Online Appendix for "Dynamic Pricing With Infrequent Inventory Replenishments"

Boxiao Chen, Menglong Li, David Simchi-Levi

Appendix EC.1: Salvage Value

We show that the deviation of the regret with a salvage value different from the unit ordering cost c is upper bounded by $O(H)$. Suppose that inventories or backlogs in epoch $T+1$ can be salvaged with a unit value s. Denote this system by $\mathcal{S}^s$ and the system with unit salvage value c by $\mathcal{S}^c$. For system $\mathcal{S}^s$, the total expected profit over the planning horizon with decisions $(y_1, p_{11}, ..., p_{1H}, ..., y_T, p_{T1}, ..., p_{TH})$ is

$$g^s(y_1, p_{11}, ..., p_{1H}, ..., y_T, p_{T1}, ..., p_{TH}) = cx_1 + \mathbb{E}\left[\sum_{t=1}^{T} g_t(y_t, p_{t1}, ..., p_{tH})\right] + (s-c)\mathbb{E}[x_{T+1}].$$

For any learning algorithm π, we can define regret $\overline{R}_T(\pi)$ for system $\mathcal{S}^s$ by

$$\mathbb{E}[g^s(\pi^*) - g^s(\pi)],$$

where π^* is an optimal inventory and pricing policy for system $\mathcal{S}^s$ and $g^s(\pi)$ is the total expected profit of π.

LEMMA EC.1. *For any policy π and learning environment $I \in \mathcal{I}$, $|R_T^I(\pi) - \overline{R}_T^I(\pi)| \leqslant O(H)$*

Proof. Since the optimal policy in system $\mathcal{S}^c$ is a feasible policy of system $\mathcal{S}^s$,

$$R_T^I(\pi) - \overline{R}_T^I(\pi) \leqslant -(s-c)\mathbb{E}[x_{T+1}^*] - (s-c)\mathbb{E}[x_{T+1}^\pi],$$

where $\mathbb{E}[x_{T+1}^*]$ and $\mathbb{E}[x_{T+1}^\pi]$ are the expected inventories or backlogs in epoch $T+1$ for the optimal policy of system $\mathcal{S}^c$ and π, respectively. Consider the last epoch, after ordering to some inventory level (must be nonnegative), the total demand in this epoch is at most $H\bar{d}$. Thus the expected inventories or backlogs in epoch $T+1$ is at most $O(H)$. Hence, $R_T^I(\pi) - \overline{R}_T^I(\pi) \leqslant O(H)$. Similarly, $\overline{R}_T^I(\pi) - R_T^I(\pi) \leqslant O(H)$. $\square$

By Lemma EC.1, the worst-case regret bound for system $\mathcal{S}^s$ is the same as that in Theorem 1 when $H \leq O(T)$.

Appendix EC.2: Proofs in Section 6

Proof of Lemma 2. Fix a problem instance I. By the definition of G_T^I, it suffices to find a policy π in Π^S under the full-information setting such that

$$\mathbb{E}\left[\sum_{t=1}^{T} g(y^*, p_1^*(\cdot), ..., p_H^*(\cdot)|D, \epsilon) - g(y^\pi, p_1^\pi(\cdot), ..., p_H^\pi(\cdot)|D, \epsilon)\right] \leqslant (\overline{p}\overline{d} + 2L\overline{d})(H - m - 1)T. \quad \text{(EC.1)}$$

Such a policy π can be constructed as follows. At the beginning of each epoch t, π orders up to the inventory level y^* of the optimal policy and then implements $p_1^*(y^*)$ for periods $1, ..., S_1 - 1$. Let x_{tS_1} be the inventory level at the beginning of period S_1 as if implementing the optimal policy $p_1^*(\cdot), ..., p_{S_1-1}^*(\cdot)$ during periods $1, ..., S_1 - 1$. For periods $S_1, ..., S_2 - 1$, policy π impliments the price $p_{S_1}^*(x_{tS_1})$. Repeat this process until the end of epoch t. The profit gap in epoch t between the optimal policy and π under the same realization of random noises is

