.

Xu and Wang (2026) study an adjacent, nonnested regime: binary feedback with possibly discontinuous demand, including jumps induced by atoms, under stochastic well-conditioned contexts. Their model relaxes link regularity, whereas ours uses Hölder smoothness to handle arbitrary contexts and bounded quantity feedback.

Algorithmically, our closest antecedent is the policy of Gong et al. (2025). Relative to its episodic offline pilot and piecewise-constant residual learning, our policy uses an online directional pilot, higher-order local polynomials, an augmented correction that makes pilot error second order, and permanent residual labels that preserve predictable sampling as the pilot evolves. Uniform-price exploration in Javanmard and Nazerzadeh (2019); Fan et al. (2024); Chen et al. (2024) and the arbitrary-context mechanisms in Tullii et al. (2024) provide additional points of contact.

Nonparametric and multimodal pricing. Without covariates, nonparametric dynamic pricing has been studied under a range of smoothness and shape conditions (Besbes and Zeevi, 2009, 2015; Wang et al., 2021). In particular, Wang et al. (2021) establish the minimax rate $\tilde{\mathcal{O}}(T^{(\beta+1)/(2\beta+1)})$ for smooth multimodal revenue using local-polynomial optimism. Their confidence indices explicitly incorporate a local approximation envelope, and thus require the corresponding smoothness radius.

Bounded quantity observations are not new by themselves. Wang et al. (2021) analyze general demand feedback without contexts, while Bu et al. (2022, 2025) study partially linear and separable contextual demand models. The present model instead places the unknown context effect inside the translated residual $p - \mathbf{x}^\top \boldsymbol{\theta}_*$. Consequently, the valuation shift and the nonparametric link must be learned jointly under an arbitrary context sequence. This translated single-index structure, rather than nonbinary feedback alone, is the relevant distinction.

Layered contextual bandits. Our analysis draws on standard confidence-based methods for contextual bandits (Auer, 2002; Chu et al., 2011; Abbasi-Yadkori et al., 2011; Lattimore and Szepesvári, 2020). Classical layered partitioning separates observations by precision level to obtain sharper confidence control. In the present problem, however, the local regression model itself depends on an evolving pilot index and is only approximately specified. The permanent labeling rule is therefore essential: it fixes the residual coordinate, local bin, and stopping layer used for each observation before the outcome is revealed. This converts the adaptively collected local samples into predictable martingale arrays while retaining global price exploration.

Organization. Section 3 introduces the model, the induced demand link, and its regularity. Section 4 presents the directional pilot, pilot-corrected local regression, and layered pricing policy. Section 5 establishes the regret upper bound and the matching minimax lower bound. The paper concludes with directions for future research.

3 Problem Formulation

Notation. Throughout the paper, we write $[n] := \{1, \dots, n\}$ for any positive integer n , denote the cardinality of a set A by $|A|$, and use $\mathbf{1}\{E\}$ for the indicator of an event E . For vectors, $\|\cdot\|_p$

denotes the standard ℓ_p norm for $1 \leq p \leq \infty$. The notation $\widetilde{O}$ suppresses absolute constants and logarithmic factors. For a positive definite matrix A , define $\|\mathbf{z}\|_A := \sqrt{\mathbf{z}^\top A \mathbf{z}}$. For an interval $I = [a, b]$, let $\Pi_I(z) := \min\{b, \max\{a, z\}\}$ denote projection onto I .

Basic Model. We consider a contextual dynamic pricing problem over a finite horizon T . In each period $t \in [T]$, a customer arrives and the seller observes a covariate vector $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$ that characterizes the observable customer, product, or market characteristics. Unlike standard i.i.d. assumptions, we do not impose any distributional assumption on the covariate sequence $\{\mathbf{x}_t\}_{t=1}^T$; it can be arbitrary and possibly dependent. The covariate space is bounded, i.e., there exists $C_x < \infty$ such that $\sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|_2 \leq C_x$. For each $t \in [T]$, let $\mathcal{F}_{t-1}$ denote the σ -field generated by all observations and policy randomization available up to the end of period $t-1$. Let U_t collect all within-period randomization used by the policy after observing $\mathbf{x}_t$. Conditional on $\mathcal{F}_{t-1} \vee \sigma(\mathbf{x}_t)$, the seed U_t is drawn independently of the current valuation shock.

The customer's valuation for the product is modeled as a linear function of the covariates plus random noise:

$$v_t = \mathbf{x}_t^\top \boldsymbol{\theta}_\star + \epsilon_t,$$

where $\boldsymbol{\theta}_\star \in \Theta = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_2 \leq C_\theta\}$ is an unknown parameter vector, and ϵ_t is an unobserved random shock. The shocks $\{\epsilon_t\}_{t=1}^T$ are independent and identically distributed according to an unknown cumulative distribution function F with zero mean. For every $t \in [T]$, ϵ_t is independent of $\mathcal{F}_{t-1} \vee \sigma(\mathbf{x}_t, U_t)$.

Assumption 3.1 (Bounded Valuations) *There exist $B > 0$ and $B_\epsilon \in (0, B/2)$ such that:*

- *The noise is uniformly bounded: $|\epsilon_t| \leq B_\epsilon$ almost surely.*
- *The linear valuation satisfies $\mathbf{x}^\top \boldsymbol{\theta}_\star \in [B_\epsilon, B - B_\epsilon]$ for all $\mathbf{x} \in \mathcal{X}$.*

Assumption 3.1 guarantees that $v_t = \mathbf{x}_t^\top \boldsymbol{\theta}_\star + \epsilon_t \in [0, B]$ almost surely, so the realized valuation and any feasible posted price lie in the same known interval $[0, B]$. This is a standard and natural requirement in practical pricing applications.

After observing $\mathbf{x}_t$, the seller posts a per-unit price $p_t \in [0, B]$. The price and every auxiliary action label constructed by the policy before demand is observed are measurable with respect to the full pre-demand information field $\mathcal{G}_t := \mathcal{F}_{t-1} \vee \sigma(\mathbf{x}_t, U_t)$. If the realized valuation satisfies $v_t \geq p_t$, a sale occurs and the seller observes a positive purchase quantity $y_t > 0$; otherwise, no sale occurs and $y_t = 0$, i.e., $\mathbf{1}\{y_t > 0\} = \mathbf{1}\{v_t \geq p_t\}$ almost surely. The observed quantity y_t may be discrete (e.g., number of units) or continuous (e.g., physical volume or usage), depending on the application. We assume there exists a constant $0 < D < \infty$ such that $0 \leq y_t \leq D$ almost surely. The realized revenue at time t is $p_t y_t$, and each round yields the observation triple $(\mathbf{x}_t, p_t, y_t)$. We set $\mathcal{F}_t := \mathcal{G}_t \vee \sigma(y_t)$, so $\mathcal{G}_t$ and $\mathcal{F}_t$ describe, respectively, all information immediately before and after current demand is observed.

Generalized Single Index Model. The basic model above only specifies the qualitative relationship between y_t and the surplus $v_t - p_t$. However, to enable statistical learning and regret minimization, the seller must quantify how the expected demand varies with the posted price and covariates. To make the problem tractable while retaining flexibility, we adopt a surplus-dependent single-index structure. This formulation generalizes the widely studied binary purchase model (Fan et al., 2026; Han et al., 2026; Luo et al., 2024; Fan et al., 2024; Gong et al., 2025) by allowing the

realized demand quantity to depend on the surplus through a function $q(\cdot)$, rather than restricting to a binary sale indicator.

Specifically, we posit that, conditional on the information available before demand is observed and on the realized valuation, the expected demand depends on the current price and covariates only through the net surplus $v_t - p_t$. This single-index form is natural: it implies that a customer's purchasing quantity decision is driven by the perceived value relative to the price, rather than by their absolute levels separately. Under this premise, there exists an unknown baseline function $q(\cdot)$ that maps the surplus to the expected quantity. Importantly, we impose no parametric form, such as a linear or logistic specification, on q , allowing it to accommodate richly heterogeneous and potentially nonmonotone demand responses. The formal statement is as follows.

Assumption 3.2 (Surplus-dependent demand response) *There exists an unknown measurable function $q : \mathbb{R} \rightarrow [0, D]$ with $q(z) = 0$ for $z < 0$ or $z > B$ and $q(z) > 0$ for $z \in [0, B]$, such that, for every $t \in [T]$,*

$$\mathbb{E}[y_t | \mathcal{G}_t \vee \sigma(v_t)] = q(v_t - p_t) \quad \text{almost surely.}$$

Assumption 3.2 formalizes the surplus-dependent single-index structure introduced above. It posits that, conditional on all pre-demand information and the realized valuation, the conditional mean demand is determined solely by the realized surplus $v_t - p_t$ and the functional form $q(\cdot)$ is left fully nonparametric. The zero restriction for negative surplus is consistent with the no-sale condition. The zero extension beyond B is only a technical convention, because such surplus cannot arise under Assumption 3.1. For an affordable purchase, the incidence condition implies positive realized quantity, and we choose the version of the conditional-mean function that is strictly positive on $[0, B]$. Crucially, we impose no monotonicity, concavity, or unimodality assumptions on q . This allows q to capture rich and potentially irregular demand patterns—such as quantity discounts, satiation effects, or Giffen-like behaviors at the micro level—that are often excluded by parametric specifications.

Define the effective residual domain $\mathcal{I}_g := [-B + B_\epsilon, B - B_\epsilon]$. The function q characterizes the expected demand conditional on the realized valuation $v_t = s$, which is unobservable to the seller. For decision-making and regret analysis, however, we require the expected demand conditional on the information available before the current demand is observed. To bridge this gap, we integrate out the noise ϵ_t from the surplus $s - p_t = \boldsymbol{\theta}_\star^\top \mathbf{x}_t + \epsilon_t - p_t$. This yields an induced link function g , defined as the convolution of q with the noise distribution F :

$$g(u) := \mathbb{E}[q(\epsilon_t - u)], \quad u \in \mathcal{I}_g,$$

where $u = p - \boldsymbol{\theta}_\star^\top \mathbf{x}$. Then, by iterated expectations and the sequential exogeneity of ϵ_t , we obtain,

$$\mathbb{E}[y_t | \mathcal{G}_t] = g(p_t - \mathbf{x}_t^\top \boldsymbol{\theta}_\star).$$

In this way, g serves as the demand link that replaces the latent q , and all subsequent estimation and regret guarantees will be stated in terms of g .

Revenue and Regret. Given covariate $\mathbf{x}$ and price p , the expected revenue is $\text{Rev}(\mathbf{x}, p) = p g(p - \boldsymbol{\theta}_\star^\top \mathbf{x})$. Let $p^\star(\mathbf{x}) \in \arg \max_{p \in [0, B]} \text{Rev}(\mathbf{x}, p)$ be an arbitrary maximizer, and write $p_t^\star := p^\star(\mathbf{x}_t)$ for period t . The cumulative regret is

$$\text{Reg}(T) = \sum_{t=1}^T (\text{Rev}(\mathbf{x}_t, p_t^\star) - \text{Rev}(\mathbf{x}_t, p_t)).$$

The goal is to design a nonanticipating pricing policy that, for each t , selects p_t based on the current covariate $\mathbf{x}_t$ and all past observations $\{(\mathbf{x}_s, p_s, y_s)\}_{s=1}^{t-1}$, so as to minimize $\text{Reg}(T)$ while learning the unknown valuation parameter $\boldsymbol{\theta}_*$ and the induced demand link g .

Regularity. To enable nonparametric estimation, we directly assume the observable induced link function g belongs to a Hölder class. This direct approach is standard in the dynamic pricing literature (Bu et al., 2025; Wang et al., 2021) and facilitates transparent convergence analysis, since g is the direct object of estimation.

Definition 3.3 (Hölder class) Let $\mathcal{I} \subset \mathbb{R}$ be a compact interval, $\beta \geq 1$, and $0 < L < \infty$. Define

$$\varpi(\beta) := \max\{k \in \mathbb{Z}_{\geq 0} : k < \beta\}.$$

A function $f : \mathcal{I} \rightarrow \mathbb{R}$ belongs to $\mathcal{H}(\beta, L; \mathcal{I})$ if f is $\varpi(\beta)$ -times differentiable on $\mathcal{I}$ and

$$\left| f^{(\varpi(\beta))}(u) - f^{(\varpi(\beta))}(u') \right| \leq L |u - u'|^{\beta - \varpi(\beta)}, \quad \forall u, u' \in \mathcal{I}.$$

Assumption 3.4 (Smoothness of the Induced Link) The induced link function satisfies $g \in \mathcal{H}(\beta, L_g; \mathcal{I}_g)$ for some $\beta \geq 1$ and $0 < L_g < \infty$.

For the contextual policy, the smoothness parameters β and L_g are treated as known. Throughout this paper, C_g denotes a fixed, class-level upper envelope for the derivatives of g up to order $\varpi(\beta)$. Because $0 \leq g \leq D$ on the interval $\mathcal{I}_g$, a standard one-dimensional interpolation inequality allows C_g to be chosen as a function only of B, B_ϵ, D, β , and L_g . Thus, the constants used by the policy do not depend on unknown instance-specific derivative values.

When the domain $\mathcal{I}$ is clear from the context, we write $\mathcal{H}(\beta)$ for simplicity.

Justification of Assumption 3.4. A natural question is whether directly assuming Hölder smoothness on g is too restrictive. In our framework, this assumption is in fact mild, because the convolutional structure $g(u) = \mathbb{E}[q(\epsilon_t - u)]$ allows g to inherit smoothness from more primitive components. The following concrete scenario shows when Assumption 3.4 automatically holds.

Smooth noise distribution. Suppose the noise distribution F itself is Hölder smooth, i.e., $F \in \mathcal{H}(\beta, L_F; [-B, B])$. Even if the nonparametric demand response q only has bounded variation (allowing jumps and discontinuities), the convolution with F regularizes g . The following proposition formalizes this.

Proposition 3.5 Under Assumptions 3.1 and 3.2, and assuming that q is of bounded variation on $[0, B]$, if $F \in \mathcal{H}(\beta, L_F; [-B, B])$, then $g \in \mathcal{H}(\beta, (D + V_q)L_F; \mathcal{I}_g)$, where V_q denotes the total variation of q on $[0, B]$.

In the special case of binary demand where $q(z) = \mathbf{1}\{0 \leq z \leq B\}$, the induced link reduces to the survival function $g(u) = 1 - F(u)$ whenever F is continuous, thereby recovering the classical binary-purchase dynamic pricing model. This special case encompasses a broad range of existing work that assumes F to be Lipschitz (Besbes and Zeevi, 2009; Gong et al., 2025), to have a Lipschitz first derivative (Wang and Chen, 2025), or to be m -times continuously differentiable (Fan et al., 2024; Wang et al., 2021). This example demonstrates that Assumption 3.4 is not an extraneous restriction: it follows from smooth valuation noise even when the quantity response itself has jumps. The model contains the bounded binary-purchase formulation as a special case while relaxing its revenue-geometry restrictions; it does not require the revenue function to be concave or the optimal price to be unique.

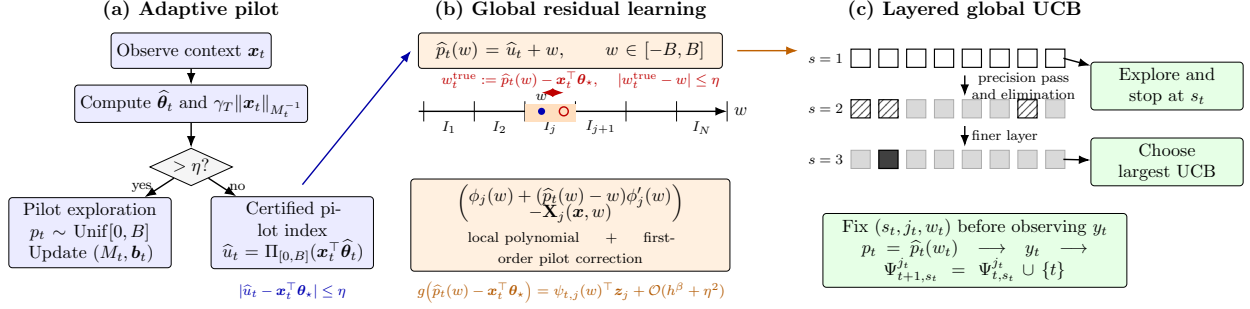


Fig. 1: Overview of the proposed adaptive semiparametric pricing policy. Panel (a) shows the uncertainty-triggered pilot module. Panel (b) illustrates global residual-space learning and the augmented first-order correction for pilot-index error. Panel (c) depicts nested survivor sets under layered UCB exploration and elimination; outline, hatching, and solid fill distinguish active, surviving, and final actions in grayscale.

Remark 3.6 (Strong unimodality) *Several recent studies impose a strong-unimodality condition: for every context $\mathbf{x}$, the function $\text{Rev}(\mathbf{x}, \cdot)$ has a unique interior maximizer, and the revenue loss is comparable to the squared distance from that maximizer (Wang and Chen, 2025; Han et al., 2026; Fan et al., 2026). This structure permits the policy to localize around a single price and, under additional smoothness, to learn a stable oracle-price map.*

We do not impose strong unimodality. Consequently, $\text{Rev}(\mathbf{x}, \cdot)$ may have multiple separated modes or nonunique maximizers, and small estimation errors may change which price region appears optimal. Localized search and oracle-price-map reductions are therefore not generally valid, motivating the global residual-space learning procedure developed in Section 4.

4 Algorithm

The policy has three coupled components, summarized in Figure 1. First, uncertainty-triggered uniform-price exploration estimates the valuation index only along the current context direction. Second, a pilot-corrected local-polynomial feature absorbs the first-order index error, leaving the usual local approximation error and a second-order pilot term. Third, a layered rule compares candidate residual actions globally, rather than localizing around one provisional optimizer.

Every main-policy observation receives its residual-bin and stopping-layer labels before demand is observed, and these labels are never recomputed. The resulting predictable datasets support martingale concentration even though prices and visited bins are selected adaptively. The following subsections define the pilot, correction, confidence radius, and global elimination rule in that order.

4.1 Adaptive Pilot Estimation

The pilot module is introduced first because the subsequent residual representation requires an accurate valuation index at the current covariate. Importantly, the policy does not require $\hat{\theta}_t$ to approximate θ_* uniformly in Euclidean norm. It only needs to certify the scalar index $\mathbf{x}_t^\top \theta_*$ along the currently observed direction. This context-specific requirement is particularly useful when the covariates need not be independently sampled: directions that are already well covered can immediately enter the main pricing module, whereas insufficiently covered directions trigger additional pilot exploration.

The pilot estimator is constructed exclusively from uniform-price exploration rounds. On such a

round, the pseudo-response $B\mathbf{1}\{y_t > 0\}$ is an unbiased observation of $\mathbf{x}_t^\top \boldsymbol{\theta}_*$. Accordingly, M_t and $\mathbf{b}_t$ form the regularized design matrix and response vector based on the pilot-exploration data, and $\hat{\boldsymbol{\theta}}_t = M_t^{-1}\mathbf{b}_t$. The quantity $\gamma_T \|\mathbf{x}_t\|_{M_t^{-1}}$ measures the remaining uncertainty in the current valuation index. If this quantity exceeds the target accuracy η , the policy collects an additional uniform-price observation. Otherwise, it invokes the layered pricing module.

Algorithm 1 Contextual Dynamic Pricing with Adaptive Pilot and Permanent Bin Labels

Require: Pilot accuracy η , horizon T , smoothness β , bounds C_θ, C_x, B, D , number of bins N , regularization $\lambda > 0$, confidence constants C_z, C_ψ, C_l , and confidence level $\delta \in (0, 1)$

- 1: Set $\gamma_T \leftarrow B\sqrt{d \log(1 + TC_x^2/d) + 2 \log(4/\delta)} + C_\theta$ and $S \leftarrow \max\{1, \lceil \log_2 \sqrt{T} \rceil\}$,
- 2: Initialize $\mathcal{T}^{\text{exp}} \leftarrow \emptyset$, $M_1 \leftarrow I_d$, $\mathbf{b}_1 \leftarrow \mathbf{0}_d$, and $\Psi_{1,s}^j \leftarrow \emptyset$ for every $s \in [S]$ and $j \in [N]$
- 3: **for** $t = 1, \dots, T$ **do**
- 4: Observe $\mathbf{x}_t$ and set $\hat{\boldsymbol{\theta}}_t \leftarrow M_t^{-1}\mathbf{b}_t$
- 5: **if** $\gamma_T \|\mathbf{x}_t\|_{M_t^{-1}} > \eta$ **then**
- 6: Set $\mathcal{T}^{\text{exp}} \leftarrow \mathcal{T}^{\text{exp}} \cup \{t\}$
- 7: Draw $p_t \sim \text{Unif}[0, B]$, post p_t , and observe $y_t \in [0, D]$
- 8: Set $M_{t+1} \leftarrow M_t + \mathbf{x}_t \mathbf{x}_t^\top$, $\mathbf{b}_{t+1} \leftarrow \mathbf{b}_t + B\mathbf{1}\{y_t > 0\}\mathbf{x}_t$
- 9: Set $\Psi_{t+1,s}^j \leftarrow \Psi_{t,s}^j$, for all $s \in [S]$ and $j \in [N]$
- 10: **else**
- 11: Run Algorithm 2 and obtain (s_t, j_t, w_t, p_t)
- 12: Record (s_t, j_t, w_t) as the permanent layer, bin, and residual-action labels of period t
- 13: Post p_t and observe $y_t \in [0, D]$
- 14: Set $M_{t+1} \leftarrow M_t$, $\mathbf{b}_{t+1} \leftarrow \mathbf{b}_t$, and

$$\Psi_{t+1,s}^j \leftarrow \begin{cases} \Psi_{t,s}^j \cup \{t\}, & s = s_t, j = j_t \\ \Psi_{t,s}^j, & \text{otherwise.} \end{cases}$$

Algorithm 1 gives the outer wrapper of the policy; the layered pricing subroutine is specified subsequently. We refer to rounds on which uniform pricing is used as pilot-exploration rounds, and to all remaining rounds as LDP rounds. Although later pilot-exploration rounds may update $\hat{\boldsymbol{\theta}}_t$, the bin and layer labels assigned to an LDP observation are permanent and are never recomputed using a later pilot.

