Supplementary Materials for "Contextual Dynamic Pricing with Unknown Noise: Explore-then-UCB Strategy and Improved Regrets"

Yiyun Luo[*], Will Wei Sun[†], and Yufeng Liu[‡]

A Supplementary Proofs

A.1 Proof of Lemma 1

Lemma 1. *Under Assumptions 1 – 3, there exists positive constants $\tilde{c}_1, \tilde{c}_2, \tilde{c}_3$ such that for any episode k with the exploration phase length $n_k \geq \tilde{c}_3(d_0+1)^3$, we have with probability at least $1 - \frac{2}{n_k} - \tilde{c}_1 e^{-\frac{\tilde{c}_2}{(d_0+1)^2} n_k}$ that*

$$||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}.$$

Proof. Denote $\mu = \mathbb{E}(z_t)$ as the expectation of the noise $z_t \sim F$. By Assumptions 3, the random valuation $v_t = x_t^\top \theta_0 + z_t$ must be bounded in $[0, B]$. By Assumption 1, we

*Ph.D. Student, Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill. Email: yiyun851@live.unc.edu.

†Assistant Professor, Krannert School of Management, Purdue University. Email: sun244@purdue.edu.

‡Professor, Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Sciences, UNC Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill. Email: yfliu@email.unc.edu.

have $|x_t^\top \theta_0| \le ||x_t||_\infty ||\theta_0||_1 \le W$. Therefore, the noise is bounded as $|z_t| \le U = W + B$. Thus we obtain $|\mu| = |\mathbb{E}(z_t)| \le U$. Remember that the estimated $\hat{\theta}_k$ in the k-th episode is derived from the optimization problem

$$(\hat{\mu}_k, \hat{\theta}_k) = \arg\min_{\mu,\theta} \frac{1}{|\mathcal{E}_k|} \sum_{t \in \mathcal{E}_k} \left(By_t - (1, x_t^\top)(\mu, \theta^\top)^\top\right)^2.$$

Denote $\tilde{x}_t = (1, x_t^\top)^\top$ and $\hat{\xi}_k = (\hat{\mu}_k, \hat{\theta}_k^\top)$. Denote $\xi^* = (\mu, \theta_0^\top)^\top$ as the true parameter. Then we have $||\xi^*||_1 = |\mu| + ||\theta_0||_1 \le U + W$. Let $L_k(\xi) = \frac{1}{n_k} \sum_{t \in \mathcal{E}_k} (By_t - \tilde{x}_t^\top \xi)^2$. Then we have $\hat{\xi}_k = \arg\min_\xi L_k(\xi)$. By applying the Taylor's theorem with Lagrange Remainder on the true value ξ^*, we obtain

$$L_k(\hat{\xi}_k) - L_k(\xi^*) = (\hat{\xi}_k - \xi^*)^\top \nabla L_k(\xi^*) + \frac{1}{2}(\hat{\xi}_k - \xi^*)^\top \nabla^2 L_k(\xi^*)(\hat{\xi}_k - \xi^*).$$

As $\hat{\xi}_k$ is the minimizer of $L_k(\xi)$, we have

$$(\hat{\xi}_k - \xi^*)^\top \nabla L_k(\xi^*) + \frac{1}{2}(\hat{\xi}_k - \xi^*)^\top \nabla^2 L_k(\xi^*)(\hat{\xi}_k - \xi^*) = L_k(\hat{\xi}_k) - L_k(\xi^*) \le 0.$$

Rearranging the terms and applying the Hölder's inequality yields

$$(\hat{\xi}_k - \xi^*)^\top \nabla^2 L_k(\xi^*)(\hat{\xi}_k - \xi^*) \le 2(\xi^* - \hat{\xi}_k)^\top \nabla L_k(\xi^*) \le 2||\xi^* - \hat{\xi}_k||_1 ||\nabla L_k(\xi^*)||_\infty. \quad (1)$$

In the following, we upper bound $||\nabla L_k(\xi^*)||_\infty$ and lower bound $\nabla^2 L_k(\xi^*)$.

We first upper bound $||\nabla L_k(\xi^*)||_\infty$ with high probability. Simple calculation yields

$$\nabla L_k(\xi^*) = \frac{1}{n_k} \sum_{t \in \mathcal{E}_k} 2(\tilde{x}_t^\top \xi^* - By_t)\tilde{x}_t.$$

A critical observation is that for any t, we have

$$\mathbb{E}(By_t|\tilde{x}_t) = \mathbb{E}\big(\mathbb{E}(B1_{\{p_t \le x_t^\top \theta_0 + z_t\}}|\tilde{x}_t, z_t)\big|\tilde{x}_t\big) = B\mathbb{E}(\frac{x_t^\top \theta_0 + z_t}{B}|\tilde{x}_t)$$
$$= x_t^\top \theta_0 + \mathbb{E}(z_t|\tilde{x}_t) = x_t^\top \theta_0 + \mu = \tilde{x}_t^\top \xi^*.$$

Let $Y_{i,t}^{(k)} = \frac{2}{n_k}(\tilde{x}_t \xi^* - By_t)\tilde{x}_{t,i}$ for $i \in [d_0+1], t \in \mathcal{E}_k$. Then we have $\big(\nabla L_k(\xi^*)\big)_i = \sum_{t \in \mathcal{E}_k} Y_{i,t}^{(k)}$. Taking expectation of $Y_{i,t}^{(k)}$ yields

$$\mathbb{E}(Y_{i,t}^{(k)}) = \mathbb{E}\big(\mathbb{E}(\frac{2}{n_k}(\tilde{x}_t \xi^* - By_t)\tilde{x}_{t,i}|\tilde{x}_t)\big) = \mathbb{E}\big(\frac{2}{n_k}\tilde{x}_{t,i}(\tilde{x}_t^\top \xi^* - \mathbb{E}(By_t|\tilde{x}_t))\big) = 0.$$

Thus for any $i \in [d_0+1]$, we have

$$\mathbb{E}\big(\nabla L_k(\xi^*)\big)_i = \mathbb{E}\big(\sum_{t \in \mathcal{E}_k} Y_{i,t}^{(k)}\big) = \sum_{t \in \mathcal{E}_k} \mathbb{E}(Y_{i,t}^{(k)}) = 0.$$

By Assumption 1, $||x_t||_\infty \le 1$. Thus we have for any $i \in [d_0+1]$ and $t \in \mathcal{E}_k$,

$$|Y_{i,t}^{(k)}| = |\frac{2}{n_k}(\tilde{x}_t \xi^* - By_t)\tilde{x}_{t,i}| \le \frac{2}{n_k}(|\tilde{x}_t^\top \xi^*| + |By_t|)$$
$$\le \frac{2}{n_k}(B + ||\tilde{x}_t||_\infty ||\xi^*||_1) \le \frac{2}{n_k}(B + U + W).$$

Namely, $Y_{i,t}^{(k)}$ is bounded as $-\frac{2}{n_k}(B+U+W) \le Y_{i,t}^{(k)} \le \frac{2}{n_k}(B+U+W)$. Furthermore, for any $i \in [d_0+1]$, $Y_{i,t}^{(k)}$ is independent across $t \in \mathcal{E}_k$. Therefore, by applying the Hoeffding's inequality for bounded variables, we obtain for any $\epsilon > 0$,

$$\mathbb{P}(|\big(\nabla L_k(\xi^*)\big)_i| \ge \epsilon) = \mathbb{P}(|\big(\nabla L_k(\xi^*)\big)_i - \mathbb{E}\big(\nabla L_k(\xi^*)\big)_i| \ge \epsilon)$$
$$\le 2\exp\Big(-\frac{2\epsilon^2}{\sum_{t=t_k+1}^{t_k+n_k}\big(\frac{2}{n_k}(B+U+W) - (-\frac{2}{n_k}(B+U+W))\big)^2}\Big)$$
$$\le 2\exp\Big(-\frac{n_k \epsilon^2}{8(B+U+W)^2}\Big).$$

Thus by applying union bound over $i \in [d_0+1]$, we have for any $\epsilon > 0$,

$$\mathbb{P}(||\nabla L_k(\xi^*)||_\infty \geq \epsilon) \leq \sum_{i\in[d_0+1]} \mathbb{P}(|\big(\nabla L_k(\xi^*)\big)_i| \geq \epsilon) = 2(d_0+1)\exp\Big(-\frac{n_k\epsilon^2}{8(B+U+W)^2}\Big).$$

By taking $\epsilon = 4(B+U+W)\sqrt{\frac{\log n_k}{n_k}}$, we obtain for $n_k \geq d_0+1$,

$$\mathbb{P}(||\nabla L_k(\xi^*)||_\infty \geq 4(B+U+W)\sqrt{\frac{\log n_k}{n_k}}) \leq \frac{2(d_0+1)}{n_k^2} \leq \frac{2}{n_k}. \tag{2}$$

Thus we have $||\nabla L_k(\xi^*)||_\infty \leq 4(B+U+W)\sqrt{\frac{\log n_k}{n_k}}$ with probability at least $1-\frac{2}{n_k}$.

We then lower bound $\nabla^2 L_k(\xi^*)$. Simple calculation yields

$$\nabla^2 L_k(\xi^*) = \frac{2}{n_k}\sum_{t\in\mathcal{E}_k} \tilde{x}_t\tilde{x}_t^\top = 2\hat{\Sigma}_k,$$

where $\hat{\Sigma}_k = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k} \tilde{x}_t\tilde{x}_t^\top$ is indeed an empirical estimate of the population matrix $\Sigma = \mathbb{E}(\tilde{x}_t\tilde{x}_t^\top)$. Let $\bar{x} = \mathbb{E}(\tilde{x}_t)$ be the expectation of $\tilde{x}_t$. Let $\dot{x}_t = \tilde{x}_t - \bar{x}$ be the difference between $\tilde{x}_t$ and its expectation. Denote $\Sigma^* = \mathbb{E}(\dot{x}_t\dot{x}_t^\top) = \mathbb{E}((\tilde{x}_t-\bar{x})(\tilde{x}_t-\bar{x})^\top)$ as the covariance matrix of $\tilde{x}_t$. Further denote $\dot{\Sigma}_k = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\dot{x}_t\dot{x}_t^\top$ as the empirical covariance matrix and $\bar{\dot{x}}_k = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\dot{x}_t, \Gamma_k = \bar{\dot{x}}_k\bar{x}^\top + \bar{x}\bar{\dot{x}}_k^\top$.

For Σ, we have the decomposation

$$\Sigma = \mathbb{E}(\tilde{x}_t\tilde{x}_t^\top) = \mathbb{E}\big((\tilde{x}_t-\bar{x})(\tilde{x}_t-\bar{x})^\top\big) + \bar{x}\bar{x}^\top = \Sigma^* + \bar{x}\bar{x}^\top.$$

For the empirical version $\hat{\Sigma}_k = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\tilde{x}_t\tilde{x}_t^\top$, we also have the decomposition

$$\begin{aligned}
\hat{\Sigma}_k &= \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\tilde{x}_t\tilde{x}_t^\top = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}(\dot{x}_t+\bar{x})(\dot{x}_t+\bar{x}) \\
&= \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\dot{x}_t\dot{x}_t^\top + \bar{x}\bar{x}^\top + \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}(\dot{x}_t\bar{x}^\top + \bar{x}\dot{x}_t^\top) \\
&= \dot{\Sigma}_k + \bar{x}\bar{x}^\top + (\bar{\dot{x}}_k\bar{x}^\top + \bar{x}\bar{\dot{x}}_k^\top) = \dot{\Sigma}_k + \bar{x}\bar{x}^\top + \Gamma_k.
\end{aligned}$$

By using Σ, we can rewrite $\hat{\Sigma}_k$ as

$$\hat{\Sigma}_k = \dot{\Sigma}_k + \bar{x}\bar{x}^\top + \Gamma_k = \Sigma - (\Sigma^* - \dot{\Sigma}_k) + \Gamma_k. \tag{3}$$

Now we tackle the difference $\dot{\Sigma}_k - \Sigma^* = \frac{1}{n_k}\sum_{t\in\mathcal{E}_k}\dot{x}_t\dot{x}_t^\top - \mathbb{E}(\dot{x}_t\dot{x}_t^\top)$. Since $||x_t||_\infty \leq 1$, we have $||\bar{x}||_\infty \leq 1$ and $||\dot{x}_t||_\infty \leq 2$. Let $\mathbb{S}^{d_0} = \{v \in \mathbb{R}^{d_0+1} \mid ||v||_2 = 1\}$ be the Euclidean unit sphere in $\mathbb{R}^{d_0+1}$. Since $\mathbb{E}(\dot{x}_t) = 0$, for any $v \in \mathbb{S}^{d_0}$, we have $\mathbb{E}(v^\top\dot{x}_t) = 0$ and $|v^\top\dot{x}_t| \leq ||v||_2||\dot{x}_t||_2 \leq 2\sqrt{d_0+1}$. Thus by Hoeffding's lemma, $v^\top\dot{x}_t$ is sub-gaussian with variance proxy $\sigma^2 = 4(d_0+1)$. Since this holds for any $v \in \mathbb{S}^{d_0}$, $\dot{x}_t$ is a sub-gaussian random vector with variance proxy $\sigma^2 = 4(d_0+1)$. Denote $|||A|||_2 = \max_{||v||_2=1}||Av||_2$ as the ℓ_2 operator norm for a matrix A. Then by Theorem 6.5 in Wainwright (2019), there are universal positive constants $\{c_i\}_{i=1}^3$ such that

$$\mathbb{P}\Big(\frac{|||\dot{\Sigma}_k - \Sigma^*|||_2}{\sigma^2} \geq c_1\big(\sqrt{\frac{d_0+1}{n_k}} + \frac{d_0+1}{n_k}\big) + \delta\Big) \leq c_2 e^{-c_3 n_k \min\{\delta,\delta^2\}}, \forall\delta > 0.$$

There exists a positive constant $\tilde{c}_0 \leq c_0$ that satisfies $\tilde{c}_0 \leq \min\{32, 64c_1\}(d_0+1)$. Let $\delta = \frac{\tilde{c}_0}{32(d_0+1)} \leq 1$. Then $\min\{\delta,\delta^2\} = \delta^2 = \frac{\tilde{c}_0^2}{1024(d_0+1)^2}$. For $n_k \geq \frac{64^2c_1^2}{\tilde{c}_0^2}(d_0+1)^3$, we have $\sqrt{\frac{d_0+1}{n_k}} \leq \frac{\tilde{c}_0}{64c_1(d_0+1)} \leq 1$. Thus $\sqrt{\frac{d_0+1}{n_k}} + \frac{d_0+1}{n_k} \leq 2\sqrt{\frac{d_0+1}{n_k}} \leq \frac{\tilde{c}_0}{32c_1(d_0+1)}$ and $\sigma^2\Big(c_1\big(\sqrt{\frac{d_0+1}{n_k}} + \frac{d_0+1}{n_k}\big) + \delta\Big) \leq \frac{\tilde{c}_0}{4}$. Therefore, we have

$$\mathbb{P}\big(|||\dot{\Sigma}_k - \Sigma^*|||_2 \geq \frac{\tilde{c}_0}{4}\big) \leq c_2 e^{-\frac{c_3\tilde{c}_0^2}{1024(d_0+1)^2}n_k}, \tag{4}$$

for $n_k \geq \frac{64^2c_1^2}{\tilde{c}_0^2}(d_0+1)^3$.

Now we proceed to handle another component in the expression (3) of $\hat{\Sigma}_k$, which is $\Gamma_k = \bar{\dot{x}}_k\bar{x}^\top + \bar{x}\bar{\dot{x}}_k^\top$. Clearly Γ_k is symmetric. Thus the ℓ_2 operator norm for Γ_k can be written as $|||\Gamma_k|||_2 = \max_{v\in\mathbb{S}^{d_0}}|v^\top\Gamma_k v|$. We aim for a high probability upper bound for the ℓ_2-operator norm $|||\Gamma_k|||_2$.