$$\begin{aligned}
&\sum_{i \neq S_k, k=0,1,...,m} [p_i^* D_i(p_i^*) - p_i^\pi D_i(p_i^\pi)] - G\left(y^* - \sum_{i=1}^{H} D_i(p_i^*) - \sum_{i=1}^{H} \epsilon_{ti}\right) \\
&\quad + G\left(y^* - \sum_{i=S_k, k=0,...,m} D_i(p_i^*) - \sum_{i \neq S_k, k=0,...,m}^{H} D_i(p_i^\pi) - \sum_{i=1}^{H} \epsilon_{ti}\right) \\
\leqslant& \sum_{i \neq S_k, k=0,1,...,m} \overline{p}(\overline{d} - \epsilon_{ti}) + L\left|\sum_{i \neq S_k, k=0,...,m} D_i(p_i^*) - D_i(p_i^\pi)\right| \\
\leqslant& \overline{p}\overline{d}(H - m - 1) - \sum_{i \neq S_k, k=0,1,...,m} \overline{p}\epsilon_{ti} + 2L\overline{d}(H - m - 1),
\end{aligned}$$

where the first inequality is from the upper bounds of demand and price and the Lipschitz continuity of function $G(\cdot)$, and the second inequality is from the upper bound of demand. The inequality (EC.1) is obtained by taking summation over t and taking the expectation of the noises. □

Proof of Lemma 4. Note that ϵ_s^k is $(\sigma\sqrt{S_k - S_{k-1}})$-sub-Gaussian. By applying Theorem 2 of Abbasi-yadkori et al. (2011) with $V = (S_k - S_{k-1})I$, we have with probability $1 - \delta/(m+1)$ uniformly over $t \geqslant 1$ that

$$\|\hat{\theta}_t^k - \theta^k\|_{\Lambda_{tk}} \leqslant \sigma\sqrt{S_k - S_{k-1}}\sqrt{l\log\left(\frac{1 + tL^2/(S_k - S_{k-1})}{\delta/(m+1)}\right)} + \sqrt{S_k - S_{k-1}}L \leqslant \gamma_k.$$

By Cauchy-Schwarz inequality,

$$|\eta(p)\cdot(\hat{\theta}_t^k-\theta^k)| \leqslant \|\eta(p)\|_{\Lambda_{tk}^{-1}}\|\hat{\theta}_t^k-\theta^k\|_{\Lambda_{tk}} \leqslant \gamma_k\|\eta(p)\|_{\Lambda_{tk}^{-1}} = \Delta_{tk}(p).$$

The statement then follows from the union bound inequality. $\square$

Proof of Lemma 5. We have

$$\begin{aligned}
\sum_{s=1}^t \Delta_{sk}(p_{sS_{k-1}}) &\leqslant \sqrt{t}\sqrt{\sum_{s=1}^t \Delta_{sk}(p_{sS_{k-1}})^2} \\
&\leqslant L\gamma_k\sqrt{t}\sqrt{\sum_{s=1}^t \min\{1, \|\eta(p_{sS_{k-1}})\|^2_{\Lambda_{sk}^{-1}}\}} \\
&\leqslant L\gamma_k\sqrt{t}\sqrt{2\sum_{s=1}^t \log(1+\|\eta(p_{sS_{k-1}})\|^2_{\Lambda_{sk}^{-1}})},
\end{aligned}$$

where the second inequality is from $\|\eta(p)\|_{\Lambda_{sk}^{-1}} \leqslant \|\eta(p)\|_2 \leqslant L$ and the third one is from $e^{x/2} \leqslant 1+x$ for $x\in[0,1]$.

Since $\Lambda_{tk} = \Lambda_{t-1,k} + \eta(p_{t-1,S_{k-1}})\eta(p_{t-1,S_{k-1}})^\top$,

$$\begin{aligned}
\det(\Lambda_{tk}) &= \det(\Lambda_{t-1,k})\det(I+\Lambda_{t-1,k}^{1/2}\eta(p_{t-1,S_{k-1}})\eta(p_{t-1,S_{k-1}})^\top\Lambda_{t-1,k}^{1/2}) \\
&= \det(\Lambda_{t-1,k})(1+\|\eta(p_{t-1,S_{k-1}})\|^2_{\Lambda_{t-1,k}^{-1}}),
\end{aligned}$$

and thus

$$\sum_{s=1}^t \log(1+\|\eta(p_{sS_{k-1}})\|^2_{\Lambda_{sk}^{-1}}) = \log(\det(\Lambda_{t+1,k})) - \log(\det(\Lambda_{1k})) \leqslant l\log(S_k - S_{k-1} + tL^2).$$

Therefore, $\sum_{s=1}^t \Delta_{sk}(p_{sS_{k-1}}) \leqslant L\gamma_k\sqrt{2l\log(S_k-S_{k-1}+tL^2)}\sqrt{t}$ with probability 1. $\square$