Algorithm 2 assigns each LDP round a stopping layer s_t and a permanent bin label j_t . For each pair (s, j) , the outer wrapper maintains a dataset $\Psi_{t,s}^j$, which stores all past LDP observations carrying that fixed label. The task of selecting (s_t, j_t) is entirely delegated to Algorithm 2; the outer algorithm only appends the current observation to the corresponding $\Psi_{t+1,s_t}^{j_t}$.

The following lemma validates the uncertainty gate in Algorithm 1. It provides a time-uniform, direction-dependent confidence bound and therefore guarantees that the valuation index is sufficiently accurate whenever the policy enters an LDP round.

Lemma 4.1 (Pilot estimation confidence) *With the value of γ_T specified in Algorithm 1, with probability at least $1 - \delta/2$, simultaneously for every $t \leq T$ and every $\mathbf{x} \in \mathbb{R}^d$,*

$$\left| \mathbf{x}^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_*) \right| \leq \gamma_T \|\mathbf{x}\|_{M_t^{-1}}. \quad (1)$$

Consequently, on the event in Lemma 4.1, every LDP round t satisfies $\left| \Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t) - \mathbf{x}_t^\top \boldsymbol{\theta}_* \right| \leq \eta$.

Therefore, for every feasible pilot-residual action w ,

$$\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star = w + \left[\Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star \right],$$

so the true residual differs from its pilot counterpart w by at most η . This certification localizes the evaluation point of g , but it does not by itself eliminate the resulting first-order error. We next incorporate this displacement directly into the local regression feature.

4.2 Pilot-Corrected Local Regression

The pilot certificate controls the valuation-index error at the current context, but does not by itself remove its effect on estimating the unknown demand link. On the pilot-confidence event, $|\hat{p}_t(w) - w - \mathbf{x}_t^\top \boldsymbol{\theta}_\star| \leq \eta$. An ordinary local-polynomial regression indexed by w ignores this displacement and therefore incurs a first-order error of order η . When $\beta > 1$, this error may dominate the target local-polynomial bias h^β , where h denotes the bin width defined below.

Existing smooth semiparametric pricing methods address this coupling by jointly refining the index parameter and the local approximation of g (Wang and Chen, 2025; Han et al., 2026). Conditional on a candidate $\boldsymbol{\theta}$, both the local regressors and their Gram matrix depend on $\boldsymbol{\theta}$. Profiling out the polynomial coefficients consequently leads to a constrained least-squares problem that is generally nonconvex and does not, in general, admit a globally certified solution.

Our key simplification is to target prediction rather than structural joint identification. Pricing requires accurate predictions of $g(p - \mathbf{x}^\top \boldsymbol{\theta}_\star)$, but does not require a separate refinement of $\boldsymbol{\theta}_\star$ within every local regression. We therefore retain the complete first-order effect of the pilot-index displacement and estimate the resulting products as composite coefficients. This prediction-preserving lifting converts the coupled estimation problem into a convex ridge regression while leaving only a second-order pilot remainder.

To formalize this construction, partition the residual space $[-B, B]^1$ into N bins of equal width $h = 2B/N$. Let $a_j = -B + \frac{2Bj}{N}$, for $j = 0, 1, \dots, N$, and define

$$I_j = [a_{j-1}, a_j), \quad j = 1, \dots, N-1, \quad I_N = [a_{N-1}, a_N].$$

For each I_j , define the local-polynomial feature anchored at a_{j-1} by

$$\phi_j(u) := (1, u - a_{j-1}, \dots, (u - a_{j-1})^{\varpi(\beta)})^\top \in \mathbb{R}^{1+\varpi(\beta)},$$

and let $\phi'_j(u)$ denote its componentwise derivative. When $\varpi(\beta) \geq 1$, define

$$\mathbf{X}_j(\mathbf{x}, u) := \begin{pmatrix} \mathbf{x} \\ 2(u - a_{j-1})\mathbf{x} \\ \vdots \\ \varpi(\beta)(u - a_{j-1})^{\varpi(\beta)-1}\mathbf{x} \end{pmatrix} \in \mathbb{R}^{\varpi(\beta)d},$$

when $\varpi(\beta) = 0$, let $\mathbf{X}_j(\mathbf{x}, u)$ be the empty vector.

¹For the analysis, we extend g from $\mathcal{I}_g$ to $[-B, B]$ by its endpoint Taylor polynomials of order $\varpi(\beta)$, and continue to denote the extension by g . The extension agrees with the original induced link on $\mathcal{I}_g$ and belongs to $\mathcal{H}(\beta, L_g; [-B, B])$. It is used only to define the comparison coefficients and does not modify the demand model or the policy.

For $w \in I_j$, expanding around the bin anchor and retaining the first-order pilot displacement gives

$$\begin{aligned} g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star) &= \phi_j(w)^\top \begin{pmatrix} g(a_{j-1}) \\ g'(a_{j-1}) \\ \vdots \\ g^{(\varpi(\beta))}(a_{j-1})/\varpi(\beta)! \end{pmatrix} \\ &\quad + (\hat{p}_t(w) - w - \mathbf{x}_t^\top \boldsymbol{\theta}_\star) \phi'_j(w)^\top \begin{pmatrix} g(a_{j-1}) \\ g'(a_{j-1}) \\ \vdots \\ g^{(\varpi(\beta))}(a_{j-1})/\varpi(\beta)! \end{pmatrix} + R_{t,j}(w), \end{aligned} \quad (2)$$

where, under $\eta \leq h$, $|R_{t,j}(w)| \leq C_l (h^\beta + \eta^2)$. The two components of the remainder correspond to the local-polynomial approximation error and the second-order pilot-displacement error, respectively.

Hence, the first-order representation in (2) can be rewritten as

$$[\phi_j(w) + (\hat{p}_t(w) - w) \phi'_j(w)]^\top \begin{pmatrix} g(a_{j-1}) \\ g'(a_{j-1}) \\ \vdots \\ g^{(\varpi(\beta))}(a_{j-1})/\varpi(\beta)! \end{pmatrix} - \mathbf{X}_j(\mathbf{x}_t, w)^\top \begin{pmatrix} g'(a_{j-1}) \boldsymbol{\theta}_\star \\ \vdots \\ \frac{g^{(\varpi(\beta))}(a_{j-1})}{\varpi(\beta)!} \boldsymbol{\theta}_\star \end{pmatrix}.$$

The second term contains products between the local derivative coefficients and $\boldsymbol{\theta}_\star$. Enforcing their common factorization through $\boldsymbol{\theta}_\star$ would recover the nonlinear coupling underlying the constrained joint estimators in prior work. Instead, we treat these products directly as composite coefficients.

Specifically, define a $1 + \varpi(\beta)(d+1)$ -dimensional column vector

$$\mathbf{z}_j := \left(g(a_{j-1}), g'(a_{j-1}), \dots, \frac{g^{(\varpi(\beta))}(a_{j-1})}{\varpi(\beta)!}, (g'(a_{j-1}) \boldsymbol{\theta}_\star)^\top, \dots, \left(\frac{g^{(\varpi(\beta))}(a_{j-1})}{\varpi(\beta)!} \boldsymbol{\theta}_\star \right)^\top \right)^\top,$$

and, for every t, j , and $w \in I_j$, define the pilot-corrected feature

$$\psi_{t,j}(w) := \begin{pmatrix} \phi_j(w) + (\hat{p}_t(w) - w) \phi'_j(w) \\ -\mathbf{X}_j(\mathbf{x}_t, w) \end{pmatrix} \in \mathbb{R}^{1+\varpi(\beta)(d+1)}. \quad (3)$$

Equation (2) then becomes

$$g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star) = \psi_{t,j}(w)^\top \mathbf{z}_j + R_{t,j}(w), \quad |R_{t,j}(w)| \leq C_l (h^\beta + \eta^2). \quad (4)$$

Thus, up to a higher-order deterministic remainder, the conditional mean belongs to a finite-dimensional linear prediction class. When $\beta = 1$, we have $\varpi(\beta) = 0$; the derivative block is absent, and the pilot displacement is absorbed into the $\mathcal{O}(h)$ approximation error under $\eta \leq h$.

For each layer-bin pair (s, j) , we estimate the composite coefficient using observations carrying the corresponding permanent labels. Define $\Lambda_{t,s}^j := \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \psi_{i,j}(w_i)^\top$. Fix a regularization parameter $\lambda > 0$. The ridge estimator is

$$\hat{\mathbf{z}}_{t,s}^j := \underset{\mathbf{z} \in \mathbb{R}^{1+\varpi(\beta)(d+1)}}{\operatorname{argmin}} \left\{ \sum_{i \in \Psi_{t,s}^j} (y_i - \psi_{i,j}(w_i)^\top \mathbf{z})^2 + \lambda \|\mathbf{z}\|_2^2 \right\} = (\lambda I + \Lambda_{t,s}^j)^{-1} \sum_{i \in \Psi_{t,s}^j} y_i \psi_{i,j}(w_i). \quad (5)$$

For $w \in I_j$, the corresponding demand estimate based on $\Psi_{t,s}^j$ is

$$\hat{g}_{t,s}^j(w) := \Pi_{[0,D]}(\psi_{t,j}(w)^\top \hat{z}_{t,s}^j). \quad (6)$$

The estimator in (5) does not approximate a local solution of the preceding nonconvex joint-estimation problem. Instead, it relaxes the factorization restrictions among the derivative coefficients and estimates the prediction-sufficient composite coefficients directly. The population vector z_j remains in the lifted class, so this relaxation introduces no additional first-order approximation error. Its deterministic misspecification is bounded by $\mathcal{O}(h^\beta + \eta^2)$.

Once the lifted features have been constructed, (5) is a standard strictly convex ridge regression and therefore has the displayed unique global solution. Moreover, each feature $\psi_{i,j}(w_i)$ is determined before y_i is observed, allowing the stochastic regression errors to be analyzed as a predictable martingale transform. Under $\eta^2 \lesssim h^\beta$, the lifting preserves the local-polynomial approximation order while avoiding nonconvex joint refinement altogether. The next subsection converts these point predictions into confidence bounds and incorporates them into the layered decision rule.

4.3 Layered Decision Making

The preceding pilot-corrected regression converts the coupled semiparametric estimation problem into a convex linear prediction problem and provides a point estimate of demand for each feasible residual action. Point predictions alone, however, are insufficient for sequential pricing: the amount of information varies across layer-bin pairs, and the lifted representation retains a deterministic approximation error of order $\mathcal{O}(h^\beta + \eta^2)$. We therefore construct confidence bounds that jointly account for stochastic estimation error, ridge regularization, and this remaining approximation error, and use them to form optimistic estimates of the corresponding revenues.

Confidence radius. The explicit form of the radius is

$$r_{t,s}^j(w) = \min \left\{ D, \underbrace{C_\psi D \sqrt{\iota_T} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}}_{\text{stochastic estimation error}} + \underbrace{C_z \sqrt{\lambda} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}}_{\text{ridge bias}} \right. \\ \left. + \underbrace{C_l(h^\beta + \eta^2) \sqrt{|\Psi_{t,s}^j|} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}}_{\text{historical approximation error}} + \underbrace{C_l(h^\beta + \eta^2)}_{\text{current-action approximation error}} \right\}, \quad (7)$$

with

$$\iota_T := 2 \log \left(\frac{4SN}{\delta} \right) + [1 + \varpi(\beta)(d+1)] \log \left(1 + \frac{T}{\lambda} \right). \quad (8)$$

If $|\Psi_{t,s}^j| = 0$, we set $r_{t,s}^j(w) = D$.

The four summands in (7) have distinct roles: self-normalized concentration controls the centered demand noise; the ridge penalty contributes $C_z \sqrt{\lambda}$; the accumulated historical remainders contribute the term proportional to $\sqrt{|\Psi_{t,s}^j|}$; and the final term controls the current action's approximation error. Projection onto $[0, D]$ cannot increase error relative to a target in that interval. Proposition 4.2 formalizes this decomposition and proves that $r_{t,s}^j(w)$ is a simultaneous confidence radius.

Proposition 4.2 (Confidence bound) *Suppose Assumptions 3.1, 3.2, and 3.4 hold. Let $\lambda > 0$, $\delta \in (0, 1)$, and $0 < \eta \leq h$. Let $C_z, C_l, C_\psi > 0$ be the finite constants specified in (25), (27), and*

(35). Then, with probability at least $1 - \delta$, simultaneously for every LDP round t , every $s \in [s_t]$, and every $(j, w) \in \mathcal{A}_{t,s}$,

$$\left| \hat{g}_{t,s}^j(w) - g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| \leq r_{t,s}^j(w). \quad (9)$$

Algorithm 2 Layered Decision Making with Pilot-Residual Actions

Require: Current round t , context $\mathbf{x}_t$, certified pilot $\hat{\boldsymbol{\theta}}_t$, permanent datasets $\{\Psi_{t,s}^j\}_{s \in [S], j \in [N]}$, bin partition $\{I_j\}_{j \in [N]}$, and constants C_ψ, C_z, C_l

```

1: Discretize the residual-action space by  $\mathcal{W} := \{-B + kT^{-1/2} : k = 0, 1, \dots, \lfloor 2B\sqrt{T} \rfloor\} \cup \{0, B\}$ .
2: Set  $\hat{u}_t \leftarrow \Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t)$ ,  $\hat{p}_t(w) \leftarrow \hat{u}_t + w$ .
3: for each  $j \in [N]$  do
4:   Set  $\mathcal{W}_{t,j} \leftarrow \{w \in \mathcal{W} \cap I_j : 0 \leq \hat{p}_t(w) \leq B\}$ .
5: Set  $s \leftarrow 1, \mathcal{A}_{t,1} \leftarrow \{(j, w) : j \in [N], w \in \mathcal{W}_{t,j}\}$ .
6: repeat
7:   Set  $\mathcal{J}_{t,s} \leftarrow \{j \in [N] : (j, w) \in \mathcal{A}_{t,s} \text{ for some } w\}$ .
8:   for each  $j \in \mathcal{J}_{t,s}$  do
9:     Compute  $\hat{\mathbf{z}}_{t,s}^j$  by (5)
10:  for each  $(j, w) \in \mathcal{A}_{t,s}$  do
11:    Compute  $\hat{g}_{t,s}^j(w) \leftarrow \Pi_{[0,D]}(\psi_{t,j}(w)^\top \hat{\mathbf{z}}_{t,s}^j)$ .
12:    Compute  $r_{t,s}^j(w)$  by (7)
13:    Set  $U_{t,s}^j(w) \leftarrow \hat{p}_t(w) \min\{D, \hat{g}_{t,s}^j(w) + r_{t,s}^j(w)\}$ ,  $\text{wid}_{t,s}^j(w) \leftarrow \hat{p}_t(w) r_{t,s}^j(w)$ .
14:  if  $s = S$  then
15:    Select  $(j_t, w_t) \in \arg\max_{(j,w) \in \mathcal{A}_{t,s}} U_{t,s}^j(w)$ ,  $s_t \leftarrow s$ .
16:  else if  $\max_{(j,w) \in \mathcal{A}_{t,s}} \text{wid}_{t,s}^j(w) \leq BD 2^{-s}$  then
17:    Set  $\mathcal{A}_{t,s+1} \leftarrow \left\{ (j, w) \in \mathcal{A}_{t,s} : U_{t,s}^j(w) \geq \max_{(j',w') \in \mathcal{A}_{t,s}} U_{t,s}^{j'}(w') - BD 2^{1-s} \right\}$ ,
18:    Set  $s \leftarrow s + 1$ .
19:  else
20:    Set  $\mathcal{A}_{t,s}^u \leftarrow \{(j, w) \in \mathcal{A}_{t,s} : \text{wid}_{t,s}^j(w) > BD 2^{-s}\}$ .
21:    Select  $(j_t, w_t) \in \arg\max_{(j,w) \in \mathcal{A}_{t,s}^u} U_{t,s}^j(w)$ ,  $s_t \leftarrow s$ .
22: until  $(j_t, w_t)$  has been selected
23: Set  $p_t \leftarrow \hat{p}_t(w_t)$ .
24: return  $(s_t, j_t, w_t, p_t)$ .
```

Layered exploration and elimination. The layered decision rule proceeds by refining the revenue resolution level by level. At layer s , the target resolution is $BD 2^{-s}$. For every action (j, w) surviving in the current active set $\mathcal{A}_{t,s}$, the algorithm maintains an optimistic revenue $U_{t,s}^j(w)$ and a width $\text{wid}_{t,s}^j(w)$ that measures the remaining uncertainty in its estimated revenue. The policy then decides whether to sample at this layer or to advance.

If there exists an action in $\mathcal{A}_{t,s}$ with width strictly greater than $BD 2^{-s}$, the policy selects one such under-explored action and assigns the current observation permanently to layer s and the corresponding bin j . This sampling step directly reduces the uncertainty of that action.

Otherwise, if every surviving action satisfies $\text{wid}_{t,s}^j(w) \leq BD 2^{-s}$, then the layer is considered sufficiently resolved. Our policy eliminates all actions whose optimistic revenue falls more than $BD 2^{1-s}$ below the largest optimistic revenue among the survivors, and passes the remaining actions to the next layer $s + 1$. The safety of this elimination follows from the confidence bounds. Crucially, this argument does not require strong unimodality, local curvature, or uniqueness of the revenue maximizer, because it compares all surviving actions globally.

If the policy proceeds to layer $s + 1$, it repeats the same check with the finer resolution $BD 2^{-(s+1)}$. Once all layers up to S have been passed, the final action is chosen as the one with the largest UCB among the remaining active set.

Algorithm 2 implements this layered procedure. It starts with the full feasible residual grid and sequentially examines the layers. At each layer, it first computes the required estimates and widths; if an under-explored action is found, the round stops and the observation is assigned to that layer. Otherwise, it performs elimination and continues without consuming the current observation. The process terminates when either an action is selected for sampling or the final layer is reached, in which case the action with the highest UCB is chosen.

5 Regret Analysis

5.1 Upper Bound

We now analyze the regret of the proposed policy. The proof follows the two operating modes of Algorithm 1. On LDP rounds, the confidence bounds translate the stopping layer into a one-period revenue guarantee. We then control how often the policy can stop at each layer by combining the layer-specific uncertainty threshold with an elliptical-potential bound for the permanently labeled observations. This yields a bound on the regret accumulated over all LDP rounds. Separately, the uncertainty gate limits the number of pilot-exploration rounds through the growth of the pilot design matrix. Combining these two components gives the finite-sample regret bound, after which we establish a matching lower bound in its dependence on the horizon.

We begin with the one-period analysis of an LDP round. The confidence event guarantees that the best action in the current discretized candidate set survives every completed layer. Consequently, the stopping layer controls the revenue loss relative to the best discretized action, while the mesh of the residual grid controls the additional loss relative to the continuous oracle price.

Lemma 5.1 *Fix an LDP round t , and suppose that the confidence bounds (9) hold at every layer visited in period t . For each visited layer s , define $V_{t,s} := \max_{(j,w) \in \mathcal{A}_{t,s}} \text{Rev}(\mathbf{x}_t, \hat{p}_t(w))$. Then:*

1. *If the algorithm passes the precision check at a layer $s < S$, then $V_{t,s+1} = V_{t,s}$.*
2. *If $s_t \geq 2$, then $V_{t,1} - \text{Rev}(\mathbf{x}_t, p_t) \leq 8BD 2^{-s_t}$.*

Moreover, there exists a finite constant $L_{\text{Rev}} > 0$, such that, for every $s_t \in [S]$,

$$\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t) \leq \frac{L_{\text{Rev}}}{\sqrt{T}} + 8BD 2^{-s_t}. \quad (10)$$

Lemma 5.1 reduces the regret over LDP rounds to the weighted layer occupancy $\sum_{s=1}^S 2^{-s} |\Psi_{T+1,s}|$. Here $\Psi_{T+1,s} := \bigcup_{j=1}^N \Psi_{T+1,s}^j$ denotes the set of all LDP rounds assigned to layer s . It therefore remains to control the number of rounds terminating at each layer. The required auxiliary results

are established in the appendix. Lemma A.2 bounds the cumulative revenue uncertainty within a layer by applying a sequential elliptical-potential argument within each permanently labeled residual bin and then aggregating across bins; the resulting bound separates statistical uncertainty from the approximation error induced by the local-polynomial remainder and the residual second-order pilot-index error. Lemma A.3 then exploits the fact that every round terminating at a nonterminal layer $s < S$ has revenue width exceeding $BD2^{-s}$ to convert the cumulative-width bound into an occupancy bound whenever the layer resolution dominates the approximation floor $\mathcal{O}(h^\beta + \eta^2)$. These results provide the ingredients needed to aggregate the layer-dependent losses.

These ingredients lead to a natural three-regime decomposition of the LDP regret. For nonterminal layers whose resolution remains above the approximation floor, the occupancy bounds apply, and the resulting contribution is governed by statistical uncertainty. Once the resolution falls to the approximation scale, a separate occupancy bound is no longer needed, because the per-round loss at these finer layers can be charged directly to the approximation error. Finally, each round reaching the terminal layer contributes at most its prescribed resolution 2^{-S} . Summing the contributions from these three regimes yields the total regret incurred over the LDP rounds.