To reduce the supremum to a finite maximum, we apply the discretization argument introduced in the proof of Theorem 6.5 in Wainwright (2019). For completeness, we present the detailed derivations of this discretization method. Let $\{v^1, \dots, v^N\}$ be a $\frac{1}{8}$-covering of the sphere $\mathbb{S}^{d_0}$ in the Euclidean norm. From Example 5.8 in Wainwright (2019), there exists such a covering with $N \leq 17^{d_0+1}$ vectors. Thus, for any $v \in \mathbb{S}^{d_0}$, there exists a v^j in the cover such that $v = v^j + \Delta$ where Δ is an error vector satisfying $||\Delta||_2 \leq \frac{1}{8}$. Thus we have

$$
\begin{aligned}
|\langle v, \Gamma_k v\rangle| = |\langle v^j + \Delta, \Gamma_k(v^j + \Delta)\rangle| &= |\langle v^j, \Gamma_k v^j\rangle + 2\langle \Delta, \Gamma_k v^j\rangle + \langle \Delta, \Gamma_k \Delta\rangle| \\
&\leq |\langle v^j, \Gamma_k v^j\rangle| + 2||\Delta||_2|||\Gamma_k|||_2||v^j||_2 + |||\Gamma_k|||_2||\Delta||_2^2 \\
&\leq |\langle v^j, \Gamma_k v^j\rangle| + \frac{1}{4}|||\Gamma_k|||_2 + \frac{1}{64}|||\Gamma_k|||_2 \\
&\leq |\langle v^j, \Gamma_k v^j\rangle| + \frac{1}{2}|||\Gamma_k|||_2.
\end{aligned}
$$

By taking the maximum of the right-hand side over $j \in [N]$, we obtain

$$
|\langle v, \Gamma_k v\rangle| \leq \max_{j\in[N]} |\langle v^j, \Gamma_k v^j\rangle| + \frac{1}{2}|||\Gamma_k|||_2.
$$

Further taking the supremum of the left-hand side over $v \in \mathbb{S}^{d_0}$, we get

$$
|||\Gamma_k|||_2 = \max_{v\in\mathbb{S}^{d_0}} |\langle v, \Gamma_k v\rangle| \leq \max_{j\in[N]} |\langle v^j, \Gamma_k v^j\rangle| + \frac{1}{2}|||\Gamma_k|||_2.
$$

Therefore, we have $|||\Gamma_k|||_2 \leq 2\max_{j\in[N]} |\langle v^j, \Gamma_k v^j\rangle|$. Consequently, we have

$$
\begin{aligned}
\mathbb{E}(e^{\lambda|||\Gamma_k|||_2}) \leq \mathbb{E}\big(\exp(2\lambda \max_{j\in[N]} |\langle v^j, \Gamma_k v^j\rangle|)\big) &\leq \sum_{j=1}^{N} \mathbb{E}(e^{2\lambda|\langle v^j, \Gamma_k v^j\rangle|}) \\
&\leq \sum_{j=1}^{N} \big(\mathbb{E}(e^{2\lambda\langle v^j, \Gamma_k v^j\rangle}) + \mathbb{E}(e^{-2\lambda\langle v^j, \Gamma_k v^j\rangle})\big).
\end{aligned} \tag{5}
$$

For any fixed unit vector $v \in \mathbb{S}^{d_0}$, we have $v^\top \Gamma_k v = v^\top(\bar{\dot{x}}_k \bar{x}^\top + \bar{x}\bar{\dot{x}}_k^\top)v = 2(v^\top \bar{x})(v^\top \bar{\dot{x}}_k)$. We can rewrite $v^\top \bar{\dot{x}}_k$ as $v^\top(\frac{1}{n_k}\sum_{t\in\mathcal{E}_k} \dot{x}_t) = \sum_{t\in\mathcal{E}_k} \frac{1}{n_k}(v^\top \dot{x}_t)$. Note that we have proved before that $v^\top \dot{x}_t$ is sub-gaussian with variance proxy $\sigma^2 = 4(d_0 + 1)$. Since $\dot{x}_t$ are independent of each other across $t \in \mathcal{E}_k$, $v^\top \dot{x}_t$ are also independent for $t \in \mathcal{E}_k$.

Thus we conclude that $v^\top \bar{\dot{x}}_k = \sum_{t\in\mathcal{E}_k} \frac{1}{n_k}(v^\top \dot{x}_t)$ is sub-gaussian with variance proxy $\sum_{t\in\mathcal{E}_k}(\frac{1}{n_k})^2\sigma^2 = \frac{4(d_0+1)}{n_k}$ by the property of sum of independent sub-gaussian variables. Therefore, $v^\top \Gamma_k v = 2(v^\top \bar{x})(v^\top \bar{\dot{x}}_k)$ is sub-gaussian with variance proxy $\frac{4^2(d_0+1)^2}{n_k} = \frac{\sigma^4}{n_k}$ as $|v^\top \bar{x}| \le ||v||_2||\bar{x}||_2 \le ||v||_2\sqrt{d_0+1}||\bar{x}||_\infty = \sqrt{d_0+1}$. Thus we obtain

$$\mathbb{E}(e^{\alpha\langle v,\Gamma v\rangle}) \le e^{\frac{1}{2}\frac{\alpha^2\sigma^4}{n_k}},$$

for $\alpha \in \mathbb{R}$ and any fixed $v \in \mathbb{S}^{d_0}$.

Thus by applying the above result in the inequality (5), we obtain for any $\lambda \in \mathbb{R}$,

$$\mathbb{E}(e^{\lambda|||\Gamma_k|||_2}) \le 2Ne^{2\frac{\lambda^2\sigma^4}{n_k}} \le 2\cdot 17^{d_0+1}\cdot e^{2\frac{\lambda^2\sigma^4}{n_k}} \le e^{2\frac{\lambda^2\sigma^4}{n_k}+4(d_0+1)}.$$

Then by applying the Chernoff bound, we obtain for any $\gamma > 0$,

$$\begin{aligned}\mathbb{P}(|||\Gamma_k|||_2 \ge \gamma) &\le \inf_{\lambda\ge 0}\frac{\mathbb{E}(e^{\lambda|||\Gamma_k|||_2})}{e^{\lambda\gamma}} \le \inf_{\lambda\ge 0}\frac{e^{2\frac{\lambda^2\sigma^4}{n_k}+4(d_0+1)}}{e^{\lambda\gamma}}\\ &= \inf_{\lambda\ge 0} e^{2\frac{\lambda^2\sigma^4}{n_k}-\lambda\gamma+4(d_0+1)} = \inf_{\lambda\ge 0} e^{\frac{2\sigma^4}{n_k}(\lambda-\frac{\gamma n_k}{4\sigma^4})^2-\frac{n_k\gamma^2}{8\sigma^4}+4(d_0+1)}\\ &= e^{-\frac{n_k\gamma^2}{8\sigma^4}+4(d_0+1)}.\end{aligned}$$

By substituting γ with $\sigma^2(\sqrt{\frac{32(d_0+1)}{n_k}}+\delta)$, we obtain for any $\delta \ge 0$,

$$\mathbb{P}\Big(\frac{|||\Gamma_k|||_2}{\sigma^2} \ge \sqrt{\frac{32(d_0+1)}{n_k}}+\delta\Big) \le e^{-\frac{n_k\sigma^4(\sqrt{\frac{32(d_0+1)}{n_k}}+\delta)^2}{8\sigma^4}+4(d_0+1)} \le e^{-\frac{n_k\delta^2}{8}}.$$

Let $\delta = \frac{c_0}{32(d_0+1)}$. Then for $n_k \ge \frac{2^{15}}{c_0^2}(d_0+1)^3$, we have $\sigma^2\big(\sqrt{\frac{32(d_0+1)}{n_k}}+\delta\big) \le \frac{c_0}{8}+\frac{c_0}{8} = \frac{c_0}{4}$. Therefore, we have

$$\mathbb{P}\big(|||\Gamma_k|||_2 \ge \frac{c_0}{4}\big) \le e^{-\frac{c_0^2}{8096(d_0+1)^2}n_k}, \tag{6}$$

for $n_k \ge \frac{2^{15}}{c_0^2}(d_0+1)^3$.

Now we are ready to lower bound $\nabla^2 L_k(\xi^*) = 2\hat{\Sigma}_k$. By Equation (4) and Equation (6), we have that for $n_k \geq \max\{\frac{64^2 c_1^2}{\tilde{c}_0^2}(d_0+1)^3, \frac{2^{15}}{c_0^2}(d_0+1)^3\}$, $|||\dot{\Sigma}_k - \Sigma^*|||_2 \leq \frac{\tilde{c}_0}{4}$ and $|||\Gamma_k|||_2 \leq \frac{c_0}{4}$ hold simultaneously with probability at least $1 - c_2 e^{-\frac{c_3 \tilde{c}_0^2}{1024(d_0+1)^2} n_k} - e^{-\frac{c_0^2}{8096(d_0+1)^2} n_k}$. Therefore, on this high probability event,

$$|||-(\Sigma^* - \dot{\Sigma}_k) + \Gamma_k|||_2 \leq |||-(\Sigma^* - \dot{\Sigma}_k)|||_2 + |||\Gamma_k|||_2 \leq \frac{\tilde{c}_0}{4} + \frac{c_0}{4} \leq \frac{c_0}{2}.$$

By Equation (3), we have $\hat{\Sigma}_k = \Sigma - (\Sigma^* - \dot{\Sigma}_k) + \Gamma_k$. By Assumption 2, $\Sigma - c_0\mathbb{I}$ is positive-definite. Thus we have on the high probability event that

$$\begin{aligned}
&(\hat{\xi}_k - \xi^*)^\top \hat{\Sigma}_k (\hat{\xi}_k - \xi^*) = (\hat{\xi}_k - \xi^*)^\top (\Sigma - (\Sigma^* - \dot{\Sigma}_k) + \Gamma_k)(\hat{\xi}_k - \xi^*) \\
&= (\hat{\xi}_k - \xi^*)^\top \Sigma (\hat{\xi}_k - \xi^*) + (\hat{\xi}_k - \xi^*)^\top (-(\Sigma^* - \dot{\Sigma}_k) + \Gamma_k)(\hat{\xi}_k - \xi^*) \\
&\geq c_0 ||\hat{\xi}_k - \xi^*||_2^2 - \frac{c_0}{2}||\hat{\xi}_k - \xi^*||_2^2 = \frac{c_0}{2}||\hat{\xi}_k - \xi^*||_2^2.
\end{aligned}$$

Since $\nabla^2 L_k(\xi^*) = 2\hat{\Sigma}_k$, we have on that high probability event,

$$(\hat{\xi}_k - \xi^*)^\top \nabla^2 L_k(\xi^*)(\hat{\xi}_k - \xi^*) \geq c_0 ||\hat{\xi}_k - \xi^*||_2^2 \geq \frac{c_0}{d_0+1}||\hat{\xi}_k - \xi^*||_1^2.$$

Combined with Equation (1) and Equation (2), we have for $n_k \geq \max\{d_0+1, \frac{64^2 c_1^2}{\tilde{c}_0^2}(d_0+1)^3, \frac{2^{15}}{c_0^2}(d_0+1)^3\}$,

$$\begin{aligned}
\frac{c_0}{d_0+1}||\hat{\xi}_k - \xi^*||_1^2 &\leq (\hat{\xi}_k - \xi^*)^\top \nabla^2 L_k(\xi^*)(\hat{\xi}_k - \xi^*) \\
&\leq 2||\xi^* - \hat{\xi}_k||_1 ||\nabla L_k(\xi^*)||_\infty \\
&\leq 8(B+U+W)\sqrt{\frac{\log n_k}{n_k}}||\xi^* - \hat{\xi}_k||_1
\end{aligned}$$

with probability at least $1 - \frac{2}{n_k} - c_2 e^{-\frac{c_3 \tilde{c}_0^2}{1024(d_0+1)^2} n_k} - e^{-\frac{c_0^2}{8096(d_0+1)^2} n_k}$.

Let $\tilde{c}_1 = c_2 + 1, \tilde{c}_2 = \min\{\frac{c_3 \tilde{c}_0^2}{1024}, \frac{c_0^2}{8096}\}, \tilde{c}_3 = \max\{\frac{64^2 c_1^2}{\tilde{c}_0^2}, \frac{2^{15}}{c_0^2}\}$. Then we obtain that for

$n_k \geq \tilde{c}_3(d_0+1)^3$,

$$||\hat{\theta}_k - \theta_0||_1 \leq ||\hat{\xi}_k - \xi^*||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}$$

with probability at least $1 - \frac{2}{n_k} - \tilde{c}_1 e^{-\frac{\tilde{c}_2}{(d_0+1)^2}n_k}$. □

A.2 Proof of Proposition 1

Proposition 1. *Under Assumptions 1 and 4, there exists positive constants $C_1^{'}, C_2^{'}$ and $C_3^{'}$ such that with probability at least $1 - \frac{1}{T_0}$, the Inner UCB Algorithm yields a discrete-part regret*

$$R_{T_0,1} \leq C_1^{'} d\sqrt{T_0}\log(C_2^{'}T_0) + C_3^{'}L||\hat{\theta} - \theta_0||_1 T_0.$$

Proof. By Lemmas 1 – 2 in Luo et al. (2021), the UCB phase pricing problem has an equivalent Perturbed Linear Bandit (PLB) formulation and the Inner UCB Algorithm is equivalent to a slightly modified version of the LinUCB Algorithm (Abbasi-Yadkori et al., 2011). By Theorem 1 in Luo et al. (2021), with the choice of $\beta_t = \beta_t^* = p_{\max}^2(1 \vee (\frac{1}{p_{\max}}\sqrt{\lambda d} + \sqrt{2\log(T_0) + d\log(\frac{d\lambda + (t-1)p_{\max}^2}{d\lambda})})^2)$, the Inner UCB Algorithm yields a discrete part regret $R_{T_0,1}$ satisfying

$$R_{T_0,1} \leq 2\sqrt{2dT_0\beta_{T_0}^* \log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} + 4T_0 L||\theta_0 - \hat{\theta}||_1 p_{\max} + 2dp_{\max} \tag{7}$$

with probability at least $1 - \frac{1}{T_0}$.