Proof of Lemma 6. Let $\mathcal{H}_{>t}^k = [t]\setminus\mathcal{H}_{<t}^k$. By definition and Lemma 5, $\alpha_k|\mathcal{H}_{>t}^k|/\sqrt{t} \leqslant \sum_{s\in\mathcal{H}_{>t}^k}\Delta_{sk}(p_{sS_{k-1}}) \leqslant \sum_{s\in[t]}\Delta_{sk}(p_{sS_{k-1}}) \leqslant L\gamma_k\sqrt{2l\log(S_k-S_{k-1}+tL^2)}\sqrt{t}$. It follows that $|\mathcal{H}_{>t}^k| \leqslant tL\gamma_k\sqrt{2l\log(S_k-S_{k-1}+tL^2)}/\alpha_k \leqslant t/2$. $\square$

Proof of Lemma 7. Let $\tilde{\mu}_t^k = \frac{1}{|\mathcal{H}_{<t}^k|}\sum_{s\in\mathcal{H}_{<t}^k}\mathbb{I}[d_s^k - \eta(p_{sS_{k-1}})\cdot\theta^k]$ be the empirical distribution of the aggregated noise ϵ_t^k. Since ϵ_t^k is $\sigma\sqrt{S_k - S_{k-1}}$-sub-Gaussian, the variance of ϵ_t^k is bounded by $\sigma^2(S_k - S_{k-1})$. By Theorem 1 of Fournier and Guillin (2015),

$$\mathbb{E}[W_1(\mu_t^k, \tilde{\mu}_t^k)] \leqslant C\sigma\sqrt{S_k - S_{k-1}}|\mathcal{H}_{<t}^k|^{-1/2} \leqslant \sqrt{2}C\sigma\sqrt{\frac{S_k - S_{k-1}}{t}}.$$

Since under event A, $|(d_s^k - \eta(p_{sS_{k-1}})\cdot\hat{\theta}_s^k) - (d_s^k - \eta(p_{sS_{k-1}})\cdot\theta^k)| = |\eta(p_{sS_{k-1}})\cdot(\hat{\theta}_s^k - \theta_k)| \leqslant \Delta_{sk}(p_{sS_{k-1}}) \leqslant \alpha_k/\sqrt{t}$ for $s\in\mathcal{H}_{<t}^k$, we have $W_1(\tilde{\mu}_t^k, \hat{\mu}_t^k) \leqslant \alpha_k/\sqrt{t}$. It follows that

$$\mathbb{E}[W_1(\tilde{\mu}_t^k, \hat{\mu}_t^k)|A] \leqslant \frac{\alpha_k}{\sqrt{t}}.$$

By the proof of Lemma 4, $\mathbb{P}(A) \geqslant 1-\delta$. Therefore,

$$\begin{aligned}\mathbb{E}[W_1(\mu_t^k, \hat{\mu}_t^k)|A] &\leqslant \mathbb{E}[W_1(\mu_t^k, \tilde{\mu}_t^k)|A] + \mathbb{E}[W_1(\tilde{\mu}_t^k, \hat{\mu}_t^k)|A] \\ &\leqslant \frac{C\sigma}{1-\delta}\sqrt{\frac{S_k - S_{k-1}}{t}} + \frac{\alpha_k}{\sqrt{t}}.\end{aligned}$$

□

Proof of Lemma 8. This is clearly true for $k = N+1$. Suppose the statement is true for $k+1$. Without loss of generality, assume $u_k(y) \geqslant u_k(y')$. Then

$$\begin{aligned}&|u_k(y) - u_k(y')| = u_k(y) - u_k(y') \\ \leqslant& \max_{p\in\mathcal{P}}\{\mathbb{E}[R_k(p) - G(y - D_k(p) - \beta_k)\mathbb{I}_{k=N} + u_{k+1}(y - D_k(p) - \beta_k)] \\ &- \mathbb{E}[R_k(p) - G(y' - D_k(p) - \beta_k)\mathbb{I}_{k=N} + u_{k+1}(y' - D_k(p) - \beta_k)]\} \\ \leqslant& L|y-y'|\mathbb{I}_{k=N} + \max_{p\in\mathcal{P}}\{\mathbb{E}[u_{k+1}(y - D_k(p) - \beta_k) - u_{k+1}(y' - D_k(p) - \beta_k)]\} \\ \leqslant& L|y-y'|.\end{aligned}$$