Proposition 5.2 (Regret of the discretized LDP step) *On the uniform confidence event in Proposition 4.2, we have*

$$\begin{aligned} \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] &\leq 64C_\omega B(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{NT\iota_T} \\ &\quad + 16C_\omega BT(h^\beta + \eta^2)\sqrt{\iota_T} + (8BD + L_{\text{Rev}})\sqrt{T}. \end{aligned} \quad (11)$$

Suppose that $\lambda > 0$ is fixed independently of T , $\eta^2 \leq c_\eta h^\beta$ for some fixed constant $c_\eta > 0$ in addition to $\eta \leq h$, $N = \left\lceil T^{\frac{1}{2\beta+1}} \right\rceil$ and $h = 2B/N$, then there exists a finite constant $C_{\text{LDP}} > 0$ such that

$$\sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq C_{\text{LDP}} \iota_T T^{\frac{\beta+1}{2\beta+1}}. \quad (12)$$

It remains to control the pilot-exploration rounds. Such a round is triggered only when the leverage score of the current covariate exceeds the target index accuracy. Each trigger produces a multiplicative increase in the determinant of the pilot design matrix. Since the determinant is also bounded above by the trace of that matrix, only a limited number of pilot-exploration rounds can occur. This argument is pathwise and does not consume any additional failure probability.

Lemma 5.3 (Pilot-exploration count and regret) *Suppose that Algorithm 1 is implemented with $\eta > 0$. Then the number of pilot-exploration rounds satisfies, pathwise,*

$$|\mathcal{T}^{\text{exp}}| \leq \min \left\{ T, \frac{d \log(1 + TC_x^2/d)}{\log(1 + \eta^2/\gamma_T^2)} \right\} \leq \min \left\{ T, d \left(1 + \frac{\gamma_T^2}{\eta^2} \right) \log \left(1 + \frac{TC_x^2}{d} \right) \right\}. \quad (13)$$

Consequently, the regret incurred on pilot-exploration rounds satisfies

$$\sum_{t \in \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq BD \min \left\{ T, d \left(1 + \frac{\gamma_T^2}{\eta^2} \right) \log \left(1 + \frac{TC_x^2}{d} \right) \right\}. \quad (14)$$

The regret decomposition is now complete. The pilot-exploration bound in Lemma 5.3 holds pathwise, whereas the LDP bound in Proposition 5.2 holds on the uniform confidence event. Hence,

their sum holds on the same event, without an additional union bound. The following theorem first states the resulting finite-sample guarantee and then specializes it to compatible choices of the residual-bin width and pilot accuracy.

Theorem 5.4 (Regret upper bound) *Suppose Assumptions 3.1, 3.2, and 3.4 hold. Let $0 < \eta \leq h$, and γ_T, ι_T be defined as in Algorithm 1 and (8), respectively. Then, with probability at least $1 - \delta$,*

$$\begin{aligned} \text{Reg}(T) \leq BD \min \left\{ T, \frac{d \log(1 + TC_x^2/d)}{\log(1 + \eta^2/\gamma_T^2)} \right\} + 64C_\omega B(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{NT\iota_T} \\ + 16C_\omega BT(h^\beta + \eta^2)\sqrt{\iota_T} + (8BD + L_{\text{Rev}})\sqrt{T}, \end{aligned} \quad (15)$$

where C_ω and L_{Rev} are the constants in Lemma A.2 and Lemma 5.1.

Moreover, let $N = \lceil T^{\frac{1}{2\beta+1}} \rceil$ and $\eta^2 = \min\{h^2, h^\beta\}$. Then with probability at least $1 - \delta$, there exists a finite constant $C > 0$ such that

$$\text{Reg}(T) \leq C \left[d(1 + \gamma_T^2) \log \left(1 + \frac{TC_x^2}{d} \right) + (1 + \sqrt{\lambda})\iota_T \right] T^{\frac{\beta+1}{2\beta+1}}. \quad (16)$$

The theorem separates the main sources of regret. Pilot exploration pays for certifying the valuation index along the covariate directions encountered by the policy; its contribution is controlled by the uncertainty-triggered determinant argument. On LDP rounds, $\sqrt{NT}$ is the statistical cost of learning across the N residual bins, whereas Th^β is the local-polynomial approximation cost. The pilot-corrected lifting leaves only the second-order contribution $T\eta^2$, so choosing $\eta^2 \leq h^\beta$ prevents pilot-index error from worsening the nonparametric rate. The residual-grid discretization and terminal layer contribute only order $\sqrt{T}$. Choosing $N \asymp T^{\frac{1}{2\beta+1}}$, $h \asymp T^{-\frac{1}{2\beta+1}}$ and $\eta^2 \leq h^\beta$ balances the statistical and approximation terms. For fixed d and fixed problem primitives, this yields the regret on the order of $\tilde{\mathcal{O}}\left(T^{\frac{\beta+1}{2\beta+1}}\right)$.

Remark 5.5 (Adaptation to an unknown time horizon) *The knowledge of T can be removed using the standard doubling trick, as commonly adopted in the dynamic-pricing literature (Javanmard and Nazerzadeh, 2019; Luo et al., 2024; Fan et al., 2024; Wang and Chen, 2025; Han et al., 2026). Specifically, the policy is restarted over epochs of geometrically increasing lengths, with all horizon-dependent parameters, including N , calibrated to the current epoch length.*

5.2 Lower Bound

We next establish a minimax lower bound that matches the horizon dependence of the regret upper bound in Theorem 5.4, up to logarithmic factors. The construction uses a constant-covariate, binary-demand subclass. Hence, the lower bound isolates the difficulty of learning an unknown smooth demand function and does not rely on estimating the contextual parameter.

A hard binary-demand subclass. Consider the admissible demand subclass

$$y_t = D\mathbf{1}\{v_t \geq p_t\}, \quad q(z) = D\mathbf{1}\{0 \leq z \leq B\}. \quad (17)$$

For every $u \in \mathcal{I}_g$, this subclass gives $g(u) = \mathbb{E}[q(\epsilon_t - u)] = D\mathbb{P}(\epsilon_t \geq u)$. Indeed, if $u \in \mathcal{I}_g$ and $\epsilon_t \geq u$, then $0 \leq \epsilon_t - u \leq B_\epsilon - (-B + B_\epsilon) = B$. Thus, on the effective residual domain $\mathcal{I}_g$, the normalized link g/D is the survival function of the valuation shock.

Fix an admissible $\theta^\circ \in \Theta$ satisfying Assumption 3.1 over $\mathcal{X}$, and fix any $\mathbf{x}^\circ \in \mathcal{X}$. Because deterministic covariate sequences are admissible, the minimax problem contains the subclass $\mathbf{x}_t \equiv \mathbf{x}^\circ, \theta_\star = \theta^\circ$. For every $p \in [0, B]$, $p - (\mathbf{x}^\circ)^\top \theta^\circ \in [-(\mathbf{x}^\circ)^\top \theta^\circ, B - (\mathbf{x}^\circ)^\top \theta^\circ] \subseteq \mathcal{I}_g$, and the corresponding expected revenue reduces to $\text{Rev}(p) = p g(p - (\mathbf{x}^\circ)^\top \theta^\circ)$.

Theorem 5.6 (Minimax lower bound) *There exist finite constants $\bar{L}_g > 0$, $c > 0$, and $T_0 < \infty$ such that, for every $L_g \geq \bar{L}_g$ and every $T \geq T_0$,*

$$\inf_{\pi} \sup_{\mathcal{P}(L_g)} \mathbb{E}^\pi[\text{Reg}(T)] \geq cT^{\frac{\beta+1}{2\beta+1}}. \quad (18)$$

Here, the infimum is over all nonanticipating, possibly randomized, pricing policies, and the supremum is over all instances in $\mathcal{P}(L_g)$. The class $\mathcal{P}(L_g)$ contains all instances satisfying Assumptions 3.1, 3.2, and 3.4 with Hölder constant at most L_g , together with all admissible covariate sequences. The expectation is taken over both the demand randomness and randomization of the pricing policy.

Corollary 5.7 (Expected-regret upper bound and minimax horizon rate) *Fix $L_g \geq \bar{L}_g$, the dimension, and the problem primitives, and suppose the confidence constants used by the policy are chosen uniformly over $\mathcal{P}(L_g)$. For all sufficiently large T , there are finite constants $0 < \bar{c} \leq \bar{C} < \infty$, independent of T , such that*

$$\bar{c}T^{\frac{\beta+1}{2\beta+1}} \leq \inf_{\pi} \sup_{\mathcal{P}(L_g)} \mathbb{E}^\pi[\text{Reg}(T)] \leq \bar{C}(\log T)^2 T^{\frac{\beta+1}{2\beta+1}}. \quad (19)$$

In particular, the upper and lower bounds match in their horizon exponent.

Construction overview. The main difficulty is to construct a hard family that simultaneously satisfies the smoothness, monotonicity, bounded-support, and zero-mean requirements imposed on the valuation shock. We first construct a smooth baseline link whose expected revenue is constant over a nondegenerate interior price interval. We then place a smooth local perturbation in one of K_T disjoint subintervals of this flat region. Each perturbation creates a hidden revenue improvement, while a smaller adjustment away from the flat region preserves the zero-mean condition.

The following lemma provides the required baseline link and the function used for the compensating adjustment.

Lemma 5.8 (A flat zero-mean baseline) *There exist constants $0 < p_L < p_U < B$, a nonnegative infinitely differentiable function $\chi : \mathbb{R} \rightarrow [0, \infty)$, and a nonincreasing function $g_0 : \mathbb{R} \rightarrow [0, D]$ satisfying the following properties:*

1. *The function g_0 is infinitely differentiable, $g_0(u) = D$ for $u \leq -B_\epsilon$, and $g_0(u) = 0$ for $u \geq B_\epsilon$.*
2. *The function g_0 belongs to $\mathcal{H}(\beta)$ on $[-B, B]$, with a finite Hölder constant.*
3. *The normalized function g_0/D is the survival function of a zero-mean distribution supported on $[-B_\epsilon, B_\epsilon]$. In view of the monotonicity and endpoint conditions above, its zero-mean property is equivalently expressed as $\int_{-B_\epsilon}^{B_\epsilon} g_0(u) du = DB_\epsilon$.*
4. *The baseline revenue satisfies*

$$p g_0(p - (\mathbf{x}^\circ)^\top \theta^\circ) \leq D \left((\mathbf{x}^\circ)^\top \theta^\circ - \frac{B_\epsilon}{2} \right), \quad p \in [0, B], \quad (20)$$

with equality for every $p \in [p_L, p_U]$.

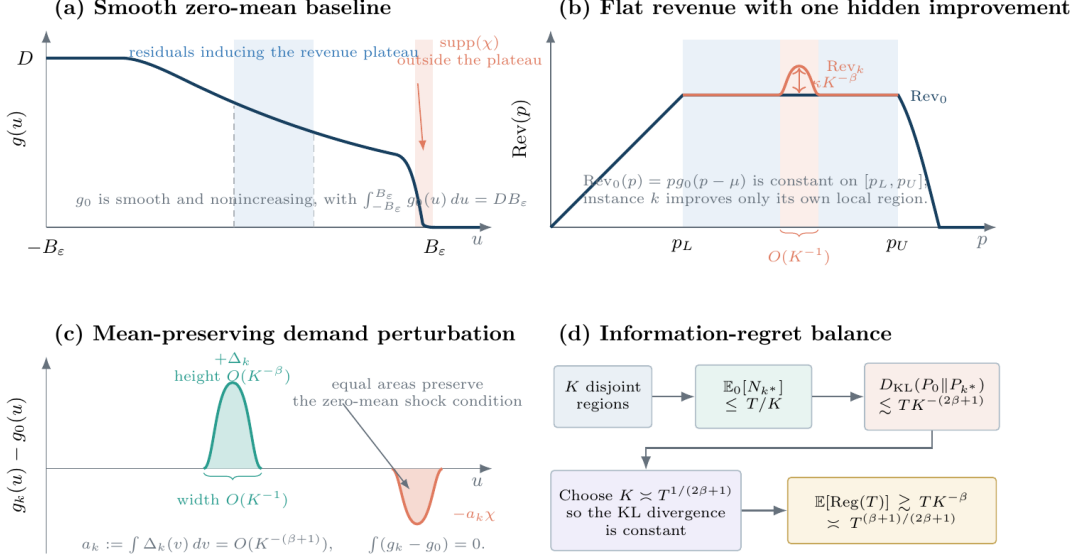


Fig. 2: Illustration of the lower-bound construction. Panel (a) constructs a smooth, nonincreasing baseline demand function g_0 satisfying the zero-mean integral condition. Panel (b) shows the induced flat revenue region and a local perturbation of width $\mathcal{O}(K_T^{-1})$ and height $\kappa K_T^{-\beta}$. Panel (c) adds the compensating term $-a_k \chi$, where $a_k = \int \Delta_k(v) dv = \mathcal{O}(K_T^{-(\beta+1)})$, so that the perturbation preserves the zero-mean shock condition. Panel (d) summarizes the information-regret balance: choosing $K_T \asymp T^{1/(2\beta+1)}$ keeps the KL divergence bounded and yields the regret lower bound $\Omega(T^{(\beta+1)/(2\beta+1)})$.

5. The function χ satisfies $\text{supp}(\chi) \subset (-B_\epsilon, B_\epsilon) \setminus [p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$, $\int_{\mathbb{R}} \chi(u) du = 1$. On neighborhoods of both $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$ and $\text{supp}(\chi)$, the function g_0 is bounded away from 0 and D , and g'_0 is uniformly negative.

Starting from Lemma 5.8, partition the flat revenue interval into K_T disjoint price regions. For the k th region, add a smooth local perturbation Δ_k and subtract the compensating term $a_k \chi$, where $a_k = \int_{\mathbb{R}} \Delta_k(v) dv$. Thus, the perturbed link is of the form $g_0(u) + \Delta_k(u) - a_k \chi(u)$. Because $\int_{\mathbb{R}} \chi(u) du = 1$, $\int_{\mathbb{R}} (\Delta_k(u) - a_k \chi(u)) du = 0$. Consequently, the perturbation does not alter the integral condition corresponding to the zero-mean valuation shock.

The local perturbation has width $\mathcal{O}(K_T^{-1})$ and raises the revenue in its associated region by order $K_T^{-\beta}$. Its integral therefore satisfies $a_k = \mathcal{O}(K_T^{-(\beta+1)})$. Hence, the compensating adjustment is smaller than the local perturbation by a factor of order K_T^{-1} . By placing its support away from the flat region and choosing the perturbation magnitude sufficiently small, all perturbed links remain nonincreasing, take values in $[0, D]$, and belong to a common Hölder ball.

Information-regret balance. Under the baseline instance, the expected total number of visits to all K_T disjoint perturbation regions is at most T ; hence some region is visited at most T/K_T times in expectation. In that region, the perturbation changes the Bernoulli success probability by $\Theta(K_T^{-\beta})$, producing a per-observation KL divergence of order $K_T^{-2\beta}$. The local KL contribution is therefore $\mathcal{O}(\frac{T}{K_T} K_T^{-2\beta}) = \mathcal{O}(TK_T^{-(2\beta+1)})$. The compensating adjustment, of magnitude $\mathcal{O}(K_T^{-(\beta+1)})$, contributes at most $\mathcal{O}(TK_T^{-2(\beta+1)})$ to the KL divergence, which is of lower order. Hence the total KL divergence remains bounded when $K_T \asymp T^{1/(2\beta+1)}$. A standard testing argument then shows

that with probability bounded away from zero, the policy fails to identify the perturbed region for a constant fraction of the horizon. Since any price outside that region incurs a revenue loss of order $K_T^{-\beta}$, the expected regret is at least of order $TK_T^{-\beta} \asymp T^{\frac{\beta+1}{2\beta+1}}$, proving Theorem 5.6. Therefore, for fixed problem primitives, the regret rate $\tilde{\mathcal{O}}\left(T^{\frac{\beta+1}{2\beta+1}}\right)$ is minimax optimal in its dependence on the horizon.

Remark 5.9 (The effect of strong unimodality) *The hard family of instances underlying Theorem 5.6 is based on a revenue function that is flat over a nondegenerate price interval. It therefore lies outside any class that satisfies the strong unimodality condition. The strong unimodality condition, however, rules out such flat alternatives and ensures that price errors translate into quadratic revenue losses. For $\beta \geq 2$, Fan et al. (2026) obtain the matching horizon exponent $\frac{2\beta-1}{4\beta-3}$. Thus, strong unimodality fundamentally changes the learning difficulty and permits higher-order smoothness to be exploited more effectively.*

6 Conclusion and Future Research

This paper studies contextual dynamic pricing with an unknown linear valuation component and an unknown nonparametric demand response. The model accommodates bounded discrete or continuous purchase quantities and imposes only Hölder smoothness on the induced demand function. In particular, the revenue function may be multimodal and its maximizer need not be unique. The framework therefore applies beyond the binary purchase model and avoids the strong shape restrictions commonly used to make contextual pricing analytically tractable.

To address the interaction between parametric estimation, nonparametric learning, and adaptive pricing, we develop a layered decision-partitioning framework that combines directional pilot learning, pilot-corrected local polynomial regression, and permanent layer-bin assignments. The pilot correction reduces the remaining index error to second order, whereas the permanent data partition preserves predictability and enables concentration under adaptively collected observations. Together with a global elimination procedure, these ingredients yield the regret bound $\text{Reg}(T) = \tilde{\mathcal{O}}\left(T^{\frac{\beta+1}{2\beta+1}}\right)$, which matches our lower bound in its dependence on the horizon.

Two directions appear particularly promising for future research. First, it would be valuable to develop a policy that adapts simultaneously to the unknown Hölder radius L_g and smoothness order β , while preserving predictable data assignment and valid confidence guarantees under adaptive sampling. An accompanying question is whether such simultaneous adaptation necessarily incurs an additional statistical cost. Second, although the present framework accommodates general, possibly multimodal revenue functions, it does not exploit favorable local revenue geometry when such structure is present. Developing a shape-adaptive policy that achieves faster regret under quadratic revenue growth or strong unimodality, while retaining the minimax guarantee over the unrestricted revenue class, would yield a desirable best-of-both-worlds result.

References

- Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. Advances in neural information processing systems, 24.
- Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research, 3(Nov):397–422.
- Ban, G.-Y. and Keskin, N. B. (2021). Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. Management Science, 67(9):5549–5568.
- Besbes, O. and Zeevi, A. (2009). Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. Operations research, 57(6):1407–1420.
- Besbes, O. and Zeevi, A. (2015). On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. Management Science, 61(4):723–739.
- Bu, J., Simchi-Levi, D., and Wang, C. (2022). Context-based dynamic pricing with partially linear demand model. Advances in Neural Information Processing Systems, 35:23780–23791.
- Bu, J., Simchi-Levi, D., and Wang, C. (2025). Context-based dynamic pricing with separable demand models. Management Science.
- Chai, J., Duan, Y., Fan, J., and Wang, K. (2026). Optimal semiparametric dynamic pricing with feature diversity. arXiv preprint arXiv:2605.04207.
- Chen, E., Chen, X., Gao, L., and Li, J. (2024). Dynamic contextual pricing with doubly non-parametric random utility models. arXiv preprint arXiv:2405.06866.
- Choi, Y.-G., Kim, G.-S., Yunseo, C., Cho, W., Paik, M. C., and Oh, M.-h. (2023). Semi-parametric contextual pricing algorithm using cox proportional hazards model. In International Conference on Machine Learning, pages 5771–5786. PMLR.
- Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandits with linear payoff functions. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 208–214. JMLR Workshop and Conference Proceedings.
- Den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. Surveys in operations research and management science, 20(1):1–18.
- Fan, J., Guo, Y., and Yu, M. (2024). Policy optimization using semiparametric models for dynamic pricing. Journal of the American Statistical Association, 119(545):552–564.
- Fan, Y., Han, Y., Lv, J., Xu, X., and Zhou, Z. (2026). Harnessing unimodality in semiparametric contextual pricing via oracle price map learning. arXiv preprint arXiv:2605.15411.
- Gong, X., You, W., and Zhang, J. (2025). Minimax optimality in contextual dynamic pricing with general valuation models. Operations Research, 74(2):879–897.
- Han, Y., Xu, X., Wen, Y., Han, Y., Lobel, I., and Zhou, Z. (2026). Semi-parametric contextual pricing with general smoothness. In The Fourteenth International Conference on Learning Representations.

- Härdle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. The Annals of Statistics, 21(1):157–178.
- Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. Journal of the American Statistical Association, 91(436):1632–1640.
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58(1–2):71–120.
- Javanmard, A. and Nazerzadeh, H. (2019). Dynamic pricing in high-dimensions. Journal of Machine Learning Research, 20(9):1–49.
- Lattimore, T. and Szepesvári, C. (2020). Bandit algorithms. Cambridge University Press.
- Luo, Y., Sun, W. W., and Liu, Y. (2022). Contextual dynamic pricing with unknown noise: Explore-then-ucb strategy and improved regrets. Advances in Neural Information Processing Systems, 35:37445–37457.
- Luo, Y., Sun, W. W., and Liu, Y. (2024). Distribution-free contextual dynamic pricing. Mathematics of Operations Research, 49(1):599–618.
- Saharan, S., Bawa, S., and Kumar, N. (2020). Dynamic pricing techniques for intelligent transportation system in smart cities: A systematic review. Computer Communications, 150:603–625.
- Tullii, M., Gaucher, S., Merlis, N., and Perchet, V. (2024). Improved algorithms for contextual dynamic pricing. In Advances in Neural Information Processing Systems, volume 37, pages 126088–126117.
- Wang, Y. and Chen, B. (2025). Tight regret bounds in contextual pricing with semi-parametric demand learning. Available at SSRN 5133677.
- Wang, Y., Chen, B., and Simchi-Levi, D. (2021). Multimodal dynamic pricing. Management Science, 67(10):6136–6152.
- Xu, J. and Wang, Y.-X. (2022). Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. In International Conference on Artificial Intelligence and Statistics, pages 9643–9662. PMLR.
- Xu, J. and Wang, Y.-X. (2026). Optimal contextual pricing under agnostic non-lipschitz demand. arXiv preprint arXiv:2605.05609.