Denote $A_1 = p_{\max}, A_2 = \sqrt{\lambda d}, A_3 = p_{\max}\sqrt{2\log(T_0) + d\log(\frac{d\lambda + (T_0-1)p_{\max}^2}{d\lambda})}$. Then $\sqrt{\beta_{T_0}^*} = A_1 \vee (A_2 + A_3)$. Let $C_4^{'} = \max\{1 + \frac{p_{\max}^2}{\lambda}, 3\}$, then

$$1 + \frac{p_{\max}^2}{\lambda}T_0 \leq (1 + \frac{p_{\max}^2}{\lambda})T_0 \leq C_4^{'}T_0 \Rightarrow \log(1 + \frac{p_{\max}^2}{\lambda}T_0) \leq \log(C_4^{'}T_0).$$

Since $C_4' \geq 3$, $\log(C_4' T_0) \geq 1$. Therefore, we have

$$\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} \leq \sqrt{\log(1 + \frac{p_{\max}^2}{\lambda} T_0)} \leq \sqrt{\log(C_4' T_0)} \leq \log(C_4' T_0). \tag{8}$$

On the other hand, we have

$$\begin{aligned}
\frac{A_3}{p_{\max}} &= \sqrt{2\log(T_0) + d\log(\frac{d\lambda + (T_0 - 1)p_{\max}^2}{d\lambda})} \\
&\leq \sqrt{d}\sqrt{2\log(T_0) + \log(\frac{d\lambda + (T_0 - 1)p_{\max}^2}{d\lambda})} \\
&\leq \sqrt{d}\sqrt{2\log(T_0) + \log(1 + \frac{p_{\max}^2}{d\lambda} T_0)} \leq \sqrt{d}\sqrt{2\log(T_0) + \log(1 + \frac{p_{\max}^2}{\lambda} T_0)} \\
\text{(By (8))} &\leq \sqrt{d}\sqrt{2\log(T_0) + \log(C_4' T_0)} \leq \sqrt{d}\sqrt{3\log(C_4' T_0)}.
\end{aligned} \tag{9}$$

Therefore, we have

$$A_1\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} \leq p_{\max}\log(C_4' T_0) \leq C_5' \log(C_4' T_0)\sqrt{d} \text{ where } C_5' = p_{\max},$$

$$A_2\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} \leq \sqrt{\lambda d}\log(C_4' T_0) \leq C_6' \log(C_4' T_0)\sqrt{d} \text{ where } C_6' = \sqrt{\lambda},$$

$$\begin{aligned}
A_3\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} &= p_{\max}\sqrt{2\log(T_0) + d\log(\frac{d\lambda + (T_0 - 1)p_{\max}^2}{d\lambda})}\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} \\
\text{(By (8) and (9))} &\leq p_{\max}\sqrt{d}\sqrt{3\log(C_4' T_0)}\sqrt{\log(C_4' T_0)} \\
&= \sqrt{3}p_{\max}\sqrt{d}\log(C_4' T_0) \leq C_7' \log(C_4' T_0)\sqrt{d} \text{ where } C_7' = \sqrt{3}p_{\max}.
\end{aligned}$$

Let $C_8' = \max\{C_5', C_6' + C_7'\}$, then we have

$$\begin{aligned}
&\sqrt{\beta_{T_0}^* \log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} = (A_1 \vee (A_2 + A_3))\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} \\
=&(A_1\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})}) \vee (A_2\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})} + A_3\sqrt{\log(\frac{d\lambda + T_0 p_{\max}^2}{d\lambda})}) \\
\leq&(C_5' \log(C_4' T_0)\sqrt{d}) \vee (C_6' \log(C_4' T_0)\sqrt{d} + C_7' \log(C_4' T_0)\sqrt{d}) \\
=&(C_5' \vee (C_6' + C_7'))\log(C_4' T_0)\sqrt{d} = C_8' \log(C_4' T_0)\sqrt{d}.
\end{aligned}$$

Thus we have

$$
\begin{aligned}
&2\sqrt{2dT_0\beta^*_{T_0}\log(\frac{d\lambda+T_0p^2_{\max}}{d\lambda})}+4T_0L||\theta_0-\hat{\theta}||_1p_{\max}+2dp_{\max}\\
&\leq 2\sqrt{2}C_8'd\sqrt{T_0}\log(C_4'T_0)+2p_{\max}d+4T_0L||\theta_0-\hat{\theta}||_1p_{\max}\\
&\leq C_1'd\sqrt{T_0}\log(C_2'T_0)+C_3'L||\hat{\theta}-\theta_0||_1T_0,
\end{aligned}
$$

where $C_1' = \max\{2\sqrt{2}C_8', 2p_{\max}\}, C_2' = C_4'$ and $C_3' = 4p_{\max}$. Therefore, by Equation (7), the discrete-part regret satisfies

$$
R_{T_0,1} \leq C_1'd\sqrt{T_0}\log(C_2'T_0)+C_3'L||\hat{\theta}-\theta_0||_1T_0
$$

with probability at least $1-\frac{1}{T_0}$. □

A.3 Proof of Theorem 1

Theorem 1. *Under Assumptions 1 – 5, by choosing $\beta = \frac{2}{3}$ and $\gamma = \frac{1}{6}$ in Algorithm 1, the expected regret satisfies $\mathbb{E}(R_T) = \tilde{O}(d_0^2T^{2/3}) = \tilde{O}(T^{2/3})$.*

Proof. Note that the last episode can be incomplete. Nevertheless, its regret will be upper bounded by that of the completed version. Moreover, in Algorithm 1, the parameters for the last episode are just set as if it would be in its full projected length. Thus without loss of generality, we can assume a complete last episode.

Denote $\ell_{k,e} = \lceil C_1\ell_k^{2/3}\rceil$ and $\ell_{k,u} = \ell_k - \lceil C_1\ell_k^{2/3}\rceil$ as the exploration phase length and the UCB phase length in episode k. Denote $R_e^{(k)}$ as the exploration phase regret and $R_u^{(k)}$ as the UCB phase regret in episode k. Denote $R^{(k)}$ as the overall regret in episode k and we have $R^{(k)} = R_e^{(k)} + R_u^{(k)}$. Further denote $R_{u,1}^{(k)}, R_{u,2}^{(k)}$ as the discrete-part and continuous-part regret in the UCB phase of episode k. Then we have $R_u^{(k)} = R_{u,1}^{(k)}+R_{u,2}^{(k)}$.

By Lemma 1, we have for $n_k = \ell_{k,e} \geq \tilde{c}_3(d_0+1)^3$,

$$
||\hat{\theta}_k-\theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}
$$

with probability at least $1 - \frac{2}{n_k} - \tilde{c}_1 e^{-\frac{\tilde{c}_2}{(d_0+1)^2} n_k}$. Denote this high probability event as $\mathcal{P}_k$. Then on $\mathcal{P}_k$, we have

$$||\hat{\theta}_k - \theta_0||_1 \le C_3 \sqrt{\frac{\log n_k}{n_k}}, \quad ||\hat{\theta}_k||_1 \le ||\hat{\theta}_k - \theta_0||_1 + ||\theta_0||_1 \le C_3,$$

where $C_3 = W + \frac{8(B+U+W)(d_0+1)}{c_0}$ is a constant. Moreover, there exists $C_4 = \tilde{c}_4(d_0 + 1)^3 \ge \tilde{c}_3(d_0+1)^3$ for some constant $\tilde{c}_4$ such that for any $n_k = \ell_{k,e} \ge C_4$, we have $\tilde{c}_1 e^{-\frac{\tilde{c}_2}{(d_0+1)^2} n_k} \le \frac{1}{n_k}$. Then the event $\mathcal{P}_k$ happens with probability at least $1 - \frac{3}{n_k}$ for the episodes with $n_k = \ell_{k,e} \ge C_4$.

We first bound the continuous-part regret at each time period t in the UCB phase of episode k. It admits the form

$$p_t^*(1 - F(p_t^* - x_t^\top \theta_0)) - \tilde{p}_t^*(1 - F(\tilde{p}_t^* - x_t^\top \theta_0)) = f_{x_t^\top \theta_0}(p_t^*) - f_{x_t^\top \theta_0}(\tilde{p}_t^*),$$

where $f_q(p) = p(1 - F(p - q))$ as defined in Assumption 5.

By our discretization approach, $\{m_i + x_t^\top \hat{\theta}\}_{i \in [d_k]}$ are a sequence of points with a special pattern that any two consecutive points have a difference $|(m_{i+1} + x_t^\top \hat{\theta}) - (m_i + x_t^\top \hat{\theta})| = \frac{|G(\hat{\theta}_k)|}{d_k}$. Moreover, the left-most point satisfies $m_1 + x_t^\top \hat{\theta} \le \frac{1}{2} \cdot \frac{|G(\hat{\theta}_k)|}{d_k}$ while the right-most point satisfies $m_d + x_t^\top \hat{\theta} \ge p_{\max} - \frac{1}{2} \cdot \frac{|G(\hat{\theta}_k)|}{d_k}$. Since $\mathcal{S}_t = \{m_j + x_t^\top \hat{\theta}_k | j \in [d_k], m_j + x_t^\top \hat{\theta}_k \in (0, p_{\max})\}$ and $p_t^* \in (0, p_{\max})$, there must be some price $\dot{p}_t \in \mathcal{S}_t$ whose distance with p_t^* is less than $\frac{|G(\hat{\theta}_k)|}{d_k}$, i.e., $|p_t^* - \dot{p}_t| \le \frac{|G(\hat{\theta}_k)|}{d_k}$. Thus by Assumption 5, there exists a constant $C > 0$ such that $f_{x_t^\top \theta_0}(p_t^*) - f_{x_t^\top \theta_0}(\dot{p}_t) \le C(p_t^* - \dot{p}_t)^2$. Then on the high probability event $\mathcal{P}_k$, we have $f_{x_t^\top \theta_0}(p_t^*) - f_{x_t^\top \theta_0}(\dot{p}_t) \le C\frac{|G(\hat{\theta}_k)|^2}{d_k^2} \le \frac{CC_6^2}{d_k^2}$ where the constant $C_6 = p_{\max} + 2C_3 \ge p_{\max} + 2||\hat{\theta}_k||_1 = |G(\hat{\theta}_k)|$. Since $\tilde{p}_t^*$ yields the maximum reward for prices in $\mathcal{S}_t$, we have $f_{x_t^\top \theta_0}(\tilde{p}_t^*) \ge f_{x_t^\top \theta_0}(\dot{p}_t)$. Thus we have

$$f_{x_t^\top \theta_0}(p_t^*) - f_{x_t^\top \theta_0}(\tilde{p}_t^*) \le f_{x_t^\top \theta_0}(p_t^*) - f_{x_t^\top \theta_0}(\dot{p}_t) \le \frac{CC_6^2}{d_k^2} = \frac{C_5}{d_k^2},$$

where the constant $C_5 = CC_6^2$. Therefore, the continous-part regret $R_{u,2}^{(k)}$ in the UCB phase of episode k can be bounded as $R_{u,2}^{(k)} \leq \frac{C_5 \ell_{k,u}}{d_k^2}$ on the high probability event $\mathcal{P}_k$, i.e., when $\hat{\theta}_k$ satisfies $||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}$.

On the other hand, by Proposition 1, the discrete-part regret in the UCB phase of episode k satisfies $R_{u,1}^{(k)} \leq C_1' d_k \sqrt{\ell_{k,u}} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u}$ with probability at least $1 - \frac{1}{\ell_{k,u}}$ conditional on $\hat{\theta}_k$. Thus we have

$$\begin{aligned}
\mathbb{E}(R_{u,1}^{(k)}|\hat{\theta}_k) &\leq C_1' d_k \sqrt{\ell_{k,u}} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u} + \frac{p_{\max}}{\ell_{k,u}} \cdot \ell_{k,u} \\
&\leq C_1' d_k \sqrt{\ell_{k,u}} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u} + p_{\max}.
\end{aligned}$$

Combined with our previously derived continuous-part regret result, we have for $\hat{\theta}_k$ such that $||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}$,

$$\begin{aligned}
\mathbb{E}(R_u^{(k)}|\hat{\theta}_k) &= \mathbb{E}(R_{u,1}^{(k)}|\hat{\theta}_k) + \mathbb{E}(R_{u,2}^{(k)}|\hat{\theta}_k) \\
&\leq C_1' d_k \sqrt{\ell_{k,u}} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u} + p_{\max} + \frac{C_5 \ell_{k,u}}{d_k^2}.
\end{aligned}$$

As we choose $d_k = \lceil C_2 \ell_{k,u}^{1/6} \rceil$ that satisfies $C_2 \ell_{k,u}^{1/6} \leq d_k \leq C_2 \ell_{k,u}^{1/6} + 1$, we obtain

$$\mathbb{E}(R_u^{(k)}|\hat{\theta}_k) \leq C_4' \ell_{k,u}^{2/3} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u},$$

where the constant $C_4' = C_1' C_2 + C_1 + \frac{C_5}{C_2^2} + p_{\max}$. Since $\mathcal{P}_k$ is in the σ-field generated by $\hat{\theta}_k$, we obtain

$$\begin{aligned}
\mathbb{E}(R_u^{(k)}) &= \mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k}) + \mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k^c}) \leq \mathbb{E}\big(\mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k}^2 | \hat{\theta}_k)\big) + \mathbb{P}(\mathcal{P}_k^c) \cdot p_{\max}\ell_{k,u} \\
&\leq \mathbb{E}\big(1_{\mathcal{P}_k}\mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k} | \hat{\theta}_k)\big) + \frac{3}{\ell_{k,e}} \cdot p_{\max}\ell_{k,u} \\
&\leq \mathbb{E}\big(1_{\mathcal{P}_k}(C_4' \log(C_2'\ell_{k,u})\ell_{k,u}^{2/3} + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u})\big) + \frac{3}{\lceil C_1 \ell_k^{2/3} \rceil} \cdot p_{\max}\ell_k \\
&\leq C_4' \log(C_2'\ell_{k,u})\ell_{k,u}^{2/3} + C_3' L (C_3 \sqrt{\frac{\log \ell_{k,e}}{\ell_{k,e}}})\ell_{k,u} + \frac{3p_{\max}}{C_1}\ell_k^{1/3} \\
&\leq C_4' \log(C_2'\ell_k)\ell_k^{2/3} + C_3' L (C_3 \sqrt{\frac{\log \ell_k}{\lceil C_1 \ell_k^{2/3} \rceil}})\ell_k + \frac{3p_{\max}}{C_1}\ell_k^{1/3} \\
&\leq \big(C_4' \log(C_2'\ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}}\sqrt{\log \ell_k} + \frac{3p_{\max}}{C_1}\big)\ell_k^{2/3} = \tilde{O}(\ell_k^{2/3}).
\end{aligned}$$

On the other hand, the expected regret in the exploration phase satisfies $\mathbb{E}(R_e^{(k)}) \leq p_{\max}\ell_{k,e} \leq p_{\max}\lceil C_1 \ell_k^{2/3} \rceil \leq C_7 \ell_k^{2/3} = \tilde{O}(\ell_k^{2/3})$ where $C_7 = p_{\max}C_1 + p_{\max}$. Therefore the overall expected regret in episode k satisfies $\mathbb{E}(R^{(k)}) = \mathbb{E}(R_u^{(k)}) + \mathbb{E}(R_e^{(k)}) = \tilde{O}(\ell_k^{2/3})$.

Note that this only happens for $n_k = \ell_{k,e} = \lceil C_1 \ell_k^{2/3} \rceil \geq C_4$. For the episode such that its length ℓ_k does not satisfy $\lceil C_1 \ell_k^{2/3} \rceil \geq C_4$, we have $\mathbb{E}(R^{(k)}) \leq p_{\max}\ell_k \leq \frac{\tilde{c}_4^{1/2} p_{\max}}{C_1^{1/2}} \ell_k^{2/3} (d_0+1)^{3/2} \leq C_5' d_0^{3/2} \ell_k^{2/3}$ where C_5' is a constant. Thus for any episode k, we have $\mathbb{E}(R^{(k)}) \leq \big(C_5' d_0^{3/2} + C_4' \log(C_2'\ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}}\sqrt{\log \ell_k} + C_6'\big)\ell_k^{2/3} = \tilde{O}(\ell_k^{2/3})$ where the constant $C_6' = \frac{3p_{\max}}{C_1} + C_7$. Here some constants contain the dimensionality d_0. We have $C_3 = W + \frac{8(B+U+W)(d_0+1)}{c_0} = O(d_0)$, $C_6 = p_{\max} + 2C_3 = O(d_0)$, $C_5 = CC_6^2 = O(d_0^2)$ and $C_4' = C_1' C_2 + C_1 + \frac{C_5}{C_2^2} + p_{\max} = O(d_0^2)$.

Denote $K = n(T, \alpha_1)$ as the number of episodes. Note that $\ell_K = 2\ell_{K-1} \leq 2T$. Now we bound the expected regret for the entire horizon as

$$
\begin{aligned}
\mathbb{E}(R_T) &= \mathbb{E}(\sum_{k=1}^{K} R^{(k)}) = \sum_{k=1}^{K} \mathbb{E}(R^{(k)}) \\
&\leq \sum_{k=1}^{K} \Big(\big(C_5' d_0^{3/2} + C_4' \log(C_2' \ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log \ell_k} + C_6'\big) \ell_k^{2/3} \Big) \\
&\leq \big(C_5' d_0^{3/2} + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6'\big) \sum_{k=1}^{K} \ell_k^{2/3} \\
&= \big(C_5' d_0^{3/2} + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6'\big) \alpha_1^{2/3} \frac{(2^{2/3})^K - 1}{2^{2/3} - 1} \\
&= \frac{2^{2/3}}{2^{2/3} - 1} \big(C_5' d_0^{3/2} + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6'\big) \ell_K^{2/3} \\
&\leq \frac{2^{4/3}}{2^{2/3} - 1} \big(C_5' d_0^{3/2} + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6'\big) T^{2/3} \\
&= \tilde{O}(d_0^2 T^{2/3}) = \tilde{O}(T^{2/3}).
\end{aligned}
$$

□

A.4 Proof of Theorem 2

Theorem 2. *Under Assumptions 1 – 4, by choosing* $\beta = \frac{3}{4}$ *and* $\gamma = \frac{1}{4}$ *in Algorithm 1, the expected regret satisfies* $\mathbb{E}(R_T) = \tilde{O}(d_0 T^{3/4}) = \tilde{O}(T^{3/4})$.