A Proofs

Proof roadmap. The appendix follows the logical dependency of the results. We first justify the induced link and the uniform-price pilot identity. We then establish pilot confidence, the pilot-corrected local approximation, and the uniform layer-bin confidence event. These ingredients feed the one-period gap and layer-counting arguments used in the contextual upper bound. The lower-bound construction is proved next.

A.1 Proofs for the Model

Proof of Theorem 3.5. *Proof.* Let $\mathcal{I}_g = [-B + B_\epsilon, B - B_\epsilon]$. Since $q : \mathbb{R} \rightarrow [0, D]$, the function $g(u) = \mathbb{E}[q(\epsilon_t - u)]$ is well-defined and satisfies $0 \leq g(u) \leq D$ for all $u \in \mathcal{I}_g$. Let $\bar{F}(u) := \mathbb{P}(\epsilon_t \geq u)$ be the survival function of the noise. Since $F \in \mathcal{H}(\beta, L_F; [-B, B])$ with $\beta \geq 1$, F is continuous. Hence ϵ_t has no atoms.

By Assumption 3.1, the support of ϵ_t is contained in $[-B_\epsilon, B_\epsilon]$. Therefore $\bar{F}$ is constant on $(-\infty, -B_\epsilon]$ and on $[B_\epsilon, \infty)$. Since $B_\epsilon < B$, the function $1 - F$ is already constant on $[-B, -B_\epsilon]$ and on $[B_\epsilon, B]$. Thus, when viewed on the interval $[-B + B_\epsilon, 2B - B_\epsilon]$, the survival function $\bar{F}$ belongs to $\mathcal{H}(\beta, L_F; [-B + B_\epsilon, 2B - B_\epsilon])$.

Because q has bounded variation on $[0, B]$, it has at most countably many discontinuities. Let q^+ be a right-continuous representative of q on $[0, B]$ with $\text{TV}_{[0, B]}(q^+) \leq V_q$, extend it by zero outside $[0, B]$, and continue to denote the resulting function by q^+ . Since ϵ_t has no atoms, for each fixed $u \in \mathcal{I}_g$, $\mathbb{E}[q(\epsilon_t - u)] = \mathbb{E}[q^+(\epsilon_t - u)]$. Hence g can be represented using q^+ . By the Lebesgue–Stieltjes representation for bounded-variation functions, there exists a finite signed measure μ_q on $(0, B]$ such that

$$q^+(w) = q^+(0) + \mu_q((0, w]), \quad w \in [0, B],$$

and moreover,

$$|\mu_q|((0, B]) \leq V_q, \quad |q^+(0)| \leq D.$$

For $u \in \mathcal{I}_g$, Assumption 3.1 implies that $\epsilon_t - u \leq B$ almost surely whenever $\epsilon_t - u \geq 0$. Therefore,

$$q^+(\epsilon_t - u) = q^+(0)\mathbf{1}\{\epsilon_t \geq u\} + \int_{(0, B]} \mathbf{1}\{\epsilon_t \geq u + a\} \mu_q(\text{d}a).$$

Taking expectations and applying Fubini's theorem for finite signed measures gives

$$g(u) = q^+(0)\bar{F}(u) + \int_{(0, B]} \bar{F}(u + a) \mu_q(\text{d}a), \quad \forall u \in \mathcal{I}_g.$$

Let $m := \varpi(\beta)$. Since μ_q is finite and $\bar{F} \in \mathcal{H}(\beta, L_F; [-B + B_\epsilon, 2B - B_\epsilon])$, differentiating under the finite signed measure yields

$$g^{(m)}(u) = q^+(0)\bar{F}^{(m)}(u) + \int_{(0, B]} \bar{F}^{(m)}(u + a) \mu_q(\text{d}a).$$

Thus, for any $u, u' \in \mathcal{I}_g$,

$$\begin{aligned} |g^{(m)}(u) - g^{(m)}(u')| &\leq |q^+(0)| |\bar{F}^{(m)}(u) - \bar{F}^{(m)}(u')| \\ &\quad + \int_{(0, B]} |\bar{F}^{(m)}(u + a) - \bar{F}^{(m)}(u' + a)| |\mu_q|(\text{d}a) \\ &\leq (|q^+(0)| + |\mu_q|((0, B])) L_F |u - u'|^{\beta-m} \\ &\leq (D + V_q) L_F |u - u'|^{\beta-m}. \end{aligned}$$

Therefore $g \in \mathcal{H}(\beta, L_g; \mathcal{I}_g)$ with $L_g \leq (D + V_q)L_F$. The proof is complete. $\square$

A.2 Proofs for the Contextual Policy

Proof of Theorem 4.1. *Proof.* On an exploratory round, $p_i \sim \text{Unif}[0, B]$ is drawn independently of ϵ_i conditional on $\mathcal{F}_{i-1} \vee \sigma(\mathbf{x}_i)$. Since $\mathbf{1}\{y_i > 0\} = \mathbf{1}\{v_i \geq p_i\}$ and $v_i \in [0, B]$, we have

$$\mathbb{E}[B\mathbf{1}\{y_i > 0\} \mid \mathcal{F}_{i-1}, \mathbf{x}_i, \{i \in \mathcal{T}^{\text{exp}}\}] = \mathbb{E}[v_i \mid \mathcal{F}_{i-1}, \mathbf{x}_i] = \mathbf{x}_i^\top \boldsymbol{\theta}_*.$$

Thus, $B\mathbf{1}\{y_i > 0\}$ is an unbiased measurement of $\mathbf{x}_i^\top \boldsymbol{\theta}_*$ on pilot rounds. The observation error is bounded by B in absolute value, and $\mathbf{1}\{i \in \mathcal{T}^{\text{exp}}\}\mathbf{x}_i$ is determined before y_i is revealed. By the standard self-normalized concentration inequality for adaptive linear regression, with probability at least $1 - \delta/2$, simultaneously for all $t \leq T$,

$$\left\| \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i (B\mathbf{1}\{y_i > 0\} - \mathbf{x}_i^\top \boldsymbol{\theta}_*) \right\|_{M_t^{-1}} \leq B \sqrt{\log \det(M_t) + 2 \log \frac{4}{\delta}}.$$

The update rules give

$$M_t = I_d + \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i \mathbf{x}_i^\top, \quad \mathbf{b}_t = \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} B\mathbf{1}\{y_i > 0\} \mathbf{x}_i.$$

Since $\|\mathbf{x}_i\|_2 \leq C_x$, standard elliptical-potential arguments yield

$$\log \det(M_t) \leq d \log \left(1 + \frac{TC_x^2}{d} \right).$$

Since $\hat{\boldsymbol{\theta}}_t = M_t^{-1} \mathbf{b}_t$, we have

$$\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_* = M_t^{-1} \mathbf{b}_t - M_t^{-1} M_t \boldsymbol{\theta}_* = M_t^{-1} (\mathbf{b}_t - M_t \boldsymbol{\theta}_*). \quad (21)$$

Moreover,

$$M_t \boldsymbol{\theta}_* = \left(I_d + \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i \mathbf{x}_i^\top \right) \boldsymbol{\theta}_* = \boldsymbol{\theta}_* + \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i (\mathbf{x}_i^\top \boldsymbol{\theta}_*). \quad (22)$$

Substituting (22) into (21) yields

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_* &= M_t^{-1} \left[\sum_{i < t, i \in \mathcal{T}^{\text{exp}}} B\mathbf{1}\{y_i > 0\} \mathbf{x}_i - \boldsymbol{\theta}_* - \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i (\mathbf{x}_i^\top \boldsymbol{\theta}_*) \right] \\ &= M_t^{-1} \left[\sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i (B\mathbf{1}\{y_i > 0\} - \mathbf{x}_i^\top \boldsymbol{\theta}_*) - \boldsymbol{\theta}_* \right]. \end{aligned}$$

For any $\mathbf{x} \in \mathbb{R}^d$, the Cauchy–Schwarz inequality in the M_t^{-1} -norm gives

$$\begin{aligned} |\mathbf{x}^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_*)| &\leq \|\mathbf{x}\|_{M_t^{-1}} \left[\left\| \sum_{i < t, i \in \mathcal{T}^{\text{exp}}} \mathbf{x}_i (B\mathbf{1}\{y_i > 0\} - \mathbf{x}_i^\top \boldsymbol{\theta}_*) \right\|_{M_t^{-1}} + \|\boldsymbol{\theta}_*\|_{M_t^{-1}} \right] \\ &\leq \|\mathbf{x}\|_{M_t^{-1}} \left[B \sqrt{d \log \left(1 + \frac{TC_x^2}{d} \right) + 2 \log \frac{4}{\delta}} + C_\theta \right], \end{aligned}$$

where the second inequality uses $\|\boldsymbol{\theta}_\star\|_{M_t^{-1}} \leq \|\boldsymbol{\theta}_\star\|_2 \leq C_\theta$, and $M_t^{-1} \preceq I_d$. The bracketed expression is precisely γ_T , completing the proof. $\square$

The proof of Proposition 4.2 uses the following martingale property.

Proposition A.1 (Martingale property) *For notational convenience, set $s_i = j_i = w_i = 0$ whenever $i \in \mathcal{T}^{\text{exp}}$. Use the chronological fields defined in Section 3 and write*

$$\mathcal{F}_i^- := \mathcal{G}_i, \quad \mathcal{F}_i = \mathcal{F}_i^- \vee \sigma(y_i), \quad i \in [T].$$

In particular, the exploration decision, price, stopping layer, bin, and residual label are all $\mathcal{F}_i^-$ -measurable. For each $i \in [T]$, define

$$\varepsilon_i := y_i - g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star).$$

Then

$$\mathbb{E}[\varepsilon_i \mid \mathcal{F}_i^-] = 0, \quad |\varepsilon_i| \leq D \quad a.s.$$

For fixed $s \in [S]$ and $j \in [N]$, define the zero-padded feature vector

$$\bar{\psi}_{i,j,s} := \begin{cases} \psi_{i,j}(w_i), & \text{if } i \notin \mathcal{T}^{\text{exp}}, s_i = s, j_i = j, \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

Then $\bar{\psi}_{i,j,s}$ is $\mathcal{F}_i^-$ -measurable, $\|\bar{\psi}_{i,j,s}\|_2 \leq C_\psi$ for some $C_\psi > 0, a.s.$, and $\mathbb{E}[\bar{\psi}_{i,j,s} \varepsilon_i \mid \mathcal{F}_i^-] = \mathbf{0}$. Consequently,

$$\mathbf{M}_0^{s,j} := \mathbf{0}, \quad \mathbf{M}_t^{s,j} := \sum_{i=1}^t \bar{\psi}_{i,j,s} \varepsilon_i, \quad t \in [T],$$

defines a vector-valued martingale with respect to $\{\mathcal{F}_t\}_{t=0}^T$, whose increments satisfy

$$\left\| \mathbf{M}_t^{s,j} - \mathbf{M}_{t-1}^{s,j} \right\|_2 \leq C_\psi D \quad a.s.$$

In particular, for every $t \in [T]$,

$$\mathbf{M}_{t-1}^{s,j} = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i.$$

Proof. By policy nonanticipativity and the sequential exogeneity of the current valuation shock, $\mathbb{P}(\epsilon_i \leq u \mid \mathcal{F}_i^-) = F(u)$. Moreover, the surplus-demand model gives $\mathbb{E}[y_i \mid \mathcal{F}_i^- \vee \sigma(v_i)] = q(v_i - p_i)$. Therefore, by the tower property,

$$\mathbb{E}[y_i \mid \mathcal{F}_i^-] = \mathbb{E}[q(v_i - p_i) \mid \mathcal{F}_i^-] = \int q(\mathbf{x}_i^\top \boldsymbol{\theta}_\star + \epsilon - p_i) dF(\epsilon) = g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star).$$

It follows that $\mathbb{E}[\varepsilon_i \mid \mathcal{F}_i^-] = 0$. Since $y_i \in [0, D]$ almost surely and $g \in [0, D]$, we have $\varepsilon_i \in [-g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star), D - g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star)]$, and hence $|\varepsilon_i| \leq D$.

Now fix $s \in [S]$ and $j \in [N]$. The exploration decision, permanent layer-bin labels, residual action, and posted price are all determined before y_i is observed. Hence $\bar{\psi}_{i,j,s}$ is $\mathcal{F}_i^-$ -measurable. Furthermore, whenever this feature is nonzero,

$$w_i \in I_j, \quad p_i - w_i = \hat{p}_i(w_i) - w_i \in [0, B], \quad \|\mathbf{x}_i\|_2 \leq C_x.$$

Since the local-polynomial degree is fixed, the definition of $\psi_{i,j}$ implies that there exists a deterministic constant $C_\psi < \infty$ such that $\|\bar{\psi}_{i,j,s}\|_2 \leq C_\psi$. Consequently,

$$\mathbb{E}[\bar{\psi}_{i,j,s}\varepsilon_i \mid \mathcal{F}_i^-] = \bar{\psi}_{i,j,s}\mathbb{E}[\varepsilon_i \mid \mathcal{F}_i^-] = \mathbf{0},$$

which proves the martingale-difference property.

To make the associated martingale explicit, define, within this proof,

$$\mathbf{M}_0^{s,j} := \mathbf{0}, \quad \mathbf{M}_t^{s,j} := \sum_{i=1}^t \bar{\psi}_{i,j,s}\varepsilon_i, \quad t \in [T].$$

The process $\{\mathbf{M}_t^{s,j}\}_{t=0}^T$ is $\{\mathcal{F}_t\}_{t=0}^T$ -adapted and integrable. Since $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t^-$, the tower property yields

$$\mathbb{E}[\mathbf{M}_t^{s,j} \mid \mathcal{F}_{t-1}] = \mathbf{M}_{t-1}^{s,j} + \mathbb{E}[\mathbb{E}[\bar{\psi}_{t,j,s}\varepsilon_t \mid \mathcal{F}_t^-] \mid \mathcal{F}_{t-1}] = \mathbf{M}_{t-1}^{s,j}.$$

Thus, $\{\mathbf{M}_t^{s,j}, \mathcal{F}_t\}_{t=0}^T$ is a vector-valued martingale, interpreted componentwise, with increments satisfying

$$\|\mathbf{M}_t^{s,j} - \mathbf{M}_{t-1}^{s,j}\|_2 \leq C_\psi D \quad \text{a.s.}$$

Finally, by the permanent-label definition,

$$\Psi_{t,s}^j = \{i < t : i \notin \mathcal{T}^{\text{exp}}, s_i = s, j_i = j\},$$

and therefore

$$\mathbf{M}_{t-1}^{s,j} = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i)\varepsilon_i.$$

This is precisely the stochastic score appearing in the layer-bin ridge estimator. $\square$

Proof of Theorem 4.2. *Proof.* We proceed in five steps.

Step 1: pilot-index control. Let

$$\mathcal{E}_{\text{pil}} := \left\{ |\mathbf{x}^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_\star)| \leq \gamma_T \|\mathbf{x}\|_{M_t^{-1}} \text{ for every } t \leq T \text{ and every } \mathbf{x} \in \mathbb{R}^d \right\}.$$

Lemma 4.1 gives $\mathbb{P}(\mathcal{E}_{\text{pil}}) \geq 1 - \delta/2$. If i is a main-policy round, then the exploration test did not trigger, and therefore $\gamma_T \|\mathbf{x}_i\|_{M_i^{-1}} \leq \eta$. On $\mathcal{E}_{\text{pil}}$, we have $|\mathbf{x}_i^\top (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_\star)| \leq \eta$. Because Euclidean projection onto $[0, B]$ is nonexpansive and $\mathbf{x}_i^\top \boldsymbol{\theta}_\star \in [0, B]$,

$$\left| \Pi_{[0,B]}(\mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right| \leq \eta. \quad (23)$$

The same inequality holds for the current LDP round t .

Step 2: local Taylor expansion and joint linearization. Since $0 \leq g \leq D$ on $\mathcal{I}_g$, $g \in \mathcal{H}(\beta, L_g; \mathcal{I}_g)$, and $\mathcal{I}_g$ is a fixed nondegenerate compact interval, a standard one-dimensional interpolation inequality implies

$$\max_{0 \leq k \leq \varpi(\beta)} \sup_{u \in \mathcal{I}_g} |g^{(k)}(u)| \leq C_g,$$

where C_g is the fixed class-level envelope specified after Assumption 3.4.

To make the Taylor coefficients at all bin endpoints well defined, extend g from $\mathcal{I}_g = [-B + B_\epsilon, B - B_\epsilon]$ to $[-B, B]$ by its endpoint Taylor polynomials of order $m := \varpi(\beta)$, while continuing to use the same notation g .

The extension agrees with g and its derivatives up to order m at both endpoints of $\mathcal{I}_g$, and is therefore m -times differentiable on $[-B, B]$. Moreover, $g^{(m)}$ is constant on each added interval and equals the corresponding endpoint value. Hence, Assumption 3.4 implies

$$\left| g^{(m)}(u) - g^{(m)}(u') \right| \leq L_g |u - u'|^{\beta-m}, \quad u, u' \in [-B, B].$$

Indeed, when u and u' lie in different regions, the original Hölder bound is applied to the corresponding endpoint or endpoints, whose separation is no greater than $|u - u'|$. Therefore, the extension belongs to $\mathcal{H}(\beta, L_g; [-B, B])$. After increasing the fixed value of C_g , if necessary, the extension therefore satisfies

$$\max_{0 \leq k \leq m} \sup_{u \in [-B, B]} |g^{(k)}(u)| \leq C_g. \quad (24)$$

This extension is used only in the analysis and leaves g unchanged on its original domain $\mathcal{I}_g$; hence it does not alter the demand model or the algorithm.

For the proof, use the convention $g^{(0)} = g$ and define the coefficient vector corresponding to the joint feature. Recall

$$\mathbf{z}_j = g(a_{j-1}).$$

if $\varpi(\beta) = 0$, and

$$\mathbf{z}_j = \begin{pmatrix} g(a_{j-1}) \\ g'(a_{j-1}) \\ \vdots \\ g^{(\varpi(\beta))}(a_{j-1})/\varpi(\beta)! \\ g'(a_{j-1})\boldsymbol{\theta}_\star \\ \vdots \\ [g^{(\varpi(\beta))}(a_{j-1})/\varpi(\beta)!]\boldsymbol{\theta}_\star \end{pmatrix}$$

if $\varpi(\beta) \geq 1$.

Define

$$C_z := C_g \left[\sum_{k=0}^m \frac{1}{(k!)^2} + C_\theta^2 \sum_{k=1}^m \frac{1}{(k!)^2} \right]^{1/2}. \quad (25)$$

Equation (24) and $\|\boldsymbol{\theta}_\star\|_2 \leq C_\theta$ imply that, uniformly over $j \in [N]$,

$$\|\mathbf{z}_j\|_2^2 = \sum_{k=0}^m \left| \frac{g^{(k)}(a_{j-1})}{k!} \right|^2 + \|\boldsymbol{\theta}_\star\|_2^2 \sum_{k=1}^m \left| \frac{g^{(k)}(a_{j-1})}{k!} \right|^2 \leq C_g^2 \left[\sum_{k=0}^m \frac{1}{(k!)^2} + C_\theta^2 \sum_{k=1}^m \frac{1}{(k!)^2} \right] = C_z^2.$$

Since the right-hand side does not depend on j , we obtain

$$\sup_{j \in [N]} \|\mathbf{z}_j\|_2 \leq C_z. \quad (26)$$

Define

$$C_l := \max \left\{ \frac{2^\beta L_g}{\varpi(\beta)!}, C_g \sum_{k=2}^{\varpi(\beta)} \frac{(2^k - k - 1)(2B)^{k-2}}{k!} \right\}. \quad (27)$$

We next establish the joint linearized representation on the event $\mathcal{E}_{\text{pil}}$. Fix $i \in \Psi_{t,s}^j$. Since the bin label assigned to period i is permanent, $w_i \in I_j$, and hence $0 \leq w_i - a_{j-1} \leq h$. Moreover, since $p_i = \hat{p}_i(w_i) = \Pi_{[0,B]}(\mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_i) + w_i$, (23) gives that

$$\left| (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right| \leq \eta. \quad (28)$$

The true residual can therefore be written exactly as

$$p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star = a_{j-1} + (w_i - a_{j-1}) + ((p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star). \quad (29)$$

Moreover, since $p_i \in [0, B]$ and $\mathbf{x}_i^\top \boldsymbol{\theta}_\star \in [B_\epsilon, B - B_\epsilon]$,

$$p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \in [-B + B_\epsilon, B - B_\epsilon] = \mathcal{I}_g.$$

Hence $g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star)$ is evaluated on the original domain of the induced demand link; the extension in Step 2 is used only to define the Taylor coefficients at a_{j-1} . Thus, relative to the expansion point a_{j-1} , the true residual contains two deviations: the within-bin deviation $w_i - a_{j-1}$, whose absolute value is at most h , and the pilot-index error $(p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star$, whose absolute value is at most η .

We next identify the function value represented by the joint feature. By the definitions of $\psi_{i,j}(w_i)$, ϕ_j , ϕ'_j , $\mathbf{X}_j$, and $\mathbf{z}_j$, we have

$$\psi_{i,j}(w_i)^\top \mathbf{z}_j = \sum_{k=0}^{\varpi(\beta)} \frac{g^{(k)}(a_{j-1})}{k!} (w_i - a_{j-1})^k + [(p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star] \sum_{k=1}^{\varpi(\beta)} \frac{g^{(k)}(a_{j-1})}{(k-1)!} (w_i - a_{j-1})^{k-1}. \quad (30)$$

The first sum in (30) is the local Taylor polynomial centered at a_{j-1} and evaluated at w_i . The second sum is the derivative of this polynomial at w_i , multiplied by the pilot-index error in (28). Hence, $\psi_{i,j}(w_i)^\top \mathbf{z}_j$ represents the local Taylor polynomial together with the complete first-order correction for the pilot-index error. When $\varpi(\beta) = 0$, the derivative term is absent.

If $\beta = 1$, then $\varpi(\beta) = 0$ and

$$\psi_{i,j}(w_i)^\top \mathbf{z}_j = g(a_{j-1}).$$

Since g is Lipschitz continuous, (29) gives

$$\begin{aligned} \left| g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star) - \psi_{i,j}(w_i)^\top \mathbf{z}_j \right| &= \left| g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star) - g(a_{j-1}) \right| \\ &\leq L_g \left(|w_i - a_{j-1}| + |(p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star| \right) \\ &\leq L_g(h + \eta) \leq 2L_g h \leq C_l(h^\beta + \eta^2), \end{aligned}$$

because $h^\beta = h$, $\eta \leq h$, and $C_l \geq 2L_g$.

Suppose now that $\beta > 1$. Applying the Hölder Taylor expansion at a_{j-1} and using (29), we obtain

$$g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star) = \sum_{k=0}^{\varpi(\beta)} \frac{g^{(k)}(a_{j-1})}{k!} \left[(w_i - a_{j-1}) + (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right]^k + R_i, \quad (31)$$

where

$$|R_i| \leq \frac{L_g}{\varpi(\beta)!} \left(|w_i - a_{j-1}| + |(p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star| \right)^\beta \leq \frac{L_g}{\varpi(\beta)!} (h + \eta)^\beta \leq \frac{2^\beta L_g}{\varpi(\beta)!} h^\beta \leq C_l h^\beta.$$

For each $k \geq 1$, applying the binomial formula to the k th power in (31) gives

$$\left[(w_i - a_{j-1}) + (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right]^k = \sum_{r=0}^k \binom{k}{r} (w_i - a_{j-1})^{k-r} ((p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star)^r.$$

The term with $r = 0$ does not contain the pilot-index error, while the term with $r = 1$ is linear in the pilot-index error. After multiplying by the Taylor coefficients and summing over k , these terms are exactly the two terms in (30). Therefore, $\psi_{i,j}(w_i)^\top \mathbf{z}_j$ contains all terms in the Taylor expansion that involve at most one factor of $(p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star$.

The remaining terms correspond to $r \geq 2$. For every $2 \leq r \leq k$, using

$$|w_i - a_{j-1}| \leq h, \quad \left| (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right| \leq \eta, \quad \eta \leq h,$$

we obtain

$$\begin{aligned} |w_i - a_{j-1}|^{k-r} \left| (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right|^r &\leq h^{k-r} \eta^r \leq h^{k-r} \eta^2 h^{r-2} \\ &= h^{k-2} \eta^2 \leq (2B)^{k-2} \eta^2, \end{aligned}$$

where the second inequality follows because $\eta \leq h$ and $r \geq 2$ and the last inequality follows because $h \leq 2B$. Moreover, (24) uniformly bounds the Taylor coefficients by C_g , and the numbers of possible values of k and r depend only on β . Hence,

$$\begin{aligned} \sum_{k=2}^{\varpi(\beta)} \frac{|g^{(k)}(a_{j-1})|}{k!} \sum_{r=2}^k \binom{k}{r} |w_i - a_{j-1}|^{k-r} \left| (p_i - w_i) - \mathbf{x}_i^\top \boldsymbol{\theta}_\star \right|^r &\leq C_g \eta^2 \sum_{k=2}^{\varpi(\beta)} \frac{(2^k - k - 1)(2B)^{k-2}}{k!} \\ &\leq C_l \eta^2, \end{aligned}$$

where we used $\sum_{r=2}^k \binom{k}{r} = 2^k - k - 1$.

Combining this bound with the Taylor remainder $|R_i| \leq C_l h^\beta$, we have

$$g(p_i - \mathbf{x}_i^\top \boldsymbol{\theta}_\star) = \psi_{i,j}(w_i)^\top \mathbf{z}_j + R_{i,j}(w_i), \quad |R_{i,j}(w_i)| \leq C_l (h^\beta + \eta^2). \quad (32)$$

When $1 < \beta \leq 2$, we have $\varpi(\beta) = 1$, so the sums over $k \geq 2$ are empty. In this case, no quadratic or higher-order term in the pilot-index error is omitted, and the additional η^2 term is unnecessary but remains a valid uniform upper bound.

The same argument applies to every current candidate action $w \in \mathcal{W}_{t,j}$. Indeed, $\hat{p}_t(w) \in [0, B]$, so Assumption 3.1 implies

$$\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star \in \mathcal{I}_g.$$

Moreover, $w \in I_j$ and Step 1 give

$$0 \leq w - a_{j-1} \leq h, \quad \left| \hat{p}_t(w) - w - \mathbf{x}_t^\top \boldsymbol{\theta}_\star \right| \leq \eta.$$

Therefore,

$$g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_\star) = \psi_{t,j}(w)^\top \mathbf{z}_j + R_{t,j}(w), \quad |R_{t,j}(w)| \leq C_l (h^\beta + \eta^2). \quad (33)$$

Step 3: ridge-error decomposition. For every $i \in \Psi_{t,s}^j$, Proposition A.1 and (32) give $y_i = \psi_{i,j}(w_i)^\top \mathbf{z}_j + R_{i,j}(w_i) + \varepsilon_i$. Thus each observed demand consists of the linear approximation $\psi_{i,j}(w_i)^\top \mathbf{z}_j$, the approximation error $R_{i,j}(w_i)$, and the demand noise ε_i .

By the definition of $\hat{\mathbf{z}}_{t,s}^j$,

$$(\Lambda_{t,s}^j + \lambda I) \hat{\mathbf{z}}_{t,s}^j = \sum_{i \in \Psi_{t,s}^j} y_i \psi_{i,j}(w_i).$$

Substituting the preceding expression for y_i yields

$$\begin{aligned} (\Lambda_{t,s}^j + \lambda I) \hat{\mathbf{z}}_{t,s}^j &= \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \left(\psi_{i,j}(w_i)^\top \mathbf{z}_j + R_{i,j}(w_i) + \varepsilon_i \right) \\ &= \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \psi_{i,j}(w_i)^\top \mathbf{z}_j + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i. \end{aligned}$$

By the definition

$$\Lambda_{t,s}^j = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \psi_{i,j}(w_i)^\top,$$

the first term on the right-hand side equals $\Lambda_{t,s}^j \mathbf{z}_j$. Hence

$$(\Lambda_{t,s}^j + \lambda I) \hat{\mathbf{z}}_{t,s}^j = \Lambda_{t,s}^j \mathbf{z}_j + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i.$$

Subtracting $(\Lambda_{t,s}^j + \lambda I) \mathbf{z}_j$ from both sides gives

$$(\Lambda_{t,s}^j + \lambda I) (\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j) = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) - \lambda \mathbf{z}_j.$$

The term $-\lambda \mathbf{z}_j$ appears because

$$\Lambda_{t,s}^j \mathbf{z}_j - (\Lambda_{t,s}^j + \lambda I) \mathbf{z}_j = -\lambda \mathbf{z}_j.$$

Since $\lambda > 0$, the matrix $\Lambda_{t,s}^j + \lambda I$ is positive definite and therefore invertible. Multiplying both sides by $(\Lambda_{t,s}^j + \lambda I)^{-1}$ gives

$$\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j = (\Lambda_{t,s}^j + \lambda I)^{-1} \left(\sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i + \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) - \lambda \mathbf{z}_j \right). \quad (34)$$

The three terms on the right-hand side arise, respectively, from the random demand noise, the approximation error in (32), and the ridge regularization.

Step 4: noise, approximation, and regularization bounds. Fix $s \in [S]$ and $j \in [N]$. Proposition A.1 shows that $\bar{\psi}_{i,j,s}$ is determined before y_i is observed and that $\mathbb{E}[\bar{\psi}_{i,j,s} \varepsilon_i \mid \mathcal{F}_i^-] = \mathbf{0}$. Moreover, conditional on $\mathcal{F}_i^-$, ε_i has mean zero and lies in an interval of length D . Hence, by the conditional Hoeffding lemma, for every $\gamma \in \mathbb{R}$,

$$\mathbb{E}[\exp(\gamma \varepsilon_i) \mid \mathcal{F}_i^-] \leq \exp\left(\frac{\gamma^2 D^2}{8}\right).$$

Thus, ε_i is conditionally $D/2$ -sub-Gaussian, and the standard time-uniform self-normalized concentration inequality applies to the features $\bar{\psi}_{i,j,s}$.

By the definition of $\bar{\psi}_{i,j,s}$, for every $t \leq T + 1$,

$$\sum_{i < t} \bar{\psi}_{i,j,s} \varepsilon_i = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i,$$

and

$$\sum_{i < t} \bar{\psi}_{i,j,s} (\bar{\psi}_{i,j,s})^\top = \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \psi_{i,j}(w_i)^\top = \Lambda_{t,s}^j.$$

Hence, for pair (s, j) , with probability at least $1 - \delta/(2SN)$, for every $t \leq T + 1$,

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq D \sqrt{\log \frac{\det(\Lambda_{t,s}^j + \lambda I)}{\det(\lambda I)} + 2 \log \left(\frac{2SN}{\delta} \right)}.$$

The matrix $\Lambda_{t,s}^j$ has dimension $1 + \varpi(\beta)(d + 1)$. Define

$$C_\psi^2 := 1 + \sum_{k=1}^{\varpi(\beta)} \left[(2B)^k + Bk(2B)^{k-1} \right]^2 + C_x^2 \sum_{k=1}^{\varpi(\beta)} k^2 (2B)^{2(k-1)}. \quad (35)$$

Since $0 \leq w_i - a_{j-1} \leq h \leq 2B$, $0 \leq p_i - w_i \leq B$, and $\|\mathbf{x}_i\|_2 \leq C_x$, the definition of the joint feature gives $\|\psi_{i,j}(w_i)\|_2 \leq C_\psi$. The product of positive numbers is no larger than the corresponding power of their average, and hence

$$\begin{aligned} \log \frac{\det(\Lambda_{t,s}^j + \lambda I)}{\det(\lambda I)} &\leq [1 + \varpi(\beta)(d + 1)] \log \left(1 + \frac{\text{tr}(\Lambda_{t,s}^j)}{\lambda[1 + \varpi(\beta)(d + 1)]} \right) \\ &\leq C_\psi^2 [1 + \varpi(\beta)(d + 1)] \log \left(1 + \frac{T}{\lambda} \right) \leq C_\psi^2 \iota_T, \end{aligned}$$

where the second inequality follows because

$$\text{tr}(\Lambda_{t,s}^j) = \sum_{i \in \Psi_{t,s}^j} \|\psi_{i,j}(w_i)\|_2^2 \leq C_\psi^2 |\Psi_{t,s}^j| \leq C_\psi^2 T.$$

It follows that, for this fixed pair (s, j) , with probability at least $1 - \delta/(2SN)$,

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq C_\psi D \sqrt{\iota_T}$$

simultaneously for every $t \leq T + 1$.

There are SN layer-bin pairs. Since the failure probability for each fixed pair is at most $\delta/(2SN)$, the probability that the preceding bound fails for at least one pair is at most

$$SN \frac{\delta}{2SN} = \frac{\delta}{2}.$$

Therefore, there exists an event $\mathcal{E}_{\text{sc}}$ with $\mathbb{P}(\mathcal{E}_{\text{sc}}) \geq 1 - \delta/2$, on which, simultaneously for every $t \leq T + 1$, $s \in [S]$, and $j \in [N]$,

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq C_\psi D \sqrt{\iota_T}. \quad (36)$$

We first control the term involving the approximation errors $R_{i,j}(w_i)$. For any vector a , the Cauchy–Schwarz inequality gives

$$\begin{aligned} \left| a^\top \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right| &= \left| \sum_{i \in \Psi_{t,s}^j} (a^\top \psi_{i,j}(w_i)) R_{i,j}(w_i) \right| \\ &\leq \left[\sum_{i \in \Psi_{t,s}^j} (a^\top \psi_{i,j}(w_i))^2 \right]^{1/2} \times \left[\sum_{i \in \Psi_{t,s}^j} R_{i,j}(w_i)^2 \right]^{1/2}. \end{aligned}$$

By the definition of $\Lambda_{t,s}^j$,

$$\sum_{i \in \Psi_{t,s}^j} (a^\top \psi_{i,j}(w_i))^2 = a^\top \Lambda_{t,s}^j a \leq a^\top (\Lambda_{t,s}^j + \lambda I) a = \|a\|_{\Lambda_{t,s}^j + \lambda I}^2.$$

Therefore,

$$\left| a^\top \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right| \leq \|a\|_{\Lambda_{t,s}^j + \lambda I} \left[\sum_{i \in \Psi_{t,s}^j} R_{i,j}(w_i)^2 \right]^{1/2}.$$

Taking

$$a = (\Lambda_{t,s}^j + \lambda I)^{-1} \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i)$$

in the preceding inequality yields

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}^2 \leq \left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \left[\sum_{i \in \Psi_{t,s}^j} R_{i,j}(w_i)^2 \right]^{1/2}.$$

Canceling the common factor when it is nonzero, while the zero case is immediate, gives

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq \left[\sum_{i \in \Psi_{t,s}^j} R_{i,j}(w_i)^2 \right]^{1/2}.$$

Since $|R_{i,j}(w_i)| \leq C_l(h^\beta + \eta^2)$, for every $i \in \Psi_{t,s}^j$, we have

$$\sum_{i \in \Psi_{t,s}^j} R_{i,j}(w_i)^2 \leq C_l^2 |\Psi_{t,s}^j| (h^\beta + \eta^2)^2.$$

It follows that

$$\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq C_l \sqrt{|\Psi_{t,s}^j|} (h^\beta + \eta^2). \quad (37)$$

It remains to control the term caused by the ridge regularization. Since $\Lambda_{t,s}^j + \lambda I \succeq \lambda I$, we have

$$(\Lambda_{t,s}^j + \lambda I)^{-1} \preceq \frac{1}{\lambda} I.$$

Therefore,

$$\|\lambda \mathbf{z}_j\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}^2 = \lambda^2 \mathbf{z}_j^\top (\Lambda_{t,s}^j + \lambda I)^{-1} \mathbf{z}_j \leq \lambda \|\mathbf{z}_j\|_2^2.$$

Taking square roots and using (26) gives

$$\|\lambda \mathbf{z}_j\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \leq \sqrt{\lambda} \|\mathbf{z}_j\|_2 \leq C_z \sqrt{\lambda}. \quad (38)$$

Step 5: uniform confidence bound and completion. We work on the event $\mathcal{E}_{\text{pil}} \cap \mathcal{E}_{\text{sc}}$, which has probability at least $1 - \delta$. Fix a main-policy round t , a layer $s \in [s_t]$, and an action $(j, w) \in \mathcal{A}_{t,s}$. Assume first that $|\Psi_{t,s}^j| \geq 1$.

By (34) and multiplying both sides by $\psi_{t,j}(w)^\top$, taking absolute values, and applying the triangle inequality gives

$$\begin{aligned} \left| \psi_{t,j}(w)^\top (\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j) \right| &\leq \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \left[\left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) \varepsilon_i \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \right. \\ &\quad \left. + \left\| \sum_{i \in \Psi_{t,s}^j} \psi_{i,j}(w_i) R_{i,j}(w_i) \right\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} + \|\lambda \mathbf{z}_j\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \right]. \end{aligned}$$

Using (36), (37), and (38), we obtain

$$\begin{aligned} \left| \psi_{t,j}(w)^\top (\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j) \right| &\leq (C_\psi D \sqrt{\iota_T} + C_z \sqrt{\lambda}) \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \\ &\quad + C_l (h^\beta + \eta^2) \sqrt{|\Psi_{t,s}^j|} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}}. \end{aligned}$$

The preceding inequality controls the error caused by replacing $\mathbf{z}_j$ with its estimator $\hat{\mathbf{z}}_{t,s}^j$. We must also account for the approximation error at the current action. By (33),

$$g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) = \psi_{t,j}(w)^\top \mathbf{z}_j + R_{t,j}(w),$$

where $|R_{t,j}(w)| \leq C_l (h^\beta + \eta^2)$. Therefore,

$$\begin{aligned} \left| \psi_{t,j}(w)^\top \hat{\mathbf{z}}_{t,s}^j - g(\hat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| &= \left| \psi_{t,j}(w)^\top (\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j) - R_{t,j}(w) \right| \\ &\leq \left| \psi_{t,j}(w)^\top (\hat{\mathbf{z}}_{t,s}^j - \mathbf{z}_j) \right| + |R_{t,j}(w)| \\ &\leq (C_\psi D \sqrt{\iota_T} + C_z \sqrt{\lambda}) \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \\ &\quad + C_l (h^\beta + \eta^2) \left[1 + \sqrt{|\Psi_{t,s}^j|} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \right]. \end{aligned}$$

The algorithm restricts the estimated demand to the interval $[0, D]$:

$$\widehat{g}_{t,s}^j(w) = \Pi_{[0,D]}(\psi_{t,j}(w)^\top \widehat{\mathbf{z}}_{t,s}^j).$$

This restriction cannot increase the error because the true mean demand also belongs to $[0, D]$. Indeed, for every $a \in \mathbb{R}$ and every $b \in [0, D]$, $|\Pi_{[0,D]}(a) - b| \leq |a - b|$. Moreover, both $\Pi_{[0,D]}(a)$ and b belong to $[0, D]$, so $|\Pi_{[0,D]}(a) - b| \leq D$. Applying these two inequalities with $a = \psi_{t,j}(w)^\top \widehat{\mathbf{z}}_{t,s}^j$ and $b = g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*)$ gives

$$\begin{aligned} & \left| \widehat{g}_{t,s}^j(w) - g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| \\ & \leq \min \left\{ D, \left(C_\psi D \sqrt{\nu_T} + C_z \sqrt{\lambda} \right) \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \right. \\ & \quad \left. + C_t(h^\beta + \eta^2) \left[1 + \sqrt{|\Psi_{t,s}^j|} \|\psi_{t,j}(w)\|_{(\Lambda_{t,s}^j + \lambda I)^{-1}} \right] \right\} \leq r_{t,s}^j(w). \end{aligned}$$

It remains to consider the case $|\Psi_{t,s}^j| = 0$. Then $\Psi_{t,s}^j = \emptyset$ and $\Lambda_{t,s}^j = \mathbf{0}$. The optimization problem reduces to $\min_{\mathbf{z}} \lambda \|\mathbf{z}\|_2^2$. Since $\lambda > 0$, its unique minimizer is $\widehat{\mathbf{z}}_{t,s}^j = \mathbf{0}$. It follows that $\widehat{g}_{t,s}^j(w) = \Pi_{[0,D]}(0) = 0$. The algorithm sets $r_{t,s}^j(w) = D$. Since the true mean demand belongs to $[0, D]$,

$$\left| \widehat{g}_{t,s}^j(w) - g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| = \left| g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| \leq D = r_{t,s}^j(w).$$

Finally, the event $\mathcal{E}_{\text{pil}}$ controls all periods simultaneously, and the event $\mathcal{E}_{\text{sc}}$ controls all periods, layers, and bins simultaneously. Once these two events occur, the preceding inequalities can be applied to every residual action w because they use only the already established vector bounds and the Cauchy–Schwarz inequality. Thus no separate probability bound is needed for each w . Therefore, (9) holds simultaneously for every main-policy round t , every $s \in [s_t]$, and every $(j, w) \in \mathcal{A}_{t,s}$. $\square$

A.3 Proofs for the Contextual Regret Bounds

Proof of Theorem 5.1.

Proof. We proceed in four steps.

Step 1: Revenue-UCB sandwich. Fix a visited layer s and an action $(j, w) \in \mathcal{A}_{t,s}$. On the confidence event, $\left| \widehat{g}_{t,s}^j(w) - g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \right| \leq r_{t,s}^j(w)$. Because the true conditional mean belongs to $[0, D]$,

$$g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) \leq \min \left\{ D, \widehat{g}_{t,s}^j(w) + r_{t,s}^j(w) \right\} \leq g(\widehat{p}_t(w) - \mathbf{x}_t^\top \boldsymbol{\theta}_*) + 2r_{t,s}^j(w).$$

Multiplying by $\widehat{p}_t(w) \geq 0$ and using the definitions of $U_{t,s}^j(w)$ and $\text{wid}_{t,s}^j(w)$ gives

$$\text{Rev}(\mathbf{x}_t, \widehat{p}_t(w)) \leq U_{t,s}^j(w) \leq \text{Rev}(\mathbf{x}_t, \widehat{p}_t(w)) + 2 \text{wid}_{t,s}^j(w). \quad (39)$$

Step 2: Preservation of the discrete benchmark. Suppose that the precision check passes at a layer $s < S$. Then

$$\text{wid}_{t,s}^j(w) \leq BD 2^{-s} \quad \text{for every } (j, w) \in \mathcal{A}_{t,s}.$$

Choose $(j^*, w^*) \in \arg\max_{(j,w) \in \mathcal{A}_{t,s}} \text{Rev}(\mathbf{x}_t, \hat{p}_t(w))$. By the lower side of (39), $U_{t,s}^{j^*}(w^*) \geq V_{t,s}$. By its upper side, for every $(j, w) \in \mathcal{A}_{t,s}$,

$$U_{t,s}^j(w) \leq V_{t,s} + 2BD 2^{-s} = V_{t,s} + BD 2^{1-s}.$$

Consequently,

$$U_{t,s}^{j^*}(w^*) \geq \max_{(j,w) \in \mathcal{A}_{t,s}} U_{t,s}^j(w) - BD 2^{1-s}.$$

By the definition of $\mathcal{A}_{t,s+1}$ in Algorithm 2, the preceding inequality implies that $(j^*, w^*) \in \mathcal{A}_{t,s+1}$. Since (j^*, w^*) attains $V_{t,s}$, it follows that

$$V_{t,s+1} \geq \text{Rev}(\mathbf{x}_t, \hat{p}_t(w^*)) = V_{t,s}.$$

On the other hand, $\mathcal{A}_{t,s+1} \subseteq \mathcal{A}_{t,s}$, and hence $V_{t,s+1} \leq V_{t,s}$. Therefore, $V_{t,s+1} = V_{t,s}$.