Proof. We use the same notation of $R^{(k)}, R_e^{(k)}, R_u^{(k)}, R_{u,1}^{(k)}, R_{u,2}^{(k)}$ and $\ell_{k,e}, \ell_{k,u}$ as in the proof for Theorem 1. Note that now $\ell_{k,e} = \lceil C_1 \ell_k^{3/4} \rceil$ and $\ell_{k,u} = \ell_k - \lceil C_1 \ell_k^{3/4} \rceil$. Similar to the proof of Theorem 1, we assume a complete last episode without loss of generality.

Similar to the proof of Theorem 1, we obtain that for any $n_k = \ell_{k,e} \geq C_4 = \tilde{c}_4(d_0 + 1)^2$, the event $\mathcal{P}_k$ that $||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0} \sqrt{\frac{\log n_k}{n_k}}$ happens with probability at least $1 - \frac{3}{n_k}$. Furthermore, on this event, we have

$$
||\hat{\theta}_k - \theta_0||_1 \leq C_3 \sqrt{\frac{\log n_k}{n_k}}, \quad ||\hat{\theta}_k||_1 \leq ||\hat{\theta}_k - \theta_0||_1 + ||\theta_0||_1 \leq C_3.
$$

We first bound the continuous-part regret at each time period t in the UCB phase of episode k. It admits the form

$$p_t^*(1 - F(p_t^* - x_t^\top \theta_0)) - \tilde{p}_t^*(1 - F(\tilde{p}_t^* - x_t^\top \theta_0)),$$

where p_t^* is the overall best price and $\tilde{p}_t^*$ is the discrete best price among the candidate set $\mathcal{S}_t = \{m_j + x_t^\top \hat{\theta}_k | j \in [d_k], m_j + x_t^\top \hat{\theta}_k \in (0, p_{\max})\}$.

Let the maximum value in an empty set be $-\infty$. Denote the price $\dot{p}_t = \max\{0, \max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\}\}$. Then there are several cases.

1. $\{m_j + x_t^\top \hat{\theta}_k | j \in [d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*\} = \emptyset$. Then $\max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} = -\infty$ and $\dot{p}_t = 0$. Moreover, we have $p_t^* \le \min_{j\in[d_k]}(m_j + x_t^\top \hat{\theta}_k) = m_1 + x_t^\top \hat{\theta}_k \le -||\hat{\theta}_k||_1 + \frac{|G(\hat{\theta}_k)|}{2d_k} + ||x_t||_\infty ||\hat{\theta}_k||_1 = \frac{|G(\hat{\theta}_k)|}{2d_k}$. Thus $p_t^* - \dot{p}_t \in [0, \frac{|G(\hat{\theta}_k)|}{d_k}]$.

2. $\{m_j + x_t^\top \hat{\theta}_k | j \in [d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*\} \ne \emptyset$ and $\max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} < m_{d_k} + x_t^\top \hat{\theta}_k$. Then we have $\max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} \le p_t^* \le \max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} + \frac{|G(\hat{\theta}_k)|}{d_k}$. Thus we have $\dot{p}_t \le p_t^* \le \dot{p}_t + \frac{|G(\hat{\theta}_k)|}{d_k}$. Namely, $p_t^* - \dot{p}_t \in [0, \frac{|G(\hat{\theta}_k)|}{d_k}]$.

3. $\{m_j + x_t^\top \hat{\theta}_k | j \in [d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*\} \ne \emptyset$ and $\max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} = m_{d_k} + x_t^\top \hat{\theta}_k$. Now since $m_{d_k} + x_t^\top \hat{\theta}_k \ge p_{\max} + ||\hat{\theta}_k||_1 - \frac{|G(\hat{\theta}_k)|}{2d_k} - ||x_t||_\infty ||\hat{\theta}_k||_1 \ge p_{\max} - \frac{|G(\hat{\theta}_k)|}{2d_k}$, we have $p_{\max} - \frac{|G(\hat{\theta}_k)|}{2d_k} \le m_{d_k} + x_t^\top \hat{\theta}_k \le p_t^* \le p_{\max}$. Thus we have $p_t^* - (m_{d_k} + x_t^\top \hat{\theta}_k) \le \frac{|G(\hat{\theta}_k)|}{2d_k} \le \frac{|G(\hat{\theta}_k)|}{d_k}$. Therefore, we have $p_t^* - \dot{p}_t \in [0, \frac{|G(\hat{\theta}_k)|}{d_k}]$ since $p_t^* \ge \dot{p}_t \ge \max_{j\in[d_k], m_j + x_t^\top \hat{\theta}_k \le p_t^*}\{m_j + x_t^\top \hat{\theta}_k\} = m_{d_k} + x_t^\top \hat{\theta}_k$.

Namely, we will always have $p_t^* - \dot{p}_t \in [0, \frac{|G(\hat{\theta}_k)|}{d_k}]$. Thus we have

$$\begin{aligned}
&p_t^*\big(1 - F(p_t^* - x_t^\top \theta_0)\big) - \dot{p}_t\big(1 - F(\dot{p}_t - x_t^\top \theta_0)\big) \\
\le\ & p_t^*\big(1 - F(\dot{p}_t - x_t^\top \theta_0)\big) - \dot{p}_t\big(1 - F(\dot{p}_t - x_t^\top \theta_0)\big) = (p_t^* - \dot{p}_t)\big(1 - F(\dot{p}_t - x_t^\top \theta_0)\big) \\
\le\ & p_t^* - \dot{p}_t \le \frac{|G(\hat{\theta}_k)|}{d_k}.
\end{aligned}$$

Now we compare the revenue of $\dot{p}_t$ and the discrete best price $\tilde{p}_t^*$. If $\dot{p}_t = 0$, then $0 = \dot{p}_t\big(1-F(\dot{p}_t-x_t^\top\theta_0)\big) \leq \tilde{p}_t^*\big(1-F(\tilde{p}_t^*-x_t^\top\theta_0)\big)$. If $\dot{p}_t \neq 0$, then $\dot{p}_t = \max_{j\in[d_k], m_j+x_t^\top\hat{\theta}_k\leq p_t^*}\{m_j+x_t^\top\hat{\theta}_k\} \in (0, p_t^*) \subseteq (0, p_{\max})$. Thus $\dot{p}_t \in \mathcal{S}_t$ and its revenue is not greater than the discrete best price $\tilde{p}_t^* \in \mathcal{S}_t$. Thus it always holds that $\dot{p}_t\big(1-F(\dot{p}_t-x_t^\top\theta_0)\big) \leq \tilde{p}_t^*\big(1-F(\tilde{p}_t^*-x_t^\top\theta_0)\big)$. Thus we can bound the continuous-part regret at time period t as

$$\begin{aligned}
&p_t^*(1-F(p_t^*-x_t^\top\theta_0)) - \tilde{p}_t^*(1-F(\tilde{p}_t^*-x_t^\top\theta_0)) \\
\leq\; &p_t^*(1-F(p_t^*-x_t^\top\theta_0)) - \dot{p}_t(1-F(\dot{p}_t-x_t^\top\theta_0)) \\
\leq\; &\frac{|G(\hat{\theta}_k)|}{d_k} \leq \frac{p_{\max}+2C_3}{d_k} = \frac{C_6}{d_k}
\end{aligned}$$

on the high probability event $\mathcal{P}_k$, where the constant $C_6 = p_{\max} + 2C_3$. Thus the overall continuous-part regret $R_{u,2}^{(k)}$ in the UCB phase of episode k can be bounded as $R_{u,2}^{(k)} \leq \frac{C_6\ell_{k,u}}{d_k}$ when $\hat{\theta}_k$ satisfies $||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}$.

On the other hand, by Proposition 1, the discrete-part regret in the UCB phase of episode k satisfies $R_{u,1}^{(k)} \leq C_1' d_k\sqrt{\ell_{k,u}}\log(C_2'\ell_{k,u}) + C_3'L||\hat{\theta}_k - \theta_0||_1\ell_{k,u}$ with probability at least $1 - \frac{1}{\ell_{k,u}}$ conditional on $\hat{\theta}_k$. Thus we have

$$\begin{aligned}
\mathbb{E}(R_{u,1}^{(k)}|\hat{\theta}_k) &\leq C_1' d_k\sqrt{\ell_{k,u}}\log(C_2'\ell_{k,u}) + C_3'L||\hat{\theta}_k - \theta_0||_1\ell_{k,u} + \frac{p_{\max}}{\ell_{k,u}}\cdot\ell_{k,u} \\
&\leq C_1' d_k\sqrt{\ell_{k,u}}\log(C_2'\ell_{k,u}) + C_3'L||\hat{\theta}_k - \theta_0||_1\ell_{k,u} + p_{\max}.
\end{aligned}$$

Combined with our previously derived continuous-part regret result, we have for $\hat{\theta}_k$ such that $||\hat{\theta}_k - \theta_0||_1 \leq \frac{8(B+U+W)(d_0+1)}{c_0}\sqrt{\frac{\log n_k}{n_k}}$,

$$\begin{aligned}
\mathbb{E}(R_u^{(k)}|\hat{\theta}_k) &= \mathbb{E}(R_{u,1}^{(k)}|\hat{\theta}_k) + \mathbb{E}(R_{u,2}^{(k)}|\hat{\theta}_k) \\
&\leq C_1' d_k\sqrt{\ell_{k,u}}\log(C_2'\ell_{k,u}) + C_3'L||\hat{\theta}_k - \theta_0||_1\ell_{k,u} + p_{\max} + \frac{C_6\ell_{k,u}}{d_k}.
\end{aligned}$$

As we choose $d_k = \lceil C_2 \ell_{k,u}^{1/4} \rceil$ that satisfies $C_2 \ell_{k,u}^{1/4} \leq d_k \leq C_2 \ell_{k,u}^{1/4} + 1$, we obtain

$$\mathbb{E}(R_u^{(k)}|\hat{\theta}_k) \leq C_4' \ell_{k,u}^{3/4} \log(C_2' \ell_{k,u}) + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u},$$

where the constant $C_4' = C_1' C_2 + C_1 + \frac{C_6}{C_2} + p_{\max}$. Since $\mathcal{P}_k$ is in the σ-field generated by $\hat{\theta}_k$, we obtain

$$\begin{aligned}
\mathbb{E}(R_u^{(k)}) &= \mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k}) + \mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k^c}) \leq \mathbb{E}\big(\mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k}^2 | \hat{\theta}_k)\big) + \mathbb{P}(\mathcal{P}_k^c) \cdot p_{\max} \ell_{k,u} \\
&\leq \mathbb{E}\big(1_{\mathcal{P}_k} \mathbb{E}(R_u^{(k)} 1_{\mathcal{P}_k} | \hat{\theta}_k)\big) + \frac{3}{\ell_{k,e}} \cdot p_{\max} \ell_{k,u} \\
&\leq \mathbb{E}\big(1_{\mathcal{P}_k} (C_4' \log(C_2' \ell_{k,u}) \ell_{k,u}^{3/4} + C_3' L ||\hat{\theta}_k - \theta_0||_1 \ell_{k,u})\big) + \frac{3}{\lceil C_1 \ell_k^{3/4} \rceil} \cdot p_{\max} \ell_k \\
&\leq C_4' \log(C_2' \ell_{k,u}) \ell_{k,u}^{3/4} + C_3' L (C_3 \sqrt{\frac{\log \ell_{k,e}}{\ell_{k,e}}}) \ell_{k,u} + \frac{3 p_{\max}}{C_1} \ell_k^{1/4} \\
&\leq C_4' \log(C_2' \ell_k) \ell_k^{3/4} + C_3' L (C_3 \sqrt{\frac{\log \ell_k}{\lceil C_1 \ell_k^{3/4} \rceil}}) \ell_k + \frac{3 p_{\max}}{C_1} \ell_k^{1/4} \\
&\leq \big(C_4' \log(C_2' \ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log \ell_k} + \frac{3 p_{\max}}{C_1}\big) \ell_k^{3/4} = \tilde{O}(\ell_k^{3/4}).
\end{aligned}$$

On the other hand, the expected regret in the exploration phase satisfies $\mathbb{E}(R_e^{(k)}) \leq p_{\max} \ell_{k,e} \leq p_{\max} \lceil C_1 \ell_k^{3/4} \rceil \leq C_5 \ell_k^{3/4} = \tilde{O}(\ell_k^{3/4})$ where $C_5 = p_{\max} C_1 + p_{\max}$. Therefore the overall expected regret in episode k satisfies $\mathbb{E}(R^{(k)}) = \mathbb{E}(R_u^{(k)}) + \mathbb{E}(R_e^{(k)}) = \tilde{O}(\ell_k^{3/4})$.

Note that this only happens for $n_k = \ell_{k,e} = \lceil C_1 \ell_k^{3/4} \rceil \geq C_4$. For the episode such that its length ℓ_k does not satisfy $\lceil C_1 \ell_k^{3/4} \rceil \geq C_4$, we have $\mathbb{E}(R^{(k)}) \leq p_{\max} \ell_k \leq \frac{\tilde{c}_4^{1/3} p_{\max}}{C_1^{1/3}} \ell_k^{3/4} (d_0 + 1) \leq C_5' d_0 \ell_k^{3/4}$ where C_5' is a constant. Thus for any episode k, we have $\mathbb{E}(R^{(k)}) \leq \big(C_5' d_0 + C_4' \log(C_2' \ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log \ell_k} + C_6'\big) \ell_k^{2/3} = \tilde{O}(\ell_k^{3/4})$ where the constant $C_6' = \frac{3 p_{\max}}{C_1} + C_5$. Here some constants contain the dimensionality d_0. We have $C_3 = W + \frac{8(B+U+W)(d_0+1)}{c_0} = O(d_0)$, $C_6 = p_{\max} + 2C_3 = O(d_0)$ and $C_4' = C_1' C_2 + C_1 + \frac{C_6}{C_2} + p_{\max} = O(d_0)$.

Denote $K = n(T, \alpha_1)$ as the number of episodes. Note that $\ell_K = 2\ell_{K-1} \leq 2T$. Now we bound the expected regret for the entire horizon as

$$
\begin{aligned}
\mathbb{E}(R_T) &= \mathbb{E}(\sum_{k=1}^{K} R^{(k)}) = \sum_{k=1}^{K} \mathbb{E}(R^{(k)}) \\
&\leq \sum_{k=1}^{K} \Big(\big(C_5' d_0 + C_4' \log(C_2' \ell_k) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log \ell_k} + C_6' \big) \ell_k^{3/4} \Big) \\
&\leq \big(C_5' d_0 + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6' \big) \sum_{k=1}^{K} \ell_k^{3/4} \\
&= \big(C_5' d_0 + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6' \big) \alpha_1^{3/4} \frac{(2^{3/4})^K - 1}{2^{3/4} - 1} \\
&= \frac{2^{3/4}}{2^{3/4} - 1} \big(C_5' d_0 + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6' \big) \ell_K^{3/4} \\
&\leq \frac{2^{3/2}}{2^{3/4} - 1} \big(C_5' d_0 + C_4' \log(2C_2' T) + \frac{C_3' C_3 L}{\sqrt{C_1}} \sqrt{\log 2T} + C_6' \big) T^{3/4} \\
&= \tilde{O}(d_0 T^{3/4}) = \tilde{O}(T^{3/4}).
\end{aligned}
$$

□

A.5 Proof of Theorem 3

Theorem 3. *For any $\delta > 0$, no policy can achieve an $O(T^{3/5-\delta})$ regret for the dynamic pricing problem under Assumptions 1 – 5.*

Proof. We first provide a brief outline of our proof. The proof can be decomposed into two steps.

Step 1. We introduce a general procedure to generate the noise CDF F. We then prove desired properties of such F that is generated from this procedure. These F will be used to construct instances in **Step 2**.

Step 2. We construct problem instances by using a bunch of F generated from the procedure introduced in **Step 1**. We then validate the assumptions for our constructed instances. Finally, we prove that no policy can perform well on all these instances.

Now we start to prove the theorem.

Step 1. Firstly, we define the basic bump function $B(v)$ as

$$B(v) = \begin{cases} 0, & \text{for } v \in (-\infty, 0], \\ 18v^2, & \text{for } v \in (0, 1/6], \\ 1 - 18(1/3 - v)^2, & \text{for } v \in (1/6, 1/3), \\ 1, & \text{for } v \in [1/3, 2/3], \\ 1 - 18(v - 2/3)^2, & \text{for } v \in (2/3, 5/6), \\ 18(1 - v)^2, & \text{for } v \in [5/6, 1), \\ 0, & \text{for } v \in [1, +\infty). \end{cases}$$

We also define the rescaled bump function for any $-\infty < a < b < +\infty$ as $B_{[a,b]}(v) = B(\frac{v-a}{b-a})$. Now we prove that $B(v)$ is Lipschitz continuous.

Lemma 2. $|B^{'}(v)| \leq 6$ *for any* $v \in \mathbb{R}$ *and* $B(v)$ *is 6-Lipschitz.*

Proof. It is obvious that $B^{'}(v) = 0$ for $v \in (-\infty, 0) \cup (1/3, 2/3) \cup (1, +\infty)$. Note that $B(v)$ is symmetric over $v = 1/2$. We consider these following cases for v.

- For $v = 0$, we have $\lim_{v \to 0^-} \frac{B(v)-B(0)}{v-0} = 0$. We also have $\lim_{v \to 0^+} \frac{B(v)-B(0)}{v-0} = \lim_{v \to 0^+} 18v = 0$. Thus $|B^{'}(0)| = 0 \leq 6$. By symmetry, we have $|B^{'}(1)| = 0 \leq 6$.

- For $v \in (0, 1/6)$, we have $|B^{'}(v)| = |36v| \leq 6$. By symmetry, we have $|B^{'}(v)| \leq 6$ for $v \in (5/6, 1)$.

- For $v = 1/6$, let $q_1(v) = 18v^2$ and $q_2(v) = 1 - 18(1/3 - v)^2$. Then $q_1(v) = B(v)$ for $v \in (0, 1/6]$ and $q_2(v) = B(v)$ for $v \in [1/6, 1/3)$. Thus we have $\lim_{v \to \frac{1}{6}^-} \frac{B(v)-B(1/6)}{v-1/6} = \lim_{v \to \frac{1}{6}^-} \frac{q_1(v)-q_1(1/6)}{v-1/6} = q_1^{'}(1/6) = 6$ and $\lim_{v \to \frac{1}{6}^+} \frac{B(v)-B(1/6)}{v-1/6} = \lim_{v \to \frac{1}{6}^+} \frac{q_2(v)-q_2(1/6)}{v-1/6} = q_2^{'}(1/6) = 6$. Thus $|B^{'}(1/6)| = 6 \leq 6$. By symmetry, we have $|B^{'}(5/6)| = 6 \leq 6$.

- For $v \in (1/6, 1/3)$, we have $|B'(v)| = |36(1/3 - v)| \leq 6$. By symmetry, we have $|B'(v)| \leq 6$ for $v \in (2/3, 5/6)$.

- For $v = 1/3$, let $q_1(v) = 1 - 18(1/3 - v)^2$ and $q_2(v) = 1$. Then $q_1(v) = B(v)$ for $v \in (1/6, 1/3]$ and $q_2(v) = B(v)$ for $v \in [1/3, 1/2)$. Thus we have $\lim_{v \to \frac{1}{3}^-} \frac{B(v)-B(1/3)}{v-1/3} = \lim_{v \to \frac{1}{3}^-} \frac{q_1(v)-q_1(1/3)}{v-1/3} = q_1'(1/3) = 0$ and $\lim_{v \to \frac{1}{3}^+} \frac{B(v)-B(1/3)}{v-1/3} = \lim_{v \to \frac{1}{3}^+} \frac{q_2(v)-q_2(1/3)}{v-1/3} = q_2'(1/3) = 0$. Thus $|B'(1/3)| = 0 \leq 6$. By symmetry, we have $|B'(2/3)| = 0 \leq 6$.

Thus we have $|B'(v)| \leq 6$ for $v \in \mathbb{R}$ and $B(v)$ is 6-Lipschitz by Lagrange's Mean Value Theorem. □

Then we prove another critical property of the bump function $B(v)$.

Lemma 3. *For $v \in [0, 1/3)$, $B(1/3) - B(v) \leq 18(1/3 - v)^2$.*

Proof. For $v \in (1/6, 1/3)$, we have $B(1/3) - B(v) = 18(1/3 - v)^2 \leq 18(1/3 - v)^2$. For $v \in [0, 1/6]$, we have $B(1/3) - B(v) = 1 - 18v^2 \leq 18(1/3 - v)^2$ since it is equivalent to $36v^2 - 12v + 1 = (6v - 1)^2 \geq 0$. □

For rescaled bump functions, Lemma 3 translates to $B_{[a,b]}(a + \frac{b-a}{3}) - B_{[a,b]}(v) = B(1/3) - B(\frac{v-a}{b-a}) \leq 18(1/3 - \frac{v-a}{b-a})^2 = \frac{18}{(b-a)^2}((a + \frac{b-a}{3}) - v)^2$ for $v \in [a, a + \frac{b-a}{3})$.

We then define a bunch of interval series $[0, 1] = [a_0, b_0] \supset [a_1, b_1] \supset \cdots \supset [a_k, b_k] \supset \ldots$, where the interval lengths satisfy $w_k = b_k - a_k = 3^{-k!}$ for $k \geq 1$ and $w_0 = b_0 - a_0 = 1$. To select $[a_k, b_k]$, we first divide the range $[a_{k-1} + \frac{w_{k-1}}{3}, b_{k-1} - \frac{w_{k-1}}{3}]$ into $Q_k = \frac{w_{k-1}}{3w_k}$ sub-intervals of the same length w_k and then pick one of these sub-intervals as $[a_k, b_k]$. Note that there are infinite such series of intervals. For each of these interval series, we are able to define the function

$$f(v) = C_f \sum_{k=0}^{\infty} w_k^2 B_{[a_k, b_k]}(v)$$

where C_f is a constant remained to be determined later. Denote $f_K(v) = C_f \sum_{k=0}^{K} w_k^2 B_{[a_k,b_k]}(v)$. We then show a few critical properties of $f(v)$.

Lemma 4. *1. $f(v) \in [0, \frac{9}{8}C_f]$ for any $v \in \mathbb{R}$.*

2. $f(v)$ is Lipschitz continuous.

3. There exists a unique $v^* \in [0,1]$ such that $f(v^*) = \max_{v\in[0,1]} f(v)$. Specifically, $\{v^*\} = \cap_{k=1}^{\infty}[a_k, b_k]$.

4. $f(v)$ is unimodal around the unique maximizer v^*.

5. For any $v \in [0,1]$, $f(v^*) - f(v) \le 18C_f(v^* - v)^2$.

Proof. 1. For any $v \in \mathbb{R}$, $B_{[a_k,b_k]}(v) \ge 0$. Thus $f_K(v)$ is non-decreasing and $f(v) = \lim_{K\to\infty} f_K(v) \ge 0$. On the other hand, $B_{[a_k,b_k]}(v) \le 1$. Thus $f_K(v) \le \sum_{k=0}^{K} C_f w_k^2$ and $f(v) = \lim_{K\to\infty} f_K(v) \le \lim_{K\to\infty} \sum_{k=0}^{K} C_f w_k^2 = \sum_{k=0}^{\infty} C_f w_k^2 \le C_f \sum_{k=0}^{\infty} 3^{-2k} \le \frac{9}{8}C_f < \infty$.

2. $f_K^{'}(v) = C_f \sum_{k=0}^{K} w_k^2 B_{[a_k,b_k]}^{'}(v) = C_f \sum_{k=0}^{K} w_k^2 \frac{1}{b_k-a_k} B^{'}(\frac{v-a_k}{b_k-a_k}) = C_f \sum_{k=0}^{K} w_k B^{'}(\frac{v-a_k}{b_k-a_k})$. Denote $\dot{f}(v) = \lim_{K\to\infty} f_K^{'}(v)$. Then we have $|\dot{f}(v)| = |\lim_{K\to\infty} f_K^{'}(v)| \le C_f \sum_{k=0}^{K} w_k |B^{'}(\frac{v-a_k}{b_k-a_k})| \le 9C_f$. Namely, $\dot{f}(v)$ exists and is finite-valued. Moreover, we have for any $v \in \mathbb{R}$,

$$|\dot{f}(v) - f_K^{'}(v)| = |C_f \sum_{k=K+1}^{\infty} w_k B^{'}(\frac{v-a_k}{b_k-a_k})| \le C_f \sum_{k=K+1}^{\infty} w_k |B^{'}(\frac{v-a_k}{b_k-a_k})| \le 3^{-(K+1)!}(9C_f)$$

since $\sum_{k=K+1}^{\infty} w_k = \sum_{k=K+1}^{\infty} 3^{-k!} \le w_{K+1} \sum_{k=0}^{\infty} 3^{-k} \le \frac{3}{2} 3^{-(K+1)!}$. Since $3^{-(K+1)!}(9C_f) \to 0$ with $K \to \infty$, we have $f_K^{'}(v)$ converges to $\dot{f}(v)$ uniformly in $\mathbb{R}$. We also have each component $w_k B^{'}(\frac{v-a_k}{b_k-a_k})$ in $\dot{f}(v)$ is continuous. Thus by the property of the function series, we have $f^{'}(v)$ exists and

$$f^{'}(v) = \lim_{K\to\infty} f_K^{'}(v) = \dot{f}(v).$$

Thus $|f'(v)| = |\dot{f}(v)| \le 9C_f$ for any $v \in \mathbb{R}$ and $f(v)$ is $9C_f$-Lipschitz by the Lagrange's Mean Value Theorem.

3. Since $f(v)$ is Lipschitz continuous, there exists the maximum for $f(v)$ in $[0,1]$. Let v^* be some maximizer of $f(v)$ in $[0,1]$. Then it is obvious that $v^* \in [a_k, b_k]$ for any $k \in \mathbb{N}$. Since $b_k - a_k = w_k = 3^{-k!} \to 0$, a_k is increasing with $a_k \le 1$ for any $k \ge 1$ and b_k is decreasing with $b_k \ge 0$ for any $k \ge 1$, we have $\lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k = v^*$ and $\{v^*\} = \cap_{k=1}^{\infty}[a_k, b_k]$.

4. For any $k \ge 0$, $B_{[a_k,b_k]}(v)$ is non-decreasing in $[0, v^*]$ and non-increasing in $[v^*, 1]$ since $v^* \in [a_{k+1}, b_{k+1}] \subset [a_k + \frac{w_k}{3}, b_k - \frac{w_k}{3}]$. Thus $f(v)$, as the sum of these functions, is non-decreasing in $[0, v^*]$ and non-increasing in $[v^*, 1]$.

5. For any $k \ge 0$, we have $B_{[a_k,b_k]}(v^*) = 1$ since $v^* \in [a_{k+1}, b_{k+1}] \subset [a_k + \frac{w_k}{3}, b_k - \frac{w_k}{3}]$. Thus we are able to calculate $f(v^*) = C_f \sum_{k=0}^{\infty} w_k^2 B_{[a_k,b_k]}(v^*) = C_f \sum_{k=0}^{\infty} w_k^2$. Note that it is enough to consider $v \in [0, v^*]$ since the reasoning is exactly the same for $v \in [v^*, 1]$. For $v = v^*$, we have any $C > 0$, $f^*(v) - f(v) = 0 \le C(v^* - v)^2 = 0$. Now we consider any fixed $v \in [0, v^*)$. Since $a_0 = 0$, a_k is increasing and $a_k \to v^*$, there exists an $i \ge 0$ such that $v \in [a_i, a_{i+1})$. Now there are two further cases.

   (a) $v \in [a_i + \frac{w_i}{3}, a_{i+1})$. In this case, we have $f(v) = C_f \sum_{k=0}^{\infty} w_k^2 B_{[a_k,b_k]}(v) = C_f \sum_{k=0}^{i} w_k^2$. Thus $f^*(v) - f(v) = C_f \sum_{k=i+1}^{\infty} w_k^2 \le C_f w_{i+1}^2 (\sum_{k=0}^{\infty} 3^{-2k}) \le \frac{9}{8} C_f w_{i+1}^2$. On the other hand, we have $v^* \ge a_{i+2} \ge a_{i+1} + \frac{w_{i+1}}{3} \ge v + \frac{w_{i+1}}{3}$. Thus $v^* - v \ge \frac{w_{i+1}}{3}$. Therefore, we have

   $$f(v^*) - f(v) \le \frac{9}{8} C_f w_{i+1}^2 \le \frac{81}{8} C_f (v^* - v)^2.$$

   (b) $v \in [a_i, a_i + \frac{w_i}{3})$. In this case, we have $f(v) = C_f \sum_{k=0}^{\infty} w_k^2 B_{[a_k,b_k]}(v) = C_f \sum_{k=0}^{i-1} w_k^2 + C_f w_i^2 B_{[a_i,b_i]}(v)$. On the other hand, let $v_i = a_i + \frac{w_i}{3}$ and we have $f(v_i) = C_f \sum_{k=0}^{\infty} w_k^2 B_{[a_k,b_k]}(v) = C_f \sum_{k=0}^{i} w_k^2$. By Lemma 3, we have $f(v_i) - f(v) = C_f w_i^2 (1 - B_{[a_i,b_i]}(v)) = C_f w_i^2 (B_{[a_i,b_i]}(v_i) - B_{[a_i,b_i]}(v)) \leq C_f w_i^2 \frac{18}{(b_i - a_i)^2}(v_i - v)^2 = 18 C_f (v_i - v)^2$. In addition, as $v_i = a_i + \frac{w_i}{3} \in [a_i + \frac{w_i}{3}, a_{i+1})$, the just-derived result in (a) yields $f(v^*) - f(v) \leq \frac{81}{8} C_f (v^* - v)^2 \leq 18 C_f (v^* - v)^2$. Thus we have

$$\begin{aligned} f(v^*) - f(v) &= (f(v^*) - f(v_i)) + (f(v_i) - f(v)) \\ &\leq 18C_f (v^* - v_i)^2 + 18 C_f (v_i - v)^2 \\ &\leq 18 C_f \big((v^* - v_i) + (v_i - v)\big)^2 = 18 C_f (v^* - v)^2. \end{aligned}$$

Thus we have $f(v^*) - f(v) \leq 18C_f(v^* - v)^2$ for $v \in [0, v^*)$. The derivation is exactly the same for $v \in (v^*, 1]$ and thus we conclude that $f(v^*) - f(v) \leq 18C_f(v^* - v)^2$ for $v \in [0, 1]$.

□

From $f(v)$, we define another function $g(v) = 1 - \frac{1}{1+f(v)}$. Then we have $|g^{'}(v)| = |\frac{f^{'}(v)}{(f(v)+1)^2}| \leq 6C_f$. Thus $g(v)$ is $6C_f$-Lipschitz by Lagrange's Mean Value Theorem. Moreover, we have $|g(v_1) - g(v_2)| = |\frac{1}{1+f(v_2)} - \frac{1}{1+f(v_1)}| \leq |f(v_1) - f(v_2)|$. In addition, $g(v)$ is unimodal with the same unique maximizer $v_g^* = v_f^*$ of $f(v)$.

Let $C_f \in (0, 1/6)$ and $b = \frac{1+6C_f}{2} \in (0, 1)$ be a constant. We further define the function $F(v)$ as

$$F(v) = \begin{cases} 0, & \text{for } v \in (-\infty, b], \\ 1 - \frac{b}{v} - \frac{1-b}{v} g(\frac{v-b}{1-b}), & \text{for } v \in (b, 1), \\ 2 - \frac{1+b}{v}, & \text{for } v \in [1, 1+b), \\ 1, & \text{for } v \in [1+b, +\infty). \end{cases}$$

Then we have the following properties for $F(v)$.