Step 3: Gap between the selected price and the initial discrete benchmark. Suppose that $s_t \geq 2$. To reach layer s_t , the algorithm must have passed the precision check at layer $s_t - 1$. Moreover, $(j_t, w_t) \in \mathcal{A}_{t,s_t}$, so this action survived the refinement from layer $s_t - 1$ to layer s_t . Hence

$$U_{t,s_t-1}^{j_t}(w_t) \geq \max_{(j,w) \in \mathcal{A}_{t,s_t-1}} U_{t,s_t-1}^j(w) - BD 2^{2-s_t} \geq V_{t,s_t-1} - BD 2^{2-s_t}. \quad (40)$$

The upper side of (39) and the precision check at layer $s_t - 1$ give

$$U_{t,s_t-1}^{j_t}(w_t) \leq \text{Rev}(\mathbf{x}_t, p_t) + 2BD 2^{-(s_t-1)} = \text{Rev}(\mathbf{x}_t, p_t) + BD 2^{2-s_t}. \quad (41)$$

Combining (40) and (41) yields $V_{t,s_t-1} - \text{Rev}(\mathbf{x}_t, p_t) \leq 2BD 2^{2-s_t}$. Every layer preceding s_t passed the precision check. Repeated application of Step 2 therefore gives $V_{t,s_t-1} = V_{t,1}$, which proves the second assertion of the lemma.

Step 4: Discretization of the continuous oracle price. If $\beta = 1$, Assumption 3.4 implies that g is L_g -Lipschitz. If $\beta > 1$, Step 2 of the proof of Proposition 4.2 gives $\sup_{u \in \mathcal{I}_g} |g'(u)| \leq C_g$. Hence g is Lipschitz on $\mathcal{I}_g$ with constant $\max\{L_g, C_g\}$. Since both residual arguments below belong to $\mathcal{I}_g$ by Assumption 3.1, for every $\mathbf{x} \in \mathcal{X}$ and $p, p' \in [0, B]$,

$$|\text{Rev}(\mathbf{x}, p) - \text{Rev}(\mathbf{x}, p')| \leq D|p - p'| + B \left| g(p - \mathbf{x}^\top \boldsymbol{\theta}_*) - g(p' - \mathbf{x}^\top \boldsymbol{\theta}_*) \right| \leq L_{\text{Rev}} |p - p'|,$$

where one may take $L_{\text{Rev}} = D + B \max\{L_g, C_g\}$.

The oracle residual $p_t^* - \Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t)$ belongs to $[-B, B]$. The feasible residual set is the interval $[-\Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t), B - \Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t)]$, which contains both zero and the oracle residual. By choosing a point of $\mathcal{W}$ between zero and the oracle residual, the mesh definition of $\mathcal{W}$, together with the explicitly included endpoints, gives a feasible $w \in \mathcal{W}$ satisfying

$$\left| w - [p_t^* - \Pi_{[0,B]}(\mathbf{x}_t^\top \hat{\boldsymbol{\theta}}_t)] \right| \leq T^{-1/2}.$$

Assigning this w to its unique bin yields an action in $\mathcal{A}_{t,1}$, and $|\hat{p}_t(w) - p_t^*| \leq T^{-1/2}$. Hence

$$\text{Rev}(\mathbf{x}_t, p_t^*) - V_{t,1} \leq \frac{L_{\text{Rev}}}{\sqrt{T}}.$$

If $s_t \geq 2$, combining this inequality with the bound established in Step 3 proves (10). If $s_t = 1$, then $0 \leq \text{Rev}(\mathbf{x}_t, p) \leq BD$ for every feasible price, so

$$\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t) \leq \frac{L_{\text{Rev}}}{\sqrt{T}} + V_{t,1} - \text{Rev}(\mathbf{x}_t, p_t) \leq \frac{L_{\text{Rev}}}{\sqrt{T}} + BD \leq \frac{L_{\text{Rev}}}{\sqrt{T}} + 8BD 2^{-s_t}.$$

Thus the same conclusion holds for $s_t = 1$. $\square$

Proof of counting lemmas.

Lemma A.2 (Cumulative revenue uncertainty within one layer) *There exists a finite constant $C_\omega \geq 1$, such that, for every $s \in [S]$, pathwise,*

$$\sum_{t \in \Psi_{T+1,s}} \text{wid}_{t,s}^{j_t}(w_t) \leq C_\omega B \left[(D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{N|\Psi_{T+1,s}| \iota_T} + (h^\beta + \eta^2) |\Psi_{T+1,s}| \sqrt{\iota_T} \right]. \quad (42)$$

The right-hand side is interpreted as zero when $\Psi_{T+1,s} = \emptyset$.

Proof. Fix $s \in [S]$. The claim is immediate if $\Psi_{T+1,s} = \emptyset$. We first work within one nonempty bin and then sum over bins.

Step 1: Sequential Gram matrices within a permanent bin. Fix $j \in [N]$ with $\Psi_{T+1,s}^j \neq \emptyset$, and enumerate its observations chronologically as

$$\Psi_{T+1,s}^j = \{t_{j,1} < \dots < t_{j,|\Psi_{T+1,s}^j|}\}.$$

Permanent labeling and the definition of $\Lambda_{t,s}^j$ imply that, for every $k \leq |\Psi_{T+1,s}^j|$,

$$\Lambda_{t_{j,k},s}^j + \lambda I = \lambda I + \sum_{\ell < k} \psi_{t_{j,\ell},j}(w_{t_{j,\ell}}) \psi_{t_{j,\ell},j}(w_{t_{j,\ell}})^\top, \quad |\Psi_{t_{j,k},s}^j| = k - 1. \quad (43)$$

Step 2: Elliptical-potential bound. The uniform feature bound established in the proof of Proposition 4.2 gives $\|\psi_{t,j}(w_t)\|_2 \leq C_\psi$. Applying the matrix determinant lemma sequentially to (43), and using $\min\{1, x\} \leq 2 \log(1+x)$ for $x \geq 0$, yields

$$\begin{aligned} & \sum_{k=1}^{|\Psi_{T+1,s}^j|} \min \left\{ 1, \left\| \psi_{t_{j,k},j}(w_{t_{j,k}}) \right\|_{(\Lambda_{t_{j,k},s}^j + \lambda I)^{-1}}^2 \right\} \\ & \leq 2 \log \frac{\det \left(\lambda I + \sum_{k=1}^{|\Psi_{T+1,s}^j|} \psi_{t_{j,k},j}(w_{t_{j,k}}) \psi_{t_{j,k},j}(w_{t_{j,k}})^\top \right)}{\det(\lambda I)} \\ & \leq 2 [1 + \varpi(\beta)(d+1)] \log \left(1 + \frac{C_\psi^2 |\Psi_{T+1,s}^j|}{\lambda [1 + \varpi(\beta)(d+1)]} \right) \\ & \leq 2 \max\{1, C_\psi^2\} \iota_T. \end{aligned} \quad (44)$$

The last inequality follows from $|\Psi_{T+1,s}^j| \leq T$, the definition of ι_T , and $\log(1+ax) \leq \max\{1, a\} \log(1+x)$ for $a, x \geq 0$.

By Cauchy–Schwarz and (44),

$$\begin{aligned} & \sum_{k=1}^{|\Psi_{T+1,s}^j|} \min \left\{ 1, \left\| \psi_{t_{j,k},j}(w_{t_{j,k}}) \right\|_{(\Lambda_{t_{j,k},s}^j + \lambda I)^{-1}} \right\} \\ & \leq \sqrt{|\Psi_{T+1,s}^j| \sum_{k=1}^{|\Psi_{T+1,s}^j|} \min \left\{ 1, \left\| \psi_{t_{j,k},j}(w_{t_{j,k}}) \right\|_{(\Lambda_{t_{j,k},s}^j + \lambda I)^{-1}}^2 \right\}} \leq \sqrt{2 \max\{1, C_\psi^2\}} \sqrt{|\Psi_{T+1,s}^j| \iota_T}. \end{aligned} \quad (45)$$

Step 3: Sum of confidence radii within one bin. The first observation in the bin satisfies $|\Psi_{t_{j,1},s}^j| = 0$, and therefore (7) gives $r_{t_{j,1},s}^j(w_{t_{j,1}}) = D$. For $k \geq 2$, we have $|\Psi_{t_{j,k},s}^j| = k - 1 \leq |\Psi_{T+1,s}^j|$. Therefore, the radius definition and its truncation at D imply

$$\begin{aligned} r_{t_{j,k},s}^j(w_{t_{j,k}}) & \leq \max\{C_\psi, C_z\} (D\sqrt{\iota_T} + \sqrt{\lambda}) \min \left\{ 1, \left\| \psi_{t_{j,k},j}(w_{t_{j,k}}) \right\|_{(\Lambda_{t_{j,k},s}^j + \lambda I)^{-1}} \right\} \\ & \quad + C_l(h^\beta + \eta^2) \left[1 + \sqrt{|\Psi_{T+1,s}^j|} \min \left\{ 1, \left\| \psi_{t_{j,k},j}(w_{t_{j,k}}) \right\|_{(\Lambda_{t_{j,k},s}^j + \lambda I)^{-1}} \right\} \right]. \end{aligned} \quad (46)$$

Summing (46) over $k \geq 2$, adding the first radius, and applying (45) gives

$$\begin{aligned} \sum_{t \in \Psi_{T+1,s}^j} r_{t,s}^j(w_t) & \leq D + \max\{C_\psi, C_z\} \sqrt{2 \max\{1, C_\psi^2\}} (D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{|\Psi_{T+1,s}^j| \iota_T} \\ & \quad + C_l(h^\beta + \eta^2) \left[|\Psi_{T+1,s}^j| + \sqrt{2 \max\{1, C_\psi^2\}} |\Psi_{T+1,s}^j| \sqrt{\iota_T} \right]. \end{aligned}$$

Since $|\Psi_{T+1,s}^j| \geq 1$, $\iota_T > 1$, and $D\sqrt{\iota_T} + \sqrt{\lambda} \geq D$, the first and third terms can be absorbed, yielding

$$\begin{aligned} \sum_{t \in \Psi_{T+1,s}^j} r_{t,s}^j(w_t) & \leq \left[1 + \max\{C_\psi, C_z\} \sqrt{2 \max\{1, C_\psi^2\}} \right] \cdot (D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{|\Psi_{T+1,s}^j| \iota_T} \\ & \quad + C_l \left[1 + \sqrt{2 \max\{1, C_\psi^2\}} \right] (h^\beta + \eta^2) |\Psi_{T+1,s}^j| \sqrt{\iota_T}. \end{aligned} \quad (47)$$

Step 4: Sum over bins and convert demand radii to revenue widths. By Cauchy–Schwarz,

$$\sum_{j=1}^N \sqrt{|\Psi_{T+1,s}^j|} \leq \sqrt{N \sum_{j=1}^N |\Psi_{T+1,s}^j|} = \sqrt{N |\Psi_{T+1,s}|},$$

and, by the permanent partition,

$$\sum_{j=1}^N |\Psi_{T+1,s}^j| = |\Psi_{T+1,s}|.$$

By choosing

$$C_\omega = \max \left\{ 1 + \max\{C_\psi, C_z\} \sqrt{2 \max\{1, C_\psi^2\}}, C_l \left(1 + \sqrt{2 \max\{1, C_\psi^2\}} \right) \right\},$$

both coefficients in (47) are bounded by C_ω . Summing over bins and using

$$\text{wid}_{t,s}^{j_t}(w_t) = \widehat{p}_t(w_t)r_{t,s}^{j_t}(w_t) \leq Br_{t,s}^{j_t}(w_t),$$

therefore proves (42). $\square$

Lemma A.3 (Occupancy of a nonterminal LDP layer) *Let C_ω be as in Lemma A.2. For every $s < S$, pathwise,*

$$D2^{-s}|\Psi_{T+1,s}| \leq C_\omega \left[(D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{N|\Psi_{T+1,s}| \iota_T} + (h^\beta + \eta^2) |\Psi_{T+1,s}| \sqrt{\iota_T} \right]. \quad (48)$$

Moreover, if

$$D2^{-s} \geq 2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}, \quad (49)$$

then

$$|\Psi_{T+1,s}| \leq 4C_\omega^2 \frac{(D\sqrt{\iota_T} + \sqrt{\lambda})^2}{D^2} 2^{2s} N \iota_T. \quad (50)$$

Proof. Fix $s < S$. If $\Psi_{T+1,s} = \emptyset$, both conclusions are immediate. Suppose that $\Psi_{T+1,s} \neq \emptyset$.

Every round $t \in \Psi_{T+1,s}$ is assigned to the nonterminal layer s . Under Algorithm 2, selection at such a layer can occur only through the under-explored branch. Therefore, $\text{wid}_{t,s}^{j_t}(w_t) > BD2^{-s}$ for $t \in \Psi_{T+1,s}$. Summing this inequality over $t \in \Psi_{T+1,s}$ and applying Lemma A.2 gives

$$\begin{aligned} BD2^{-s}|\Psi_{T+1,s}| &< \sum_{t \in \Psi_{T+1,s}} \text{wid}_{t,s}^{j_t}(w_t) \\ &\leq C_\omega B \left[(D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{N|\Psi_{T+1,s}| \iota_T} + (h^\beta + \eta^2) |\Psi_{T+1,s}| \sqrt{\iota_T} \right]. \end{aligned}$$

Since $B > 0$, dividing by B and weakening the resulting strict inequality proves (48).

Suppose now that (49) holds. Then

$$C_\omega(h^\beta + \eta^2)\sqrt{\iota_T} \leq \frac{D}{2}2^{-s}.$$

It follows from (48) that

$$\frac{D}{2}2^{-s}|\Psi_{T+1,s}| \leq \left[D2^{-s} - C_\omega(h^\beta + \eta^2)\sqrt{\iota_T} \right] |\Psi_{T+1,s}| \leq C_\omega(D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{N|\Psi_{T+1,s}| \iota_T}.$$

Both sides are nonnegative. Squaring this inequality gives

$$\frac{D^2}{4}2^{-2s}|\Psi_{T+1,s}|^2 \leq C_\omega^2(D\sqrt{\iota_T} + \sqrt{\lambda})^2 N |\Psi_{T+1,s}| \iota_T.$$

Because $|\Psi_{T+1,s}| > 0$, canceling this factor and rearranging yields

$$|\Psi_{T+1,s}| \leq 4C_\omega^2 \frac{(D\sqrt{\iota_T} + \sqrt{\lambda})^2}{D^2} 2^{2s} N \iota_T,$$

which proves (50). $\square$

Proof of Theorem 5.2.

Proof. We proceed in five steps. Throughout the proof, C denotes a finite constant depending only on the fixed problem primitives and on C_ψ, C_z, C_l ; its value may increase from line to line.

Step 1: Reduction to a dyadic layer sum. Note that, the sets $\Psi_{T+1,1}, \dots, \Psi_{T+1,S}$ are disjoint and contain all LDP rounds. Therefore,

$$\sum_{s=1}^S |\Psi_{T+1,s}| \leq T. \quad (51)$$

By Lemma 5.1, every LDP round t satisfies

$$\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t) \leq \frac{L_{\text{Rev}}}{\sqrt{T}} + 8BD 2^{-s_t}.$$

Summing over the LDP rounds and using their layer partition gives

$$\sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq L_{\text{Rev}} \sqrt{T} + 8BD \sum_{s=1}^S 2^{-s} |\Psi_{T+1,s}|. \quad (52)$$

Step 2: Nonterminal layers dominated by statistical uncertainty. For every nonterminal layer $s < S$ satisfying (49), Lemma A.3 gives

$$|\Psi_{T+1,s}| \leq 4C_\omega^2 \frac{(D\sqrt{\iota_T} + \sqrt{\lambda})^2}{D^2} 2^{2s} N \iota_T.$$

The deterministic bound $|\Psi_{T+1,s}| \leq T$ also holds. Note that for every $a > 0$, we have

$$\sum_{s \geq 1} 2^{-s} \min\{T, a2^{2s}\} \leq 4\sqrt{aT}, \quad (53)$$

because if $a \geq T$, the left-hand side is at most $T \sum_{s \geq 1} 2^{-s} \leq T \leq \sqrt{aT}$, while if $a < T$, choose an integer $s_0 \geq 0$ such that $2^{s_0} \leq \sqrt{T/a} < 2^{s_0+1}$, then

$$\sum_{s \geq 1} 2^{-s} \min\{T, a2^{2s}\} \leq a \sum_{s \leq s_0} 2^s + T \sum_{s > s_0} 2^{-s} \leq 2a2^{s_0} + T2^{-s_0} \leq 4\sqrt{aT}.$$

Applying (53) with

$$a = 4C_\omega^2 \frac{(D\sqrt{\iota_T} + \sqrt{\lambda})^2}{D^2} N \iota_T$$

yields

$$\sum_{s < S: D2^{-s} \geq 2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}} 2^{-s} |\Psi_{T+1,s}| \leq 4\sqrt{aT} = 8C_\omega \frac{D\sqrt{\iota_T} + \sqrt{\lambda}}{D} \sqrt{NT\iota_T}. \quad (54)$$

Step 3: Nonterminal layers dominated by approximation bias. For every remaining nonterminal layer,

$$D2^{-s} < 2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T},$$

and hence

$$2^{-s} < \frac{2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}}{D}.$$

Consequently,

$$\sum_{\substack{s < S: \\ D2^{-s} < 2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}}} 2^{-s} |\Psi_{T+1,s}| \leq \frac{2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}}{D} \sum_{s < S} |\Psi_{T+1,s}| \leq \frac{2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}}{D} T. \quad (55)$$

Step 4: Terminal layer and completion of the explicit bound. For the terminal layer,

$$2^{-S}|\Psi_{T+1,S}| \leq T2^{-S}. \quad (56)$$

Combining (54), (55), and (56), we obtain

$$\sum_{s=1}^S 2^{-s}|\Psi_{T+1,s}| \leq 8C_\omega \frac{D\sqrt{\iota_T} + \sqrt{\lambda}}{D} \sqrt{NT\iota_T} + \frac{2C_\omega(h^\beta + \eta^2)\sqrt{\iota_T}}{D} T + T2^{-S}.$$

Substituting this inequality into (52) gives

$$\begin{aligned} & \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \\ & \leq 64C_\omega B (D\sqrt{\iota_T} + \sqrt{\lambda}) \sqrt{NT\iota_T} + 16C_\omega B T (h^\beta + \eta^2) \sqrt{\iota_T} + 8BDT2^{-S} + L_{\text{Rev}}\sqrt{T}. \end{aligned}$$

For $S = \max \{1, \lceil \log_2 \sqrt{T} \rceil\}$, we have $T2^{-S} \leq \sqrt{T}$. Therefore, (11) follows.

Step 5: Simplified rate under algorithmic tuning. Let $\lambda > 0$ be fixed independently of T , let $\eta^2 \leq c_\eta h^\beta$, and let $N = \lceil T^{\frac{1}{2\beta+1}} \rceil$. Since $T \geq 1$, we have $T^{\frac{1}{2\beta+1}} \leq N \leq 2T^{\frac{1}{2\beta+1}}$. It follows that $\sqrt{NT} \leq \sqrt{2}T^{\frac{\beta+1}{2\beta+1}}$, and, because $h = 2B/N$,

$$T(h^\beta + \eta^2) \leq (1 + c_\eta)Th^\beta = (1 + c_\eta)(2B)^\beta TN^{-\beta} \leq (1 + c_\eta)(2B)^\beta T^{\frac{\beta+1}{2\beta+1}}.$$

Moreover, $\sqrt{T} \leq T^{\frac{\beta+1}{2\beta+1}}$ for every finite $\beta \geq 1$. By the definition of ι_T , we have $\iota_T > 1$. Since $\lambda > 0$ is fixed independently of T ,

$$(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{\iota_T} = D\iota_T + \sqrt{\lambda\iota_T} \leq (D + \sqrt{\lambda})\iota_T,$$

and $\sqrt{\iota_T} \leq \iota_T$.

Applying these inequalities to (11) gives

$$\begin{aligned} & \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \\ & \leq \left[64\sqrt{2}C_\omega B (D + \sqrt{\lambda}) + 16C_\omega B(1 + c_\eta)(2B)^\beta + 8BD + L_{\text{Rev}} \right] \iota_T T^{\frac{\beta+1}{2\beta+1}}. \end{aligned}$$

One may take

$$C_{\text{LDP}} := 64\sqrt{2}C_\omega B (D + \sqrt{\lambda}) + 16C_\omega B(1 + c_\eta)(2B)^\beta + 8BD + L_{\text{Rev}}.$$

This proves (12). □

Proof of Theorem 5.3.