Lemma 5. *1. $F(v)$ is a Cumulative Distribution Function (CDF) of some $\mathbb{R}$-valued random variable.*

2. *$F(v)$ is Lipschitz continuous on $\mathbb{R}$.*

3. *There is a unique minimizer $v_r^* \geq 0$ for the function $r(v) = v(1 - F(v))$ on $\mathbb{R}^+ \cup \{0\}$.*

4. *For any $v \geq 0$, we have $r(v_r^*) - r(v) \leq \frac{36}{(1-6C_f)^2}(v_r^* - v)^2$.*

Proof. 1. It is easy to see that $\lim_{v\to-\infty} F(v) = 0$ and $\lim_{v\to+\infty} F(v) = 1$. It is also easy to see that $F(v)$ is continuous on $(-\infty, b), (1, 1+b)$ and $(1+b, +\infty)$. Since $g(v)$ is continuous on $(0,1)$, $g(\frac{v-b}{1-b})$ is continous on $v \in (b,1)$. Thus $F(v)$ is continous on $v \in (b,1)$. Moreover, we have $\lim_{v\to b^+} F(v) = \lim_{v\to b^+}(1 - \frac{b}{v} - \frac{1-b}{v}g(\frac{v-b}{1-b})) = 1 - 1 - \frac{1-b}{b}g(0) = 0 = \lim_{v\to b^-} F(v) = F(b)$ and thus $F(v)$ is continuous on $v = b$. Also, we have $\lim_{v\to 1^-} F(v) = \lim_{v\to 1^-}(1 - \frac{b}{v} - \frac{1-b}{v}g(\frac{v-b}{1-b})) = 1 - b - (1-b)g(0) = 1 - b = \lim_{v\to 1^+} F(v) = F(1)$ and thus $F(v)$ is continuous on $v = 1$. In addition, we have $\lim_{v\to(1+b)^-} F(v) = 1 = \lim_{v\to(1+b)^+} F(v) = F(1+b)$ and thus $F(v)$ is continuous on $v = 1+b$. Thus $F(v)$ is continous on $\mathbb{R}$.

On the other hand, it is easy to see that $F(v)$ is non-decreasing on $(-\infty, b] \cup [1, +\infty)$. For any $v \in (b,1)$, we have $F'(v) = \frac{b - vg'(\frac{v-b}{1-b}) + (1-b)g(\frac{v-b}{1-b})}{v^2}$. Then we have $b - vg'(\frac{v-b}{1-b}) \geq b - v|g'(\frac{v-b}{1-b})| \geq b - 6C_f = \frac{1+6C_f}{2} - 6C_f = \frac{1-6C_f}{2} > 0$. Thus we have $F(v)$ is non-decreasing on $(b,1)$ and $F(b) \leq F(v) \leq F(1)$ for $v \in (b,1)$. Thus $F(v)$ is non-decreasing on $\mathbb{R}$. Therefore, $F(v)$ is a CDF for some $\mathbb{R}$-valued random variable.

2. It is easy to see that $F(v)$ is 1-Lipschitz on $(-\infty, b]$ or $[1+b, +\infty)$. Since $|F'(v)| = |\frac{1+b}{v^2}| \leq 2$ for $v \in (1, 1+b)$ and $F(v)$ is continuous on $[1, 1+b]$, we have $F(v)$ is 2-Lipschitz on $[1, 1+b]$ by Lagrange's Mean Value Theorem. Since $|F'(v)| = |\frac{b - vg'(\frac{v-b}{1-b}) + (1-b)g(\frac{v-b}{1-b})}{v^2}| \leq \frac{b + 6C_f + (1-b)}{b^2} = 2 + 12C_f \leq 14$ for $v \in (b,1)$ and $F(v)$ is continuous on $[b,1]$, we have $F(v)$ is 14-Lipschitz on $[b,1]$ by Lagrange's Mean Value Theorem. Therefore, by triangular inequalities, we have $F(v)$ is 14-Lipschitz on $\mathbb{R}$.

3. For $v \geq 0$, simple calculation yields

$$r(v) = v(1 - F(v)) = \begin{cases} v, & \text{for } v \in [0, b), \\ b + (1-b)g(\frac{v-b}{1-b}), & \text{for } v \in [b, 1], \\ 1 + b - v, & \text{for } v \in (1, 1+b], \\ 0, & \text{for } v \in (1+b, +\infty). \end{cases}$$

Since $g(v) \geq 0$, we have $r(v) \geq b > r(v')$ for any $v \in [b, 1]$ and $v' \in (\mathbb{R}^+ \cup \{0\}) \setminus [b, 1]$. Since $g(v)$ has the same unique maximizer $v_g^* = v_f^*$ for $f(v)$, we obtain that $v_r^* = b + (1-b)v_g^* \in [b, 1]$ is the unique maximizer for $r(v)$ on $\mathbb{R}^+ \cup \{0\}$.

4. For any $v \in [b, 1]$, we have

$$\begin{aligned} r(v_r^*) - r(v) &= (1-b)\big(g(\frac{v_r^* - b}{1-b}) - g(\frac{v-b}{1-b})\big) = (1-b)|g(v_g^*) - g(\frac{v-b}{1-b})| \\ &\leq (1-b)|f(v_f^*) - f(\frac{v-b}{1-b})| \leq (1-b)18C_f(v_f^* - \frac{v-b}{1-b})^2 \\ &= \frac{18C_f}{1-b}((1-b)v_f^* + b - v)^2 = \frac{36C_f}{1-6C_f}(v_r^* - v)^2. \end{aligned}$$

For any $v \in (\mathbb{R}^+ \cup \{0\}) \setminus [1, b]$, we have $r(v_r^*) - r(v) \leq r(v_r^*) \leq b + (1-b) = 1$ since $g(v) \leq 1$ for any $v \in [0, 1]$. On the other hand, since $v_g^* = v_f^* \in [1/3, 2/3]$, we have $v_r^* = b + (1-b)v_g^* \in [b + \frac{1-b}{3}, 1 - \frac{1-b}{3}]$ and thus $|v_r^* - v| \geq \frac{1-b}{3}$ for any $v \in (\mathbb{R}^+ \cup \{0\}) \setminus [1, b]$. Thus we have $r(v_r^*) - r(v) \leq 1 \leq \frac{9}{(1-b)^2}(v_r^* - v)^2 = \frac{36}{(1-6C_f)^2}(v_r^* - v)^2$. Since $\frac{36}{(1-6C_f)^2} \geq \frac{36C_f}{1-6C_f}$, we have for any $v \geq 0$, $r(v_r^*) - r(v) \leq \frac{36}{(1-6C_f)^2}(v_r^* - v)^2$.

$\square$

Step 2. Now we specify the problem setting and validate the assumptions. Let $\theta_0 = (0, 0, \ldots, 0)^\top$ be the d_0-dimensional zero vector. Let x_t be i.i.d. samples from a distribution $\mathbb{P}_x$ such that each component of x_t are independent identically distributed as Unif$(-1, 1)$. Let the noise CDF be any $F(v)$ generated from the procedure introduced in **Step 1**. Then the support of x_t is $\mathcal{X} = (-1, 1)^{d_0}$. Then $||x_t||_\infty \leq 1, ||\theta_0||_1 \leq 1$ and thus Assumption 1 is satisfied with the constant $W = 1$. Also, $\Sigma = \mathbb{E}\big((1, x_t^\top)(1, x_t^\top)\big)$ is a diagonal matrix $\mathrm{Diag}(1, 1/3, \ldots, 1/3)$ and thus Assumption 2 is satisfied with the constant $c_0 = 1/4$. Let $F(v)$ be the CDF for the market noise z_t. Then by the construction of $F(v)$, the noise is bounded in $[b, 1+b] \subseteq (0, 2)$ since $b \in (0, 1)$. In addition, any realized valuations v_t satisfies $v_t = x_t^\top \theta_0 + z_t = z_t \in [b, 1+b] \subseteq [0, 2]$ and thus Assumption 3 is satisfied with $B = 2$. By Lemma 5, $F(v)$ is 14-Lipschitz and thus Assumption 4 is satisfied with $L = 14$. Now we prove that Assumption 5 is also satisfied. For any $x \in \mathcal{X}$, we have $q = x^\top \theta_0 = 0$. Thus $f_q(p) = p(1 - F(p - x^\top \theta_0)) = p(1 - F(p)) = r(p)$. Namely, $r(v)$ is just the revenue function. Thus by Lemma 5, the optimal price $p^*(x) = v_r^* \in [b, 1]$. In addition, we have for any $p \in [0, p_{\max}]$ and $x \in \mathcal{X}$,

$$f_q(p^*(x)) - f_q(p) = r(v_r^*) - r(p) \le \frac{36}{(1 - 6C_f)^2}(v_r^* - p)^2 = \frac{36}{(1 - 6C_f)^2}(p^*(x) - p)^2.$$

Therefore, Assumption 5 is satisfied. Thus under any noise CDF $F(v)$ that is constructed through our introduced procedure, all Assumptions 1 – 5 are satisfied.

Now we construct instances to prove the main theorem. Note that each interval series $\{[a_k, b_k]\}_{k \ge 0}$ that is introduced in **Step 1** corresponds to a triplet $(f(v), g(v), F(v))$. Thus each interval series corresponds to a noise CDF $F(v)$ and thus a problem instance constructed as above. We will construct an infinite sequence of well-formulated groups of such instances. For each of these instance group, we will prove that no policy can perform well on all the instances in this group. Specifically, let $n_k = \lceil \frac{w_{k-1}}{w_k^5} \rceil$ for $k \ge 1$. For each k, we first define $f_0(v) = C_f \sum_{j=0}^{k-1} w_j^2 B_{[a_j, b_j]}(v)$, which corresponds to the finite series of intervals $\{[a_j, b_j]\}_{0 \le j \le k-1}$ and no choice of $[a_k, b_k]$. We further define $g_0(v)$ and $F_0(v)$ based on $f_0(v)$ through the previously introduced procedure in **Step 1**. Then we consider the possible $Q_k = \frac{w_{k-1}}{3w_k}$ choices of $[a_k, b_k]$. Denote these intervals as $I_j, j = 1, 2, \ldots, Q_k$ and the corresponding triplet $(f(v), g(v), F(v))$ for $[a_k, b_k] = I_j$ as $(f_j(v), g_j(v), F_j(v))$. These noise CDF $\{F_j(v)\}_{j \in [Q_k]}$ form the k-th group of instances.

Now consider any policy π. For simplicity, we denote n_k as n. The policy π would interact with the noise distribution $F_j(v)$ and generate the price and response sequence $u_n = (p_1, y_1, p_2, y_2, \ldots, p_n, y_n)$. Denote the distribution of u_n under $F_j(v)$ as $\mathbb{P}_j$ for $j \in \{0, 1, \ldots, Q_k\}$. We may assume $p_t \in [b, 1]$ as any price $p' \in [0, p_{\max}] \setminus [b, 1]$ is suboptimal and dominated by any price $p \in [b, 1]$ by the property of the revenue function $r(v)$.

Now we derive the KL-divergence $\mathrm{KL}(\mathbb{P}_0||\mathbb{P}_j)$ for any $j \in [Q_k]$. For $j \in [Q_k] \cup \{0\}$, denote $P_j(y|p)$ as the Bernoulli distribution of the binary response y given the price p under the noise distribution F_j. In particular, $P_j(y|p) = \mathrm{Ber}(1 - F_j(p))$. Then we have $\mathbb{P}_0(u_n) = \Pi_{t=1}^n\big(\pi(p_t|p_1, y_1, \ldots, p_{t-1}, y_{t-1})P_0(y_t|p_t)\big)$. Similarly, we have $\mathbb{P}_j(u_n) = \Pi_{t=1}^n\big(\pi(p_t|p_1, y_1, \ldots, p_{t-1}, y_{t-1})P_j(y_t|p_t)\big)$. Thus we obtain

$$
\begin{aligned}
\mathrm{KL}(\mathbb{P}_0||\mathbb{P}_j) &= \mathbb{E}_{\mathbb{P}_0}\big(\log\frac{\mathbb{P}_0(u_n)}{\mathbb{P}_j(u_n)}\big) = \mathbb{E}_{\mathbb{P}_0}\Big(\log\frac{\Pi_{t=1}^n\big(\pi(p_t|p_1, y_1, \ldots, p_{t-1}, y_{t-1})P_0(y_t|p_t)\big)}{\Pi_{t=1}^n\big(\pi(p_t|p_1, y_1, \ldots, p_{t-1}, y_{t-1})P_j(y_t|p_t)\big)}\Big) \\
&= \mathbb{E}_{\mathbb{P}_0}\big(\log\frac{\Pi_{t=1}^n P_0(y_t|p_t)}{\Pi_{t=1}^n P_j(y_t|p_t)}\big) = \mathbb{E}_{\mathbb{P}_0}\big(\sum_{t=1}^n \log\frac{P_0(y_t|p_t)}{P_j(y_t|p_t)}\big) \\
&= \sum_{t=1}^n \mathbb{E}_{\mathbb{P}_0}\Big(\mathbb{E}_{\mathbb{P}_0}\big(\log\frac{P_0(y_t|p_t)}{P_j(y_t|p_t)}|p_t\big)\Big) = \sum_{t=1}^n \mathbb{E}_{\mathbb{P}_0}\Big(\mathrm{KL}\big(P_0(\cdot|p_t)||P_j(\cdot|p_t)\big)\Big) \\
&= \sum_{t=1}^n \mathbb{E}_{\mathbb{P}_0}\Big(\mathrm{KL}\big(\mathrm{Ber}(1 - F_0(p_t))||\mathrm{Ber}(1 - F_j(p_t))\big)\Big) \\
&= \sum_{t=1}^n \mathbb{E}_{\mathbb{P}_0}\Big(1_{\frac{p_t - b}{1-b} \notin I_j}\mathrm{KL}\big(\mathrm{Ber}(1 - F_0(p_t))||\mathrm{Ber}(1 - F_j(p_t))\big) \\
&+ 1_{\frac{p_t - b}{1-b} \in I_j}\mathrm{KL}\big(\mathrm{Ber}(1 - F_0(p_t))||\mathrm{Ber}(1 - F_j(p_t))\big)\Big).
\end{aligned} \tag{10}
$$

Now we present a lemma on the KL-divergence of two Bernoulli distributions.

Lemma 6. *For Bernoulli distributions Ber(p) and Ber$(p + \epsilon)$ with $1/2 \le p \le p + \epsilon \le 1/2 + C$, we have*

$$KL(Ber(p)||Ber(p + \epsilon)) \le \frac{4}{1 - 4C^2}\epsilon^2.$$

Proof. For $v > 0$, we have $\frac{v}{1+v} \le \ln(1 + v) \le v$ since $(\frac{v}{1+v})' = \frac{1}{(1+v)^2} \le \big(\ln(1 + v)\big)' =$

$\frac{1}{1+v} \leq (v)' = 1$ and $(\frac{v}{1+v})\big|_{v=0} = \big(\ln(1+v)\big)\big|_{v=0} = (v)\big|_{v=0} = 0$. Direct calculation yields

$$\begin{aligned}
\mathrm{KL}(\mathrm{Ber}(p)||\mathrm{Ber}(p+\epsilon)) &= p\log(\frac{p}{p+\epsilon}) + (1-p)\log(\frac{1-p}{1-p-\epsilon}) \\
&= p\big(-\ln(1+\frac{\epsilon}{p})\big) + (1-p)\ln(1+\frac{\epsilon}{1-p-\epsilon}) \\
&\leq p(-\frac{\frac{\epsilon}{p}}{1+\frac{\epsilon}{p}}) + (1-p)\frac{\epsilon}{1-p-\epsilon} \\
&= \frac{-p\epsilon(1-p-\epsilon) + (1-p)\epsilon(p+\epsilon)}{(p+\epsilon)(1-p-\epsilon)} \\
&= \frac{\epsilon^2}{(p+\epsilon)(1-p-\epsilon)} \\
&\leq \frac{\epsilon^2}{(\frac{1}{2}+C)(\frac{1}{2}-C)} = \frac{4}{1-4C^2}\epsilon^2.
\end{aligned}$$

□

For simplicity, denote $q_t = \frac{p_t-b}{1-b}$. Then for p_t such that $\frac{p_t-b}{1-b} \notin I_j$, we have $F_0(p_t) = F_j(p_t)$ by their constructions and thus $\mathrm{KL}\big(\mathrm{Ber}(1-F_0(p_t))||\mathrm{Ber}(1-F_j(p_t))\big) = 0$.

Then we focus on the p_t such that $q_t = \frac{p_t-b}{1-b} \in I_j$, we have

$$\begin{aligned}
(1-F_j(p_t)) - (1-F_0(p_t)) &= \frac{b+(1-b)g_j(\frac{p_t-b}{1-b})}{p_t} - \frac{b+(1-b)g_0(\frac{p_t-b}{1-b})}{p_t} \\
&= \frac{1-b}{p_t}(g_j(q_t) - g_0(q_t)) \\
&= \frac{1-b}{p_t}(\frac{1}{1+f_0(q_t)} - \frac{1}{1+f_j(q_t)}) \\
&= \frac{1-b}{p_t}(\frac{C_f\sum_{i=k}^{\infty} w_i^2 B_{[a_i,b_i]}(q_t)}{(1+f_0(q_t))(1+f_j(q_t))}) \\
&\leq \frac{1-b}{b}\frac{C_f w_k^2(\sum_{i=0}^{\infty} 3^{-2i})}{1\times 1} \\
&\leq \frac{1-1/2}{1/2}\frac{9}{8}C_f w_k^2 = \frac{9}{8}C_f w_k^2.
\end{aligned}$$

Note that from the above derivation, we also have

$$(1 - F_j(p_t)) - (1 - F_0(p_t)) = \frac{1-b}{p_t}\left(\frac{C_f \sum_{i=k}^{\infty} w_i^2 B_{[a_i,b_i]}(q_t)}{(1 + f_0(q_t))(1 + f_j(q_t))}\right) \geq 0.$$

Moreover, we have $1-F_0(p_t) = \frac{b+(1-b)g_0(q_t)}{p_t} \geq b = \frac{1+6C_f}{2} \geq \frac{1}{2}$ since $g_0(q_t) = 1-\frac{1}{1+f_0(q_t)} \geq 0$. Note that by Lemma 5, $F_j(v)$ is a CDF and thus is non-decreasing. Since $k \geq 1$, we have $q_t = \frac{p_t - b}{1-b} \in I_j \subset [a_0 + 1/3, b_0 - 1/3] = [1/3, 2/3]$. Thus $p_t \geq b + \frac{1-b}{3}$. Therefore, we obtain

$$\begin{aligned} 1 - F_j(p_t) &\leq 1 - F_j(b + \frac{1-b}{3}) = \frac{b + (1-b)g_j(1/3)}{b + \frac{1-b}{3}} \\ &= 3\frac{b + (1-b)\frac{f_j(1/3)}{1+f_j(1/3)}}{2b+1} = 3\frac{b + (1-b)\frac{C_f}{1+C_f}}{2b+1} = \frac{340}{427} < \frac{5}{6} \end{aligned}$$

with the choice of $C_f = \frac{1}{60}$ and $b = \frac{1+6C_f}{2} = \frac{11}{20}$. Thus we have

$$\frac{1}{2} \leq 1 - F_0(p_t) \leq 1 - F_j(p_t) \leq \frac{1}{2} + \frac{1}{3}.$$

Thus by Lemma 6, we obtain

$$\begin{aligned} \mathrm{Ber}(1 - F_0(p_t))||\mathrm{Ber}(1 - F_j(p_t)) &\leq \frac{4}{1 - 4 \cdot (\frac{1}{3})^2}\big((1 - F_j(p_t)) - (1 - F_0(p_t))\big)^2 \\ &\leq \frac{36}{5} \cdot (\frac{9}{8})^2 \cdot (\frac{1}{60})^2 \cdot w_k^4 \leq \frac{1}{300} w_k^4. \end{aligned}$$

Then we can further bound $\mathrm{KL}(\mathbb{P}_0||\mathbb{P}_j)$ through the Equation (10) as

$$\begin{aligned}\mathrm{KL}(\mathbb{P}_0||\mathbb{P}_j) = &\sum_{t=1}^{n} \mathbb{E}_{\mathbb{P}_0}\Big(1_{\frac{p_t-b}{1-b}\notin I_j}\mathrm{KL}\big(\mathrm{Ber}(1-F_0(p_t))||\mathrm{Ber}(1-F_j(p_t))\big)\\ &+ 1_{\frac{p_t-b}{1-b}\in I_j}\mathrm{KL}\big(\mathrm{Ber}(1-F_0(p_t))||\mathrm{Ber}(1-F_j(p_t))\big)\Big)\\ \le &\sum_{t=1}^{n} \mathbb{E}_{\mathbb{P}_0}(1_{\frac{p_t-b}{1-b}\notin I_j}\cdot 0 + 1_{\frac{p_t-b}{1-b}\in I_j}\cdot \frac{1}{300}w_k^4)\\ = &\frac{1}{300}w_k^4 \sum_{t=1}^{n} \mathbb{P}_0(\frac{p_t-b}{1-b}\in I_j) = \frac{1}{300}w_k^4 \sum_{t=1}^{n}\mathbb{P}_0(q_t\in I_j).\end{aligned} \tag{11}$$

Now, consider any function h on the price and response sequence $u_n = (p_1, y_1, \ldots, p_n, y_n)$ that has a bounded value range $[0, M]$. Define the reference measure $\mathbb{Q}_j = \frac{1}{2}(\mathbb{P}_j + \mathbb{P}_0)$. Then both $\mathbb{P}_j$ and $\mathbb{P}_0$ are absolute continuous with respect to $\mathbb{Q}_j$. Thus the Radon-Nikodym derivatives $\frac{d\mathbb{P}_j}{d\mathbb{Q}_j} = m_j$ and $\frac{d\mathbb{P}_0}{d\mathbb{Q}_j} = m_{j,0}$ exist. Denote the set $O_j = \{u : m_j(u) - m_{j,0}(u) \ge 0\}$. Then we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_j}(h(u_n)) - \mathbb{E}_{\mathbb{P}_0}(h(u_n)) &= \int h d\mathbb{P}_j - \int h d\mathbb{P}_0 = \int h m_j d\mathbb{Q}_j - \int h m_{j,0} d\mathbb{Q}_j\\ &= \int h(m_j - m_{j,0})d\mathbb{Q}_j \le \int_{m_j - m_{j,0}\ge 0} h(m_j - m_{j,0})d\mathbb{Q}_j\\ &\le \int_{O_j} M(m_j - m_{j,0})d\mathbb{Q}_j = M\big(\int_{O_j} d\mathbb{P}_j - \int_{O_j} d\mathbb{P}_0\big)\\ &= M(\mathbb{P}_j(O_j) - \mathbb{P}_0(O_j)) \le M\sup_{O}|\mathbb{P}_j(O) - \mathbb{P}_0(O)|\\ &= M||\mathbb{P}_j - \mathbb{P}_0||_1 \le M\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_j||\mathbb{P}_0)}.\end{aligned} \tag{12}$$

We use the Pinsker's inequality for the last step, which demonstrates the relationship between the total variantion distance between two probability measures and their KL-divergence.

Now, for each $j \in [Q_k]$, denote $N_j = |\{t|q_t = \frac{p_t-b}{1-b} \in I_j, t \in [n_k]\}|$. Then N_j is a function of the price and response sequence u_n and is bounded in the range $[0, n_k]$.

Thus by combining the Equation (11) – (12), we obtain

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_j}(N_j) - \mathbb{E}_{\mathbb{P}_0}(N_j) &\le n_k\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_j||\mathbb{P}_0)} \le n_k\sqrt{\frac{1}{2}\frac{1}{300}w_k^4\sum_{t=1}^{n}\mathbb{P}_0(q_t\in I_j)} \\ &\le \frac{1}{20}n_k w_k^2\sqrt{\sum_{t=1}^{n}\mathbb{P}_0(q_t\in I_j)} = \frac{1}{20}n_k w_k^2\sqrt{\mathbb{E}_{\mathbb{P}_0}(N_j)}.\end{aligned}$$

Thus $\mathbb{E}_{\mathbb{P}_j}(N_j) \le \mathbb{E}_{\mathbb{P}_0}(N_j) + \frac{1}{20}n_k w_k^2\sqrt{\mathbb{E}_{\mathbb{P}_0}(N_j)}$. For $k\ge 3$, we sum over $j\in[Q_k]$ and take the average to obtain

$$\begin{aligned}\frac{1}{Q_k}\sum_{j=1}^{Q_k}\mathbb{E}_{\mathbb{P}_j}(N_j) &\le \frac{1}{Q_k}\sum_{j=1}^{Q_k}\mathbb{E}_{\mathbb{P}_0}(N_j) + \frac{1}{Q_k}\frac{1}{20}n_k w_k^2\sum_{j=1}^{Q_k}\sqrt{\mathbb{E}_{\mathbb{P}_0}(N_j)} \\ &= \frac{n_k}{Q_k} + \frac{1}{20}\frac{1}{Q_k}n_k w_k^2\sum_{j=1}^{Q_k}\sqrt{\mathbb{E}_{\mathbb{P}_0}(N_j)} \\ \text{(Cauchy-Schwarz)} &\le \frac{n_k}{Q_k} + \frac{1}{20}\frac{n_k}{Q_k}w_k^2\sqrt{Q_k\sum_{j=1}^{Q_k}\mathbb{E}_{\mathbb{P}_0}(N_j)} \\ &= \frac{n_k}{Q_k} + \frac{1}{20}\frac{n_k}{Q_k}w_k^2\sqrt{Q_k n_k} \\ &= n_k\left(\frac{3w_k}{w_{k-1}} + \frac{1}{20}\frac{3w_k^3}{w_{k-1}}\sqrt{\frac{w_{k-1}}{3w_k}\lceil\frac{w_{k-1}}{w_k^5}\rceil}\right) \\ &\le n_k\left(\frac{1}{27} + \frac{1}{20}\frac{3w_k^3}{w_{k-1}}\sqrt{\frac{w_{k-1}}{3w_k}\frac{2w_{k-1}}{w_k^5}}\right) \\ &\le n_k\left(\frac{1}{27} + \frac{3}{20}\right) \le \frac{1}{5}n_k.\end{aligned}$$

Therefore, there exists some $j\in[Q_k]$ such that $\mathbb{E}_{\mathbb{P}_j}(N_j)\le\frac{1}{5}n_k$. Define the corresponding expected revenue function under the noise distribution F_j as $r_j(p) = p(1-F_j(p))$. Recall that $f_j(v) = C_f\sum_{i=0}^{k-1}w_i^2 B_{[a_i,b_i]}(v) + w_k^2 B_{I_j}(v) + \sum_{i=k+1}^{\infty}w_i^2 B_{[a_i,b_i]}(v)$. Then the unique optimal price p_j^* satisfies that $q_j^* = \frac{p_j^*-b}{1-b}$ is the unique maximum for both f_j and g_j. Thus $q_j^* = \frac{p_j^*-b}{1-b}\in I_j = [a_k,b_k]$. For any price $p\in[b,1]$ such that $q=\frac{p-b}{1-b}\notin I_j$, we have $f_j(q_j^*) - f_j(q) \geq w_k^2$. Thus we have

$$g_j(q_j^*) - g_j(q) = \frac{f_j(q_j^*) - f_j(q)}{(1+f_j(q_j^*))(1+f_j(q))} \geq \frac{1}{(1+\frac{9}{8}C_f)^2} w_k^2 \geq \frac{1}{2} w_k^2.$$

Thus we have

$$r_j(p_j^*) - r_j(p) = (1-b)(g_j(q_j^*) - g_j(q)) \geq \frac{1-6C_f}{2}\frac{1}{2} w_k^2 \geq \frac{1}{5} w_k^2.$$

For $k \geq 2$, we have $n_k = \lceil \frac{w_{k-1}}{w_k^5} \rceil = \lceil \big(\frac{1}{w_k}\big)^{5-\frac{1}{k}} \rceil \leq 2\big(\frac{1}{w_k}\big)^{5-\frac{1}{k}}$. Thus we have

$$n_k^{\frac{3}{5}-\frac{1}{5k}} \leq 2\big(\frac{1}{w_k}\big)^{3-\frac{1}{k}-\frac{3}{5k}+\frac{1}{5k^2}} \leq 2\big(\frac{1}{w_k}\big)^{3-\frac{1}{k}}.$$

Therefore, for $k \geq 2$, we obtain

$$\begin{aligned}
\text{Regret} &= \mathbb{E}_{\mathbb{P}_j}\Big(\sum_{t=1}^{n_k}\big(r_j(p_j^*) - r_j(p_t)\big)\Big) \\
&= \sum_{t=1}^{n_k}\mathbb{E}_{\mathbb{P}_j}\Big(1_{\frac{p_t-b}{1-b}\in I_j}\big(r_j(p_j^*) - r_j(p_t)\big) + 1_{\frac{p_t-b}{1-b}\notin I_j}\big(r_j(p_j^*) - r_j(p_t)\big)\Big) \\
&\geq \sum_{t=1}^{n_k}\mathbb{E}_{\mathbb{P}_j}\Big(1_{\frac{p_t-b}{1-b}\notin I_j}\big(r_j(p_j^*) - r_j(p_t)\big)\Big) \\
&\geq \sum_{t=1}^{n_k}\mathbb{E}_{\mathbb{P}_j}\big(1_{\frac{p_t-b}{1-b}\notin I_j}\frac{1}{5}w_k^2\big) = \frac{1}{5}w_k^2\mathbb{E}_{\mathbb{P}_j}\big(\sum_{t=1}^{n_k}1_{\frac{p_t-b}{1-b}\notin I_j}\big) \\
&= \frac{1}{5}w_k^2\Big(n_k - \mathbb{E}_{\mathbb{P}_j}\big(\sum_{t=1}^{n_k}1_{\frac{p_t-b}{1-b}\in I_j}\big)\Big) = \frac{1}{5}w_k^2\big(n_k - \mathbb{E}_{\mathbb{P}_j}(N_j)\big) \\
&\geq \frac{1}{5}w_k^2\frac{4}{5}n_k = \frac{4}{25}w_k^2\lceil\frac{w_{k-1}}{w_k^5}\rceil \geq \frac{4}{25}w_k^2\frac{w_{k-1}}{w_k^5} \\
&= \frac{4}{25}w_k^{\frac{1}{k}-3} = \frac{4}{25}\big(\frac{1}{w_k}\big)^{3-\frac{1}{k}} \geq \frac{4}{25}\cdot\frac{1}{2}n_k^{\frac{3}{5}-\frac{1}{5k}} \\
&= \frac{2}{25}n_k^{\frac{3}{5}-\frac{1}{5k}}.
\end{aligned}$$

Now we claim that for any $\delta > 0$, no policy can achieve the regret of $O(T^{\frac{3}{5}-\delta})$. We prove by contradiction. Suppose that there is some $\delta > 0$ and a policy π such that its

T-period regret $\mathrm{Reg}_T(\pi) \leq C_1 T^{\frac{3}{5}-\delta}$ for any noise distribution and $T \in \mathbb{N}^+$ where C_1 is a constant. Since $n_k = \lceil \frac{w_{k-1}}{w_k^5} \rceil = \lceil 3^{5\cdot k! - (k-1)!} \rceil \geq 3^{4\cdot k!}$, there exists some sufficiently large $k > \frac{1}{\delta}$ such that $n_k^{\frac{1}{2k}} \geq 3^{2\cdot(k-1)!} > \frac{25C_1}{2}$. Then we obtain

$$\mathrm{Reg}_{n_k}(\pi) \leq C_1 n_k^{\frac{3}{5}-\delta} < C_1 n_k^{\frac{3}{5}-\frac{1}{k}} < \frac{2}{25} n_k^{\frac{3}{5}-\frac{1}{2k}} < \frac{2}{25} n_k^{\frac{3}{5}-\frac{1}{5k}} \leq \mathrm{Reg}_{n_k}(\pi).$$

This forms a contradiction.