Proof. Fix an arbitrary realization of the policy. If $\mathcal{T}^{\text{exp}} = \emptyset$, all conclusions are immediate. Otherwise, enumerate the pilot-exploration rounds chronologically as

$$\mathcal{T}^{\text{exp}} = \{\tau_1 < \dots < \tau_{|\mathcal{T}^{\text{exp}}|}\}.$$

Step 1: Leverage on exploration rounds. For every $k = 1, \dots, |\mathcal{T}^{\text{exp}}|$, period τ_k is a pilot-exploration round. Therefore, by the uncertainty gate in Algorithm 1, $\gamma_T \|\mathbf{x}_{\tau_k}\|_{M_{\tau_k}^{-1}} > \eta$. Since $\gamma_T > 0$, squaring both sides gives

$$\|\mathbf{x}_{\tau_k}\|_{M_{\tau_k}^{-1}}^2 > \frac{\eta^2}{\gamma_T^2}. \quad (57)$$

Step 2: Exact determinant recursion. The matrix M_t is updated only on pilot-exploration rounds. In particular, $M_{\tau_1} = M_1 = I_d$, and, for every $k = 1, \dots, |\mathcal{T}^{\text{exp}}| - 1$,

$$M_{\tau_{k+1}} = M_{\tau_k} + \mathbf{x}_{\tau_k} \mathbf{x}_{\tau_k}^\top,$$

because M_t remains unchanged on all LDP rounds strictly between τ_k and τ_{k+1} . Similarly,

$$M_{T+1} = M_{\tau_{|\mathcal{T}^{\text{exp}}|}} + \mathbf{x}_{\tau_{|\mathcal{T}^{\text{exp}}|}} \mathbf{x}_{\tau_{|\mathcal{T}^{\text{exp}}|}}^\top.$$

Since $M_{\tau_k} \succeq I_d$, it is positive definite. The matrix determinant lemma therefore gives

$$\det(M_{\tau_k} + \mathbf{x}_{\tau_k} \mathbf{x}_{\tau_k}^\top) = \det(M_{\tau_k}) \left(1 + \mathbf{x}_{\tau_k}^\top M_{\tau_k}^{-1} \mathbf{x}_{\tau_k}\right) = \det(M_{\tau_k}) \left(1 + \|\mathbf{x}_{\tau_k}\|_{M_{\tau_k}^{-1}}^2\right).$$

Iterating this identity and using $\det(M_{\tau_1}) = \det(I_d) = 1$ yields

$$\log \det(M_{T+1}) = \sum_{k=1}^{|\mathcal{T}^{\text{exp}}|} \log \left(1 + \|\mathbf{x}_{\tau_k}\|_{M_{\tau_k}^{-1}}^2\right). \quad (58)$$

Step 3: Lower bound on the determinant growth. By (57), every summand in (58) satisfies

$$\log \left(1 + \|\mathbf{x}_{\tau_k}\|_{M_{\tau_k}^{-1}}^2\right) > \log \left(1 + \frac{\eta^2}{\gamma_T^2}\right).$$

Summing over k gives,

$$|\mathcal{T}^{\text{exp}}| \log \left(1 + \frac{\eta^2}{\gamma_T^2}\right) < \log \det(M_{T+1}). \quad (59)$$

Step 4: Upper bound on the determinant. By the update rule,

$$M_{T+1} = I_d + \sum_{t \in \mathcal{T}^{\text{exp}}} \mathbf{x}_t \mathbf{x}_t^\top.$$

Therefore, we have

$$\text{tr}(M_{T+1}) = d + \sum_{t \in \mathcal{T}^{\text{exp}}} \|\mathbf{x}_t\|_2^2 \leq d + TC_x^2.$$

Since M_{T+1} is positive definite, by the arithmetic-geometric mean inequality applied to the eigenvalues of M_{T+1} ,

$$\det(M_{T+1})^{1/d} \leq \frac{\text{tr}(M_{T+1})}{d}.$$

Consequently,

$$\log \det(M_{T+1}) \leq d \log \left(\frac{\text{tr}(M_{T+1})}{d}\right) \leq d \log \left(1 + \frac{TC_x^2}{d}\right). \quad (60)$$

Combining (59) and (60) yields

$$|\mathcal{T}^{\text{exp}}| < \frac{d \log(1 + TC_x^2/d)}{\log(1 + \eta^2/\gamma_T^2)}.$$

Together with the trivial bound $|\mathcal{T}^{\text{exp}}| \leq T$, this proves the first inequality in (13).

For every $x \geq 0$, $\log(1 + x) \geq x/(1 + x)$. Applying this inequality with $x = \eta^2/\gamma_T^2$ gives

$$\frac{1}{\log(1 + \eta^2/\gamma_T^2)} \leq 1 + \frac{\gamma_T^2}{\eta^2}.$$

This proves the second inequality in (13).

Step 5: Pilot-exploration regret. Since $0 \leq g(u) \leq D$ and $0 \leq p \leq B$, $0 \leq \text{Rev}(\mathbf{x}, p) = p g(p - \mathbf{x}^\top \boldsymbol{\theta}_*) \leq BD$. Therefore, for every $t \in \mathcal{T}^{\text{exp}}$,

$$0 \leq \text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t) \leq \text{Rev}(\mathbf{x}_t, p_t^*) \leq BD.$$

Summing over $t \in \mathcal{T}^{\text{exp}}$ and applying (13) proves (14). $\square$

Proof of Theorem 5.4.

Proof. We work throughout on the uniform confidence event in Proposition 4.2, which occurs with probability at least $1 - \delta$.

Step 1: Decomposition of the regret. The pilot-exploration rounds and the LDP rounds form a disjoint partition of the horizon. Therefore,

$$\text{Reg}(T) = \sum_{t \in \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] + \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)]. \quad (61)$$

Step 2: Finite-sample bound. Lemma 5.3 gives the pathwise

$$\sum_{t \in \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq BD \min \left\{ T, \frac{d \log(1 + TC_x^2/d)}{\log(1 + \eta^2/\gamma_T^2)} \right\}. \quad (62)$$

Proposition 5.2 and the choice $S = \max \{1, \lceil \log_2 \sqrt{T} \rceil\}$ give

$$\begin{aligned} & \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \\ & \leq 64C_\omega B(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{NT\iota_T} + 16C_\omega BT(h^\beta + \eta^2)\sqrt{\iota_T} + (8BD + L_{\text{Rev}})\sqrt{T}. \end{aligned} \quad (63)$$

Substituting (62) and (63) into (61) proves (15).

Step 3: LDP regret under the prescribed tuning. Let $N = \lceil T^{\frac{1}{2\beta+1}} \rceil$ and $\eta^2 = \min\{h^2, h^\beta\}$. Because $\eta^2 \leq h^2$, we have $\eta \leq h$. Moreover, since $\eta^2 \leq h^\beta$, we have $h^\beta + \eta^2 \leq 2h^\beta$. Thus the prescribed tuning satisfies the conditions used in Proposition 4.2 and Proposition 5.2.

Since $T \geq 1$, we have

$$T^{\frac{1}{2\beta+1}} \leq N \leq T^{\frac{1}{2\beta+1}} + 1 \leq 2T^{\frac{1}{2\beta+1}}.$$

It follows that

$$\sqrt{NT} \leq \sqrt{2} T^{\frac{1}{2} + \frac{1}{2(2\beta+1)}} = \sqrt{2} T^{\frac{\beta+1}{2\beta+1}}. \quad (64)$$

Since $N \geq T^{1/(2\beta+1)}$ and $h = 2B/N$,

$$Th^\beta = T \left(\frac{2B}{N} \right)^\beta \leq (2B)^\beta T^{1 - \frac{\beta}{2\beta+1}} = (2B)^\beta T^{\frac{\beta+1}{2\beta+1}}. \quad (65)$$

Furthermore, since $(\beta+1)/(2\beta+1) > 1/2$, we have $\sqrt{T} \leq T^{\frac{\beta+1}{2\beta+1}}$. By the definition of ι_T , $\iota_T \geq 2 \log 4 > 1$. Therefore, $\sqrt{\iota_T} \leq \iota_T$. In addition,

$$(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{NT\iota_T} = (D\iota_T + \sqrt{\lambda\iota_T})\sqrt{NT} \leq (D+1)(1+\sqrt{\lambda})\iota_T\sqrt{NT}. \quad (66)$$

Combining (66) with (64) gives

$$64C_\omega B(D\sqrt{\iota_T} + \sqrt{\lambda})\sqrt{NT\iota_T} \leq 64\sqrt{2}C_\omega B(D+1)(1+\sqrt{\lambda})\iota_T T^{\frac{\beta+1}{2\beta+1}}.$$

Next, by $h^\beta + \eta^2 \leq 2h^\beta$, (65), and $\sqrt{\iota_T} \leq \iota_T$,

$$16C_\omega B T(h^\beta + \eta^2)\sqrt{\iota_T} \leq 32C_\omega B T h^\beta \iota_T \leq 32C_\omega B (2B)^\beta (1+\sqrt{\lambda})\iota_T T^{\frac{\beta+1}{2\beta+1}}.$$

Finally, by $\sqrt{T} \leq T^{\frac{\beta+1}{2\beta+1}}$ and $(1+\sqrt{\lambda})\iota_T \geq 1$,

$$(8BD + L_{\text{Rev}})\sqrt{T} \leq (8BD + L_{\text{Rev}})(1+\sqrt{\lambda})\iota_T T^{\frac{\beta+1}{2\beta+1}}.$$

Consequently, we have

$$\begin{aligned} & \sum_{t \in [T] \setminus \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \\ & \leq \left[64\sqrt{2}C_\omega B(D+1) + 32C_\omega B(2B)^\beta + 8BD + L_{\text{Rev}} \right] (1+\sqrt{\lambda})\iota_T T^{\frac{\beta+1}{2\beta+1}}. \end{aligned} \quad (67)$$

Step 4: Pilot-exploration regret under the prescribed tuning. For every $x \geq 0$, $\log(1+x) \geq x/(1+x)$. Applying this inequality with $x = \eta^2/\gamma_T^2$ gives

$$\frac{1}{\log(1 + \eta^2/\gamma_T^2)} \leq 1 + \frac{\gamma_T^2}{\eta^2}.$$

Thus, by (62),

$$\sum_{t \in \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq BDd \left(1 + \frac{\gamma_T^2}{\eta^2} \right) \log \left(1 + \frac{TC_x^2}{d} \right). \quad (68)$$

Since $N \leq 2T^{\frac{1}{2\beta+1}}$, we have $h = 2B/N \geq BT^{-\frac{1}{2\beta+1}}$. Because $\eta^2 = \min\{h^2, h^\beta\}$, it follows that

$$\frac{1}{\eta^2} = \max\{h^{-2}, h^{-\beta}\} \leq \max\left\{B^{-2}T^{\frac{2}{2\beta+1}}, B^{-\beta}T^{\frac{\beta}{2\beta+1}}\right\} \leq \max\{B^{-2}, B^{-\beta}\}T^{\frac{\beta+1}{2\beta+1}}. \quad (69)$$

Since $T^{(\beta+1)/(2\beta+1)} \geq 1$,

$$1 + \frac{\gamma_T^2}{\eta^2} \leq \left[1 + \max\{B^{-2}, B^{-\beta}\}\gamma_T^2 \right] T^{\frac{\beta+1}{2\beta+1}} \leq \max\{1, B^{-2}, B^{-\beta}\}(1 + \gamma_T^2)T^{\frac{\beta+1}{2\beta+1}}.$$

Substituting this inequality into (68) gives

$$\sum_{t \in \mathcal{T}^{\text{exp}}} [\text{Rev}(\mathbf{x}_t, p_t^*) - \text{Rev}(\mathbf{x}_t, p_t)] \leq BD \max\{1, B^{-2}, B^{-\beta}\} d(1 + \gamma_T^2) \log \left(1 + \frac{TC_x^2}{d} \right) T^{\frac{\beta+1}{2\beta+1}}. \quad (70)$$

Step 5: Completion of the tuned bound. Define

$$C := \max \left\{ BD \max\{1, B^{-2}, B^{-\beta}\}, 64\sqrt{2}C_\omega B(D+1) + 32C_\omega B(2B)^\beta + 8BD + L_{\text{Rev}} \right\}.$$

By Lemma A.2 and Lemma 5.1, C is finite and depends only on the fixed problem primitives.

Combining (67) and (70) yields

$$\text{Reg}(T) \leq C \left[d(1 + \gamma_T^2) \log \left(1 + \frac{TC_x^2}{d} \right) + (1 + \sqrt{\lambda})\iota_T \right] T^{\frac{\beta+1}{2\beta+1}},$$

which proves (16).

Finally, the definitions of γ_T and ι_T imply $\gamma_T^2 = \mathcal{O}(\log T)$, $\iota_T = \mathcal{O}(\log T)$. Consequently,

$$d(1 + \gamma_T^2) \log \left(1 + \frac{TC_x^2}{d} \right) + (1 + \sqrt{\lambda})\iota_T = \mathcal{O}(\log^2 T),$$

and it follows that

$$\text{Reg}(T) = \tilde{\mathcal{O}} \left(T^{\frac{\beta+1}{2\beta+1}} \right)$$

with probability at least $1 - \delta$. This completes the proof. $\square$

Proof of Lemma 5.8.

Proof. We construct g_0 in four steps. We first build a nonincreasing envelope that generates a flat globally optimal revenue region and has more mass than required by the zero-mean condition. We then smooth its only kink without changing the flat region, truncate its right tail so that the zero-mean condition holds exactly, and finally verify the remaining properties and construct the compensation function χ .

Step 1: Constructing a flat-revenue envelope. We first construct a nonincreasing function whose associated revenue is constant at its global maximum over an interior price interval. The construction initially has more total mass than required by the zero-mean condition; this excess mass will be removed in Step 3.

Define

$$\bar{g}(u) := \min \left\{ D, \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u} \right\}, \quad u \in [-B_\epsilon, B_\epsilon],$$

where the ratio is interpreted as $+\infty$ when $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u = 0$. Then

$$\bar{g}(u) = D, \quad \forall u \leq -\frac{B_\epsilon}{2} \quad \text{and} \quad \bar{g}(u) = \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u}, \quad \forall u > -\frac{B_\epsilon}{2}.$$

Consequently,

$$\int_{-B_\epsilon}^{B_\epsilon} \bar{g}(u) du = \frac{DB_\epsilon}{2} + D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right) \log \left(1 + \frac{3B_\epsilon}{2((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)} \right). \quad (71)$$

For every $a > 0$, the function $x \log(1 + a/x)$ is strictly increasing in x on $(0, \infty)$, because

$$\frac{d}{dx} \left[x \log \left(1 + \frac{a}{x} \right) \right] = \log \left(1 + \frac{a}{x} \right) - \frac{a}{x+a} > 0.$$

Since $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 \geq B_\epsilon/2$, take $a = 3B_\epsilon/2$ and equation (71) implies

$$\int_{-B_\epsilon}^{B_\epsilon} \bar{g}(u) du \geq \frac{DB_\epsilon}{2} + \frac{DB_\epsilon}{2} \log 4 > DB_\epsilon. \quad (72)$$

Set

$$p_L := (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{3B_\epsilon}{8}, \quad p_U := (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{4}.$$

Because $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in [B_\epsilon, B - B_\epsilon]$, we have $0 < p_L < p_U < B$ and

$$[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ] = \left[-\frac{3B_\epsilon}{8}, -\frac{B_\epsilon}{4}\right] \subset (-B_\epsilon, B_\epsilon).$$

For every $p \in [p_L, p_U]$, we have $p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \geq -3B_\epsilon/8 > -B_\epsilon/2$. Hence,

$$p \bar{g}(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = p \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{p} = D\left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2}\right). \quad (73)$$

Thus, the revenue induced by $\bar{g}$ is constant on $[p_L, p_U]$.

Step 2: Smoothing the envelope without changing the flat region. The envelope $\bar{g}$ is continuous but has a kink at $-B_\epsilon/2$. We smooth this kink over a short interval while preserving monotonicity, remaining below $\bar{g}$, and leaving the flat-revenue region unchanged.

Fix a nondecreasing infinitely differentiable function $\rho : \mathbb{R} \rightarrow [0, 1]$ satisfying

$$\rho(z) = 0 \quad \text{for } z \leq 0, \quad \rho(z) = 1 \quad \text{for } z \geq 1, \quad \rho'(z) > 0 \quad \text{for } z \in (0, 1).$$

Choose $\zeta > 0$ sufficiently small that

$$\zeta < B_\epsilon/64 \quad \text{and} \quad 6D\zeta < \int_{-B_\epsilon}^{B_\epsilon} \bar{g}(u) du - DB_\epsilon. \quad (74)$$

There exists $z_0 \in (-B_\epsilon/2 - 2\zeta, -B_\epsilon/2)$ such that

$$\int_{-B_\epsilon/2-2\zeta}^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} \rho\left(\frac{v - z_0}{\zeta}\right) dv = D - \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 + 2\zeta}. \quad (75)$$

Since $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \geq B_\epsilon$, we have $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 - 2\zeta > 0$, and the denominator above is strictly positive. Indeed, the left-hand side is continuous in z_0 , whereas the right-hand side equals

$$\int_{-B_\epsilon/2}^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} dv.$$

When $z_0 = -B_\epsilon/2 - 2\zeta$, we have

$$\rho\left(\frac{v - z_0}{\zeta}\right) = 1 \quad \text{for every } v \geq -\frac{B_\epsilon}{2} - \zeta.$$

Hence the left-hand side of (75) is at least

$$\int_{-B_\epsilon/2-\zeta}^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} dv,$$

which is strictly larger than the right-hand side since the additional integral over $[-B_\epsilon/2 - \zeta, -B_\epsilon/2]$ is strictly positive.

In contrast, when $z_0 = -B_\epsilon/2$, the multiplier is zero for $v \leq -B_\epsilon/2$, lies strictly between zero and one for $v \in (-B_\epsilon/2, -B_\epsilon/2 + \zeta)$, and equals one for $v \geq -B_\epsilon/2 + \zeta$. Therefore, the left-hand side is strictly smaller than

$$\int_{-B_\epsilon/2}^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} dv,$$

which is precisely the right-hand side. Applying the intermediate value theorem therefore yields a $z_0 \in (-B_\epsilon/2 - 2\zeta, -B_\epsilon/2)$ for which (75) holds.

Define

$$\tilde{g}(u) := \begin{cases} D, & u \leq -B_\epsilon/2 - 2\zeta, \\ D - \int_{-B_\epsilon/2-2\zeta}^u \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} \rho\left(\frac{v - z_0}{\zeta}\right) dv, & -B_\epsilon/2 - 2\zeta < u < -B_\epsilon/2 + 2\zeta, \\ \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u}, & u \geq -B_\epsilon/2 + 2\zeta. \end{cases}$$

The choice of z_0 in (75) ensures that the middle piece and the fractional piece of $\tilde{g}$ have the same value at $u = -B_\epsilon/2 + 2\zeta$. At the left junction point, the multiplier involving ρ equals zero on a neighborhood, so the middle piece coincides locally with the constant D . At the right junction point, the multiplier equals one on a neighborhood, and (75) ensures that the middle piece coincides locally with the fractional piece. Therefore, all derivatives match at both junction points, and $\tilde{g}$ is infinitely differentiable on $[-B_\epsilon, B_\epsilon]$.

We verify the monotonicity and range of $\tilde{g}$. On the middle interval, we have

$$\tilde{g}'(u) = -\frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u)^2} \rho\left(\frac{u - z_0}{\zeta}\right) \leq 0.$$

Hence the middle piece is nonincreasing. It starts from D at $u = -B_\epsilon/2 - 2\zeta$ and, by (75), ends at

$$\frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 + 2\zeta} > 0. \quad (76)$$

Since the left piece is constant at D and the right piece is positive and decreasing, $\tilde{g}$ is nonincreasing on $[-B_\epsilon, B_\epsilon]$ and satisfies $0 \leq \tilde{g}(u) \leq D$ for $u \in [-B_\epsilon, B_\epsilon]$.

We next show that $\tilde{g} \leq \bar{g}$. If $u \leq -B_\epsilon/2$, then $\bar{g}(u) = D$ and $\tilde{g}(u) \leq D$, so the desired inequality holds. If $u \geq -B_\epsilon/2 + 2\zeta$, we have $\tilde{g}(u) = \bar{g}(u)$ by definition. It only remains to consider the case $u \in [-B_\epsilon/2, -B_\epsilon/2 + 2\zeta]$. For every such u , by (76), we have

$$\begin{aligned} \tilde{g}(u) &= \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 + 2\zeta} + \int_u^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} \rho\left(\frac{v - z_0}{\zeta}\right) dv \\ &\leq \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2 + 2\zeta} + \int_u^{-B_\epsilon/2+2\zeta} \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + v)^2} dv \\ &= \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u} = \bar{g}(u). \end{aligned}$$

Consequently, we have $0 \leq \tilde{g}(u) \leq \bar{g}(u)$ for $u \in [-B_\epsilon, B_\epsilon]$.

Finally, the smoothing modification does not affect the flat-revenue region. Indeed, by (74),

$$-\frac{B_\epsilon}{2} + 2\zeta < -\frac{3B_\epsilon}{8} = p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ.$$

Thus the smoothing interval lies strictly to the left of $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$. It follows from the definition of $\tilde{g}$ that

$$\tilde{g}(u) = \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u}$$

on a neighborhood of $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$. Hence the smoothing operation leaves the flat-revenue region unchanged.

Step 3: Enforcing the zero-mean condition. Although $\tilde{g}$ is smooth and preserves the flat-revenue region, its integral remains larger than DB_ϵ . Indeed, $\tilde{g}$ differs from $\bar{g}$ only on $[-B_\epsilon/2 - 2\zeta, -B_\epsilon/2 + 2\zeta]$, whose length is 4ζ . Since both functions take values in $[0, D]$, equations (72) and (74) give

$$\int_{-B_\epsilon}^{B_\epsilon} \tilde{g}(u) du \geq \int_{-B_\epsilon}^{B_\epsilon} \bar{g}(u) du - 4D\zeta > DB_\epsilon.$$

We therefore remove part of the right tail of $\tilde{g}$ smoothly, while leaving the flat-revenue region unchanged.

For every $\xi \in [p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + \zeta, B_\epsilon - 2\zeta]$, define

$$g_\xi(u) := \tilde{g}(u) \left[1 - \rho\left(\frac{u - \xi}{\zeta}\right) \right], \quad u \in [-B_\epsilon, B_\epsilon].$$

For $u \leq \xi$, the multiplicative factor equals one, whereas for $u \geq \xi + \zeta$, it equals zero. On the interval $(\xi, \xi + \zeta)$, it decreases smoothly from one to zero. Thus g_ξ agrees with $\tilde{g}$ to the left of ξ and vanishes to the right of $\xi + \zeta$.

Since $\tilde{g}$ is nonnegative and nonincreasing, and since ρ is nondecreasing,

$$g'_\xi(u) = \tilde{g}'(u) \left[1 - \rho\left(\frac{u - \xi}{\zeta}\right) \right] - \frac{\tilde{g}(u)}{\zeta} \rho'\left(\frac{u - \xi}{\zeta}\right) \leq 0.$$

Hence g_ξ is nonincreasing. Moreover, we have $0 \leq g_\xi(u) \leq \tilde{g}(u) \leq \bar{g}(u)$. Because $\xi \geq p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + \zeta$, the multiplicative factor equals one throughout $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$. Therefore, on this interval,

$$g_\xi(u) = \tilde{g}(u) = \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u}.$$

In addition, $g_\xi = D$ on a neighborhood of $-B_\epsilon$ and $g_\xi = 0$ on a neighborhood of B_ϵ .

For every fixed u , the function $g_\xi(u)$ is continuous and nondecreasing in ξ . Since $0 \leq g_\xi(u) \leq D$, the dominated convergence theorem implies that

$$\xi \mapsto \int_{-B_\epsilon}^{B_\epsilon} g_\xi(u) du$$

is continuous and nondecreasing.

At $\xi = p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + \zeta$, the function g_ξ vanishes for $u \geq p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + 2\zeta$. Consequently,

$$\int_{-B_\epsilon}^{B_\epsilon} g_\xi(u) du \leq D \left(B_\epsilon + p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + 2\zeta \right) = D \left(\frac{3B_\epsilon}{4} + 2\zeta \right) < DB_\epsilon.$$

At $\xi = B_\epsilon - 2\zeta$, the functions g_ξ and $\bar{g}$ may differ only on

$$\left[-\frac{B_\epsilon}{2} - 2\zeta, -\frac{B_\epsilon}{2} + 2\zeta\right] \cup [B_\epsilon - 2\zeta, B_\epsilon],$$

whose total length is at most 6ζ . Since both functions take values in $[0, D]$, equations (72) and (74) imply

$$\int_{-B_\epsilon}^{B_\epsilon} g_{B_\epsilon - 2\zeta}(u) du \geq \int_{-B_\epsilon}^{B_\epsilon} \bar{g}(u) du - 6D\zeta > DB_\epsilon.$$

Therefore, by continuity, there exists $\xi_\star \in [p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + \zeta, B_\epsilon - 2\zeta]$ such that

$$\int_{-B_\epsilon}^{B_\epsilon} g_{\xi_\star}(u) du = DB_\epsilon.$$

Step 4: Verifying the baseline properties and constructing the compensation function.

We now set $g_0 = g_{\xi_\star}$. We first verify that g_0 corresponds to an admissible zero-mean noise distribution and generates the desired flat globally optimal revenue region. We then construct χ , which will be used to preserve the zero-mean condition in the subsequent perturbation construction.

Set $g_0 = g_{\xi_\star}$ on $[-B_\epsilon, B_\epsilon]$, and extend it to $\mathbb{R}$ by $g_0(u) = D$ for $u \leq -B_\epsilon$, and $g_0(u) = 0$ for $u \geq B_\epsilon$. Because $g_{\xi_\star}$ is constant on neighborhoods of both endpoints, this extension is infinitely differentiable on $\mathbb{R}$. In particular, $g_0 \in \mathcal{H}(\beta)$ on $[-B, B]$, with a finite Hölder constant.

The function g_0/D is continuous, nonincreasing, equals one to the left of $-B_\epsilon$, and equals zero to the right of B_ϵ . It is therefore the survival function of a continuous distribution supported on $[-B_\epsilon, B_\epsilon]$. Moreover, by the choice of $\xi_\star$ in Step 3,

$$\mathbb{E}[\epsilon_t] = \int_{-B_\epsilon}^{B_\epsilon} \frac{g_0(u)}{D} du - B_\epsilon = 0.$$

We next verify the revenue properties. If $p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in [-B_\epsilon, B_\epsilon]$ and $p > 0$, then $g_0 \leq \bar{g}$, and hence

$$p g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) \leq p \min \left\{ D, \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{p} \right\} \leq D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right).$$

If $p = 0$, the same inequality is immediate. If $p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \leq -B_\epsilon$, then $g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = D$ and

$$p g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = Dp \leq D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon) < D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right).$$

If $p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \geq B_\epsilon$, then $g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = 0$. This proves the inequality in (20).

For every $p \in [p_L, p_U]$, Step 3 gives

$$g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{p}.$$

Therefore,

$$p g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) = D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right), \quad p \in [p_L, p_U],$$

which proves the equality in (20).

Furthermore, for every $p \in [p_L, p_U]$,

$$0 < \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{p_U} \leq g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) \leq \frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{p_L} < D.$$

Because the same representation holds on a neighborhood of $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$, the function g_0 is bounded away from both 0 and D , and

$$g'_0(u) = -\frac{D((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - B_\epsilon/2)}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u)^2} < 0$$

throughout that neighborhood.

It remains to construct the compensation function. Define

$$\chi(u) := \frac{2}{\zeta} \rho' \left(\frac{2(u - \xi_\star)}{\zeta} - \frac{1}{2} \right).$$

Since $\rho' \geq 0$, we have $\chi(u) \geq 0$. By the definition of χ ,

$$\int_{\mathbb{R}} \chi(u) du = \int_{\mathbb{R}} \frac{2}{\zeta} \rho' \left(\frac{2(u - \xi_\star)}{\zeta} - \frac{1}{2} \right) du = \left[\rho \left(\frac{2(u - \xi_\star)}{\zeta} - \frac{1}{2} \right) \right]_{u=-\infty}^{u=\infty} = 1 - 0 = 1.$$

Moreover, since ρ' vanishes outside $(0, 1)$, by the choice of $\xi_\star$ in Step 3

$$\text{supp}(\chi) \subset \left[\xi_\star + \frac{\zeta}{4}, \xi_\star + \frac{3\zeta}{4} \right] \subset (-B_\epsilon, B_\epsilon) \setminus [p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ].$$

On a sufficiently small neighborhood of $\text{supp}(\chi)$, the factor $1 - \rho \left(\frac{u - \xi_\star}{\zeta} \right)$ is bounded away from both zero and one. Since $\tilde{g}$ is also bounded away from both zero and D there, the same holds for g_0 .

Finally,

$$g'_0(u) = \tilde{g}'(u) \left[1 - \rho \left(\frac{u - \xi_\star}{\zeta} \right) \right] - \frac{\tilde{g}(u)}{\zeta} \rho' \left(\frac{u - \xi_\star}{\zeta} \right) < 0,$$

because $\tilde{g}'(u) < 0$, $\tilde{g}(u) > 0$, and $\rho' \geq 0$ there. Since g'_0 is continuous, it is uniformly negative on a sufficiently small closed neighborhood of $\text{supp}(\chi)$. $\square$

Proof of Theorem 5.6.

Proof.

Step 1: Constructing the perturbed instances. Let g_0, p_L, p_U , and χ be as in Lemma 5.8. Set $K_T := \lceil T^{1/(2\beta+1)} \rceil$ and, for every $k \in [K_T]$, define

$$\bar{p}_k := p_L + \left(k - \frac{1}{2} \right) \frac{p_U - p_L}{K_T}.$$

Thus, $\bar{p}_1, \dots, \bar{p}_{K_T}$ are equally spaced in $[p_L, p_U]$. For brevity, write $u_k := \bar{p}_k - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ$.

Choose a sufficiently small constant $\kappa > 0$, depending only on the fixed problem primitives. For every $k \in [K_T]$, define

$$\Delta_k(u) := \begin{cases} \frac{\kappa K_T^{-\beta}}{((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u) \rho'(1/2)} \rho' \left(\frac{4K_T(u - u_k)}{p_U - p_L} + \frac{1}{2} \right), & |u - u_k| < \frac{p_U - p_L}{8K_T}, \\ 0, & \text{otherwise,} \end{cases} \quad (77)$$

and

$$g_k(u) := g_0(u) + \Delta_k(u) - \chi(u) \int_{\mathbb{R}} \Delta_k(v) dv. \quad (78)$$

The term Δ_k creates a local revenue increase near $\bar{p}_k$, while the term involving χ removes the same total mass away from this region.

Because $\rho'(1/2) > 0$, Δ_k is well defined and nonnegative. Moreover,

$$\text{supp}(\Delta_k) \subset \left[\bar{p}_k - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{p_U - p_L}{8K_T}, \bar{p}_k - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + \frac{p_U - p_L}{8K_T} \right].$$

These supports are pairwise disjoint and contained in $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$. Since $(\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ + u \geq p_L > 0$ on their union, the height of Δ_k is of order $K_T^{-\beta}$, while the length of its support is of order K_T^{-1} . Hence

$$0 \leq \int_{\mathbb{R}} \Delta_k(u) du \leq C_{\text{lb}} \kappa K_T^{-(\beta+1)} \quad (79)$$

uniformly in k and T , where $C_{\text{lb}} < \infty$ denotes a constant that may change from line to line.

Because ρ is infinitely differentiable and constant outside $[0, 1]$, all derivatives of ρ' vanish at zero and one. Therefore, the piecewise-defined function Δ_k is infinitely differentiable on $\mathbb{R}$.

Since the supports of Δ_k and χ lie in $(-B_\epsilon, B_\epsilon)$, and $\int_{\mathbb{R}} \chi(u) du = 1$,

$$\int_{-B_\epsilon}^{B_\epsilon} g_k(u) du = \int_{-B_\epsilon}^{B_\epsilon} g_0(u) du + \int_{\mathbb{R}} \Delta_k(u) du - \left(\int_{\mathbb{R}} \chi(u) du \right) \left(\int_{\mathbb{R}} \Delta_k(u) du \right) = DB_\epsilon.$$

Thus, g_k and g_0 have the same integral on $[-B_\epsilon, B_\epsilon]$.

We first verify that the constructed functions satisfy the required smoothness condition. By enlarging C_{lb} , if necessary, repeated application of the product and chain rules gives

$$\sup_{u \in \mathbb{R}} \left| \Delta_k^{(r)}(u) \right| \leq C_{\text{lb}} \kappa K_T^{r-\beta}, \quad r = 0, 1, \dots, \varpi(\beta) + 1. \quad (80)$$

We now verify the Hölder condition. Fix $u, u' \in \mathbb{R}$. If $|u - u'| \leq K_T^{-1}$, the mean-value theorem gives

$$\begin{aligned} \left| \Delta_k^{(\varpi(\beta))}(u) - \Delta_k^{(\varpi(\beta))}(u') \right| &\leq C_{\text{lb}} \kappa K_T^{\varpi(\beta)+1-\beta} |u - u'| \\ &= C_{\text{lb}} \kappa |u - u'|^{\beta-\varpi(\beta)} (K_T |u - u'|)^{\varpi(\beta)+1-\beta} \\ &\leq C_{\text{lb}} \kappa |u - u'|^{\beta-\varpi(\beta)}. \end{aligned}$$

If $|u - u'| > K_T^{-1}$, then

$$\left| \Delta_k^{(\varpi(\beta))}(u) - \Delta_k^{(\varpi(\beta))}(u') \right| \leq 2 \sup_{v \in \mathbb{R}} \left| \Delta_k^{(\varpi(\beta))}(v) \right| \leq C_{\text{lb}} \kappa K_T^{\varpi(\beta)-\beta} \leq C_{\text{lb}} \kappa |u - u'|^{\beta-\varpi(\beta)}.$$

Therefore, we have

$$\left| \Delta_k^{(\varpi(\beta))}(u) - \Delta_k^{(\varpi(\beta))}(u') \right| \leq C_{\text{lb}} \kappa |u - u'|^{\beta-\varpi(\beta)}. \quad (81)$$

By Lemma 5.8, g_0 satisfies the required Hölder condition. Moreover, χ is fixed and infinitely differentiable, while (79) uniformly bounds the coefficient multiplying χ . Combining these facts with (81), there exists a finite constant $\bar{L}_g$, independent of k and T , such that $g_k \in \mathcal{H}(\beta, \bar{L}_g; [-B, B])$ for every $k \in [K_T]$.

We next verify the range and monotonicity requirements. The supports of Δ_k and χ are disjoint. By Lemma 5.8, g_0 is bounded away from both 0 and D , and g'_0 is uniformly negative, on neighborhoods of these supports.

By enlarging C_{lb} , if necessary, (80) and (79) imply

$$\|\Delta_k\|_\infty + \|\Delta'_k\|_\infty + \left\| \chi(\cdot) \int_{\mathbb{R}} \Delta_k(v) dv \right\|_\infty + \left\| \chi'(\cdot) \int_{\mathbb{R}} \Delta_k(v) dv \right\|_\infty \leq C_{\text{lb}} \kappa.$$

On $\text{supp}(\Delta_k)$, $g_k = g_0 + \Delta_k$, whereas on $\text{supp}(\chi)$,

$$g_k = g_0 - \chi(\cdot) \int_{\mathbb{R}} \Delta_k(v) dv.$$

Outside these two supports, $g_k = g_0$. Combining these observations with the preceding bounds, for all sufficiently small $\kappa > 0$, uniformly in k and T , $0 \leq g_k(u) \leq D$ and $g'_k(u) \leq 0$ for $u \in \mathbb{R}$.

Since the supports of Δ_k and χ lie strictly inside $(-B_\epsilon, B_\epsilon)$, $g_k(u) = D$ for $u \leq -B_\epsilon$ and $g_k(u) = 0$ for $u \geq B_\epsilon$. Therefore, g_k/D is the survival function of a distribution supported on $[-B_\epsilon, B_\epsilon]$. Moreover, the preceding integral identity gives

$$\mathbb{E}[\epsilon_t] = \int_{-B_\epsilon}^{B_\epsilon} \frac{g_k(u)}{D} du - B_\epsilon = 0.$$

Under instance k , the shocks are i.i.d. with the constructed distribution and are independent of the policy randomization. Moreover, because $y_t = D\mathbf{1}\{v_t \geq p_t\}$, the demand is measurable with respect to $\sigma(p_t, v_t)$. Therefore,

$$\mathbb{E}[y_t | \mathcal{F}_{t-1} \vee \sigma(\mathbf{x}_t, p_t, v_t)] = D\mathbf{1}\{v_t \geq p_t\} = q(v_t - p_t).$$

Hence, Assumption 3.2 holds. Furthermore, by (17), for every $u \in \mathcal{I}_g$,

$$\mathbb{E}[q(\epsilon_t - u)] = D\mathbb{P}(\epsilon_t \geq u) = g_k(u).$$

Thus, each constructed instance satisfies the original model assumptions provided that $L_g \geq \bar{L}_g$.

Step 2: Establishing the revenue gap and statistical closeness. Under instance k , the perturbation raises the revenue near $\bar{p}_k$, while prices outside its support receive no such increase.

Since the support of χ is disjoint from $[p_L - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ, p_U - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ]$, the definition of Δ_k gives

$$\bar{p}_k g_k \left(\bar{p}_k - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \right) = D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right) + \kappa K_T^{-\beta}.$$

If $p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \notin \text{supp}(\Delta_k)$, then $\Delta_k = 0$. Since the term involving χ is nonpositive,

$$p g_k \left(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \right) \leq p g_0 \left(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \right) \leq D \left((\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ - \frac{B_\epsilon}{2} \right).$$

Consequently,

$$\max_{p' \in [0, B]} p' g_k \left(p' - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \right) - p g_k \left(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \right) \geq \kappa K_T^{-\beta}, \quad p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \notin \text{supp}(\Delta_k). \quad (82)$$

Fix an arbitrary nonanticipating policy π . Let $\mathbb{P}_k$ denote the law of the complete observed history, including posted prices and demands, under instance k , and let $\mathbb{P}_0$ denote the corresponding law under the baseline instance g_0 . Conditional on the past and p_t , the normalized observation y_t/D is Bernoulli with success probability $g_k(p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ)/D$. Since the policy uses the same conditional decision rule under every instance, the chain rule for relative entropy gives

$$D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_k) = \mathbb{E}_{\mathbb{P}_0} \left[\sum_{t=1}^T D_{\text{KL}} \left(\text{Bern} \left(\frac{g_0(p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ)}{D} \right) \parallel \text{Bern} \left(\frac{g_k(p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ)}{D} \right) \right) \right]. \quad (83)$$

By the properties established in Step 1, the relevant Bernoulli probabilities on the supports of Δ_k and χ lie in a fixed subinterval of $(0, 1)$. By enlarging C_{lb} , if necessary,

$$D_{\text{KL}}(\text{Bern}(a) \parallel \text{Bern}(b)) \leq C_{\text{lb}}(a - b)^2$$

for all probabilities appearing above. Moreover,

$$\begin{aligned} \left| g_k(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) - g_0(p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ) \right| &\leq C_{\text{lb}} \kappa K_T^{-\beta} \mathbf{1} \{p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_k)\} \\ &\quad + C_{\text{lb}} \kappa K_T^{-(\beta+1)} \mathbf{1} \{p - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\chi)\}. \end{aligned}$$

Using (80) with $r = 0$ and (79), and enlarging C_{lb} if necessary, we obtain

$$D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_k) \leq C_{\text{lb}} \kappa^2 K_T^{-2\beta} \mathbb{E}_0 \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_k)\} \right] + C_{\text{lb}} \kappa^2 K_T^{-2\beta-2} T. \quad (84)$$

Because the supports of $\Delta_1, \dots, \Delta_{K_T}$ are pairwise disjoint,

$$\sum_{k=1}^{K_T} \sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_k)\} \leq T$$

pathwise. Hence there exists $k^* \in [K_T]$ such that

$$\mathbb{E}_0 \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_{k^*})\} \right] \leq \frac{T}{K_T}.$$

For this k^* , we have

$$D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_{k^*}) \leq C_{\text{lb}} \kappa^2 T K_T^{-(2\beta+1)} + C_{\text{lb}} \kappa^2 T K_T^{-(2\beta+2)}.$$

Since $K_T \geq 1$, by enlarging C_{lb} , if necessary, we have

$$D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_{k^*}) \leq C_{\text{lb}} \kappa^2 T K_T^{-(2\beta+1)}. \quad (85)$$

Step 3: Deriving the regret lower bound. The KL bound in Step 2 shows that instance k^* is difficult to distinguish from the baseline. We now use this fact to show that the policy misses the corresponding perturbation region sufficiently often. Under instance k^* , (82) gives

$$\text{Reg}(T) \geq \kappa K_T^{-\beta} \sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \notin \text{supp}(\Delta_{k^*})\}. \quad (86)$$

For every t , apply the Bretagnolle–Huber inequality to the event $\{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_{k^\star})\}$. Summing over t gives

$$\begin{aligned} & \mathbb{E}_0 \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_{k^\star})\} \right] + \mathbb{E}_{k^\star} \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \notin \text{supp}(\Delta_{k^\star})\} \right] \\ & \geq \frac{T}{2} \exp \{-D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_{k^\star})\}. \end{aligned} \quad (87)$$

By the definition of K_T , we have $TK_T^{-(2\beta+1)} \leq 1$. By further decreasing κ , if necessary, assume that $C_{\text{lb}}\kappa^2 \leq \log 2$. This choice preserves all the range, monotonicity, and smoothness properties established above. Equation (85) then implies $D_{\text{KL}}(\mathbb{P}_0 \parallel \mathbb{P}_{k^\star}) \leq \log 2$. Increasing T_0 , if necessary, also guarantees $K_T \geq 8$. Consequently,

$$\mathbb{E}_0 \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \in \text{supp}(\Delta_{k^\star})\} \right] \leq \frac{T}{K_T} \leq \frac{T}{8}.$$

Combining this bound with (87) gives

$$\mathbb{E}_{k^\star} \left[\sum_{t=1}^T \mathbf{1} \{p_t - (\mathbf{x}^\circ)^\top \boldsymbol{\theta}^\circ \notin \text{supp}(\Delta_{k^\star})\} \right] \geq \frac{T}{4} - \frac{T}{8} = \frac{T}{8}.$$

Taking $\mathbb{E}_{k^\star}$ on both sides of (86) therefore yields

$$\mathbb{E}_{k^\star}[\text{Reg}(T)] \geq \frac{\kappa}{8} TK_T^{-\beta}.$$

Finally, since $K_T \leq 2T^{1/(2\beta+1)}$, we obtain

$$\mathbb{E}_{k^\star}[\text{Reg}(T)] \geq \frac{\kappa}{2^{\beta+3}} T^{\frac{\beta+1}{2\beta+1}}.$$

The policy π was arbitrary. Taking the infimum over policies and the supremum over admissible instances completes the proof. $\square$