Therefore, for any $\delta > 0$, no policy can achieve the regret of $O(T^{\frac{3}{5}-\delta})$. Namely, the lowest rate of any regret upper bound we could prove for this dynamic pricing problem is $O(T^{3/5})$. In this work, we use $\tilde{\Omega}(T^{3/5})$ to refer to this lower bound on any valid regret upper bound rates. In two previous works (Kleinberg, 2004; Xu and Wang, 2022) with the same type of proved lower bound results (no policy can achieve the regret of $O(T^{\alpha-\delta})$ for any $\delta > 0$), Kleinberg (2004) used $\Omega(T^\alpha)$ while Xu and Wang (2022) used $\tilde{\Omega}(T^\alpha)$. As $\Omega(T^\alpha)$ sometimes refers to a stronger lower bound claim that $\mathrm{Reg}_T(\pi) \geq CT^\alpha, \forall T \in \mathbb{N}^+$, we choose the notation $\tilde{\Omega}(\cdot)$ as paralleled with Xu and Wang (2022). $\square$

B Assumption Verifications for Simulation Settings

B.1 Simulation Setting for Case (A)

For Case **(A)**, we constructed a simulation setting with $\theta_0 = 30, x_t \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(1/2, 1)$, and a uniform mixture noise distribution $\frac{3}{4}\mathrm{Unif}(-15, 0) + \frac{1}{4}\mathrm{Unif}(0, 15)$. Thus the noise CDF F has the form

$$F = \begin{cases} 0, & \text{for } x \in (-\infty, 15], \\ 3/4 + x/20, & \text{for } x \in (-15, 0], \\ 3/4 + x/60, & \text{for } x \in (0, 15], \\ 1, & \text{for } x \in [15, +\infty). \end{cases}$$

Now we verify the seven assumptions.

Assumption 1. It is easy to see that Assumption 1 is satisfied with a constant $W = 30 \geq ||\theta_0||_1$.

Assumption 2. The matrix $\Sigma = \mathbb{E}\big((1, x_t^\top)^\top(1, x_t^\top)\big) = \begin{pmatrix} 1 & 3/4 \\ 3/4 & (3/4)^2 + 1/48 \end{pmatrix}$. Thus Σ has two eigenvalues $\frac{19+\sqrt{349}}{24}, \frac{19-\sqrt{349}}{24}$ and both of them is larger than $1/100$. Thus $\Sigma - \frac{1}{100}\mathbb{I}$ is positive-definite and Assumption 2 is satisfied with a positive constant $c_0 = 1/100$.

Assumption 3. The noise $z_t \in (-15, 15)$ and $x_t^\top\theta_0 \in (15, 30)$. Thus $x_t^\top\theta_0 + z_t \in (0, 45) \subseteq [0, 50]$. Thus Assumption 3 is satisfied with a constant $B = 50$.

Assumption 4. It is easy to see that F is 1-Lipschitz. Thus Assumption 4 is satisfied with the constant $L = 1$.

Assumption 5. For any $x \in \mathcal{X}$, denote $q = x^\top\theta_0 \in (15, 30)$. Then the expected revenue function $f_q(p) = p(1 - F(p - q))$ has the form

$$f_q(p) = \begin{cases} p, & \text{for } p \in [0, q-15], \\ p(1/4 - (p-q)/20) = -\frac{1}{20}(p - \frac{5+q}{2})^2 + \frac{(5+q)^2}{80}, & \text{for } p \in (q-15, q], \\ p(1/4 - (p-q)/60) = -\frac{1}{60}(p - \frac{15+q}{2})^2 + \frac{(15+q)^2}{240}, & \text{for } p \in (q, q+15], \\ 0, & \text{for } p \in (q+15, +\infty). \end{cases}$$

Therefore, for $p \in [0, q-15]$, the optimal price is $p_1^* = q - 15$ with the maximal expected revenue $f_q(p_1^*) = q - 15$. For $p \in [q-15, q]$, the optimal price is $p_2^* = \frac{5+q}{2} \in (q-15, q)$ since $q \in (15, 30)$. The corresponding maximal expected revenue is $f_q(p_2^*) = \frac{(5+q)^2}{80}$. For $p \in [q, q+15]$, the optimal price is $p_3^* = q$ since $\frac{15+q}{2} < q$ for $q \in (15, 30)$. The corresponding maximal expected revenue is $f_q(p_3^*) = \frac{q}{4}$. Note that p_1^* is at the right boundary of the left range $[0, q-15]$

and p_3^* is at the left boundary of the right range $[q, q+15]$. Thus $(p_1^*, f_q(p_1^*))$ and $(p_3^*, f_q(p_3^*))$ are both on the middle quadratic function $-\frac{1}{20}(p-\frac{5+q}{2})^2+\frac{(5+q)^2}{80}$. Let $r=\frac{5}{2}$. Then we have $\min\{p_2^*-p_1^*, p_3^*-p_2^*\}=\min\{\frac{35-q}{2}, \frac{q-5}{2}\}\geq r$. Thus the maximum of the middle quadratic function $f_q(p_2^*)>\max\{f_q(p_1^*), f_q(p_3^*)\}$. Therefore, $p^*(x)=p_2^*=\frac{5+q}{2}$. Actually, by the property of quadratic functions, $f_q(p)$ is non-decreasing on $[0, p^*(x)]$ and non-increasing on $[p^*(x), p_{\max}]$ where $p_{\max}=50$.

For any $p\in[p^*(x)-r, p^*(x)+r]$, we have $p\in[q-15, q]$ and thus

$$f_q(p^*(x))-f_q(p)=\frac{1}{20}(p^*(x)-p)^2. \tag{13}$$

On the other hand, for $p\in[0, p_{\max}]\setminus(p^*(x)-r, p^*(x)+r)$, we have

$$\begin{aligned} f_q(p^*(x))-f_q(p)\leq f_q(p^*(x))=\frac{(5+q)^2}{80}\leq 20 \\ \leq 4\cdot r^2\leq 4(p^*(x)-p)^2 \end{aligned} \tag{14}$$

Combining Equation (13) and Equation (14), we obtain for any x and $p\in[0, p_{\max}]$,

$$f_q(p^*(x))-f_q(p)\leq C(p^*(x)-p)^2$$

where $C=4$. Thus Assumption 5 is satisfied with constant $C=4$.

Therefore, our simulation setting for Case **(A)** satisfy Assumptions 1 – 5, which include the Lipschitz and 2nd-order smoothness assumptions.

B.2 Simulation Setting for Case (B)

For Case **(B)**, we constructed a simulation setting with $\theta_0=30, x_t\overset{\text{i.i.d.}}{\sim}\text{Unif}(1/2,1)$, and a uniform mixture noise distribution $\frac{1}{4}\text{Unif}(-15,0)+\frac{3}{4}\text{Unif}(0,15)$. Thus the noise

CDF F has the form

$$F = \begin{cases} 0, & \text{for } x \in (-\infty, 15], \\ 1/4 + x/60, & \text{for } x \in (-15, 0], \\ 1/4 + x/20, & \text{for } x \in (0, 15], \\ 1, & \text{for } x \in [15, +\infty). \end{cases}$$

Similar to the simulation setting for Case **(A)**, we can verify Assumptions 1 – 4.

Now we prove that Assumption 5 is not satisfied. Namely, there exists a covariate $x \in \mathcal{X} = (1/2, 1)$, such that for any constant C, $f_q(p^*(x)) - f_q(p) \le C(p^*(x) - p)^2$ does not hold for some $p \in [0, p_{\max}]$. Here we use the notation $q = x^\top \theta_0$. Actually, we will prove that for any covariate $x \in \mathcal{X} = (1/2, 1)$ and any constant C, $f_q(p^*(x)) - f_q(p) \le C(p^*(x) - p)^2$ does not hold for some $p \in [0, p_{\max}]$.

Consider any $x \in \mathcal{X} = (1/2, 1)$. Denote $q = x^\top \theta_0 \in (15, 30)$. Then we have

$$f_q(p) = \begin{cases} p, & \text{for } p \in [0, q - 15], \\ p(3/4 - (p - q)/60) = -\frac{1}{60}(p - \frac{45+q}{2})^2 + \frac{(45+q)^2}{240}, & \text{for } p \in (q - 15, q], \\ p(3/4 - (p - q)/20) = -\frac{1}{20}(p - \frac{15+q}{2})^2 + \frac{(15+q)^2}{80}, & \text{for } p \in (q, q + 15], \\ 0, & \text{for } p \in (q + 15, +\infty). \end{cases}$$

Therefore, for $p \in [0, q-15]$, the optimal price is $p_1^* = q-15$ with the maximal expected revenue $f_q(p_1^*) = q - 15$. For $p \in [q - 15, q]$, the optimal price is $p_2^* = q$ since $\frac{45+q}{2} > q$ for $q \in (15, 30)$. The corresponding maximal expected revenue is $f_q(p_2^*) = \frac{3q}{4}$. For $p \in [q, q + 15]$, the optimal price is $p_3^* = q = p_2^*$ since $\frac{15+q}{2} < q$ for $q \in (15, 30)$. The corresponding maximal expected revenue is $f_q(p_3^*) = \frac{3q}{4}$. Note that p_1^* is at the right boundary of the left range $[0, q - 15]$, p_2^* is at the right boundary of the middle range $[q-15, q]$ and p_3^* is at the left boundary of the right range $[q, q+15]$. Thus $(p_1^*, f_q(p_1^*))$ is on the middle quadratic function $-\frac{1}{60}(p - \frac{45+q}{2})^2 + \frac{(45+q)^2}{240}$ and $f_q(p_1^*) < f_q(p_2^*) = f_q(p_3^*)$.

Therefore, the optimal price $p^*(x) = p_2^* = p_3^* = q$.

The left derivative of $f_q(p)$ at the optimal price $p^*(x) = q$ is

$$f'_{q,-}(p^*(x)) = 3/4 - \frac{p^*(x)}{30} + \frac{q}{60} = \frac{45-q}{60} \geq \frac{1}{4}.$$

Denote $C_1 = f'_{q,-}(p^*(x)) \geq \frac{1}{4}$, then we have

$$\lim_{p \to q^-} \frac{f_q(p^*(x)) - f_q(p)}{p^*(x) - p} = C_1.$$

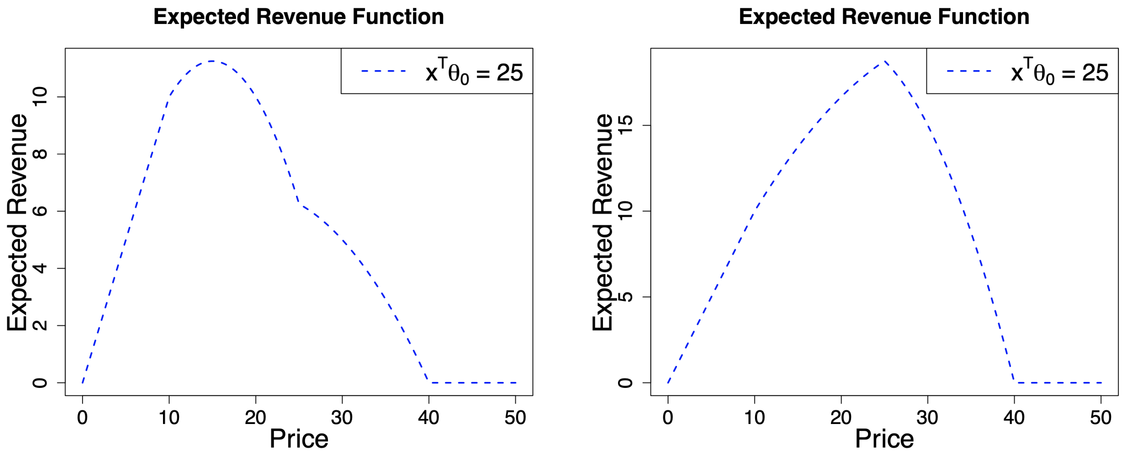
Thus there exists a constant $C_2 \leq 1$ such that for any $p \in (p^*(x) - C_2, p^*(x))$, $\frac{f_q(p^*(x)) - f_q(p)}{p^*(x)-p} \geq \frac{C_1}{2} \geq \frac{1}{8}$. Thus for any constant $C > 0$, we can select $p \in (\max\{p^*(x) - C_2, p^*(x) - \frac{1}{8C}\}, p^*(x)) \in [0, p_{\max}]$ and obtain

$$f_q(p^*(x)) - f_q(p) \geq \frac{1}{8}(p^*(x) - p) > C(p^*(x) - p)^2.$$

Therefore, for any covariate $x \in \mathcal{X} = (1/2, 1)$ and any constant C, $f_q(p^*(x)) - f_q(p) \leq C(p^*(x) - p)^2$ does not hold for some $p \in [0, p_{\max}]$. Thus, the 2nd-order smoothness Assumption 5 is not satisfied under the simulation setting designed for Case **(B)**.

B.3 Expected Revenue Function Plots

In Figure 1, we plot the expected revenue function $f_q(p) = p(1 - F(p - q))$ for $q = x^\top \theta_0 = 25$ under the two simulation settings with different noise distributions. It matches our theoretical verifications that the simulation setting for Case **(A)** satisfies the 2nd-order smoothness assumption, while the simulation setting for Case **(B)** does not.



(a) Simulation setting for Case **(A)**. (b) Simulation setting for Case **(B)**.

Figure 1: Expected revenue functions for $q = x^\top\theta_0 = 25$ under the two simulation settings for Case **(A)** and Case **(B)**.

C Background Review: Perturbed Linear Bandit

The Perturbed Linear Bandit (PLB) is introduced in (Luo et al., 2021). Here we present the formal definition of PLB.

Definition 1. *The rewards Z_t, parameters ξ_t and action sets $\mathcal{A}_t$ form a perturbed linear bandit with a perturbation constant C_p, if*

$$Z_t = \langle \xi_t, A_t \rangle + \eta_t$$

for any selected action $A_t \in \mathcal{A}_t$, and any two parameters are close to each other, i.e., $||\xi_s - \xi_t||_\infty \leq C_p, \forall s, t \in \mathbb{N}^+$. Here η_t is σ-sub-Gaussian conditional on the filtration $\mathcal{F}_{t-1} = \sigma(\xi_1, A_1, Z_1, \ldots, \xi_t, A_t)$.

The "all-close-to-each-other" condition on the linear parameters ξ_t's implies the existence of a "central" parameter ξ^* such that $||\xi_t - \xi^*||_\infty \leq \frac{C_p}{2}$ for any t. Thus the reward structure at time t regulated by ξ_t can be viewed as a perturbation from that regulated by ξ^*. The linear bandit (Abbasi-Yadkori et al., 2011; Chu et al., 2011; Agrawal and Goyal, 2013) is a zero-perturbation PLB with $\xi_t = \xi^*$ for any t.

D Ethic Issues

In this work, we considered a dynamic pricing problem where some sales-relevant contextual information, such as product features and market environments, are available at each selling period. In practical settings of dynamic pricing, the available contextual information may contain customer characteristics. Recently, the study of personalized pricing has garnered some research interest (Ban and Keskin, 2021; Elmachtoub et al., 2021; Chen and Gallego, 2022). In some cases, it would be embarassing if the seller could make use of the customers' personal information and set different prices for the same product towards different customers. However, the customized pricing approaches are common and widely accepted by consumers in insurance and lending industries. Such first-degree price discriminations are also practiced on many popular e-commerce platforms. In fact, the classical sales measure of tailor-made discount coupons also results in different prices for different cunsumers.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320.

Agrawal, S. and Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135.

Ban, G.-Y. and Keskin, N. B. (2021). Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science* **67,** 5549–5568.

Chen, N. and Gallego, G. (2022). A primal–dual learning algorithm for personalized dynamic pricing with an inventory constraint. *Mathematics of Operations Research* .

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.

Elmachtoub, A. N., Gupta, V., and Hamilton, M. L. (2021). The value of personalized pricing. *Management Science* **67,** 6055–6070.

Kleinberg, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems* **17,**.

Luo, Y., Sun, W. W., et al. (2021). Distribution-free contextual dynamic pricing. *arXiv preprint arXiv:2109.07340* .

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

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.