□

Proof of Lemma 9. We prove by induction on k. This clearly holds for $k = H+1$. Suppose this holds for $k+1$. For a fixed y, let p_k be an optimal solution for $u_k(y)$. By assumption, there exists $p_k' \in \mathcal{P}$ such that $\bar{D}_k(p_k') = D_k(p_k)$ and $\bar{R}_k(p_k') \geqslant R_k(p_k)$. Then

$$\bar{u}_k(y) \geqslant \mathbb{E}[\bar{R}_k(p_k') - G(y - \bar{D}_k(p_k') - \beta_k)\mathbb{I}_{k=N} + \bar{u}_{k+1}(y - \bar{D}_k(p_k') - \beta_k)]$$

$$
\begin{aligned}
&= \mathbb{E}[\bar{R}_k(p'_k) - G(y - D_k(p_k) - \beta_k)\mathbb{I}_{k=N} + \bar{u}_{k+1}(y - D_k(p_k) - \beta_k)] \\
&\geqslant \mathbb{E}[R_k(p_k) - G(y - D_k(p_k) - \beta_k)\mathbb{I}_{k=N} + \bar{u}_{k+1}(y - D_k(p_k) - \beta_k)] \\
&\geqslant \mathbb{E}[R_k(p_k) - G(y - D_k(p_k) - \beta_k)\mathbb{I}_{k=N} + u_{k+1}(y - D_k(p_k) - \beta_k)] \\
&= u_k(y).
\end{aligned}
$$

□

Proof of Lemma 10. It is easy to see that

$$
\begin{aligned}
&|u_k(y) - \hat{v}_k(y)| \\
\leqslant &\inf_{\xi}\{\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|G(y - D_k(p_k^*(y)) - \beta_k) - G(y - D_k(p_k^*(y)) - \hat{\beta}_k)|\mathbb{I}_{k=N}\} \\
&+ \mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \beta_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)| \\
\leqslant &LW_1(\beta_k, \hat{\beta}_k)\mathbb{I}_{k=N} + \mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \beta_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)|.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \beta_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)| \\
\leqslant &\inf_{\xi}\{\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y - D_k(p_k^*(y)) - \beta_k) - u_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)|\} \\
&+ \mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)| \\
\leqslant &LW_1(\beta_k, \hat{\beta}_k)\mathbb{I}_{k\neq N} + \mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)|.
\end{aligned}
$$

The last inequality is from Lemma 8. Hence,

$$|u_k(y) - \hat{v}_k(y)| \leqslant LW_1(\beta_k, \hat{\beta}_k) + \mathbb{E}|u_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k) - \hat{v}_{k+1}(y - D_k(p_k^*(y)) - \hat{\beta}_k)|.$$

By induction,

$$|u_k(y) - \hat{v}_k(y)| \leqslant L\sum_{i=k}^{N} W_1(\beta_i, \hat{\beta}_i).$$

□

Proof of Lemma 11. The proof is similar to that of Lemma 10.

$$
\begin{aligned}
&|u_k(y)-\tilde{v}_k(y)|\\
\leqslant&|R_k(p_k^*)-\bar{R}_k(p_k^*)|+\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|G(y-D_k(p_k^*)-\beta_k)-G(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\mathbb{I}_{k=N}\\
&+\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y-D_k(p_k^*)-\beta_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\\
\leqslant&|R_k(p_k^*)-\bar{R}_k(p_k^*)|+L|D_k(p_k^*)-\bar{D}_k(p_k^*)|\mathbb{I}_{k=N}+L\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|\beta_k-\hat{\beta}_k|\mathbb{I}_{k=N}\\
&+\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y-D_k(p_k^*)-\beta_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\\
\leqslant&|R_k(p_k^*)-\bar{R}_k(p_k^*)|+L|D_k(p_k^*)-\bar{D}_k(p_k^*)|\mathbb{I}_{k=N}+LW_1(\beta_k,\hat{\beta}_k)\mathbb{I}_{k=N}\\
&+\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y-D_k(p_k^*)-\beta_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|.
\end{aligned}
$$

We only need to bound the last term.

$$
\begin{aligned}
&\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y-D_k(p_k^*)-\beta_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\\
\leqslant&\inf_{\xi}\mathbb{E}_{\xi\sim(\beta_k,\hat{\beta}_k)}|u_{k+1}(y-D_k(p_k^*)-\beta_k)-u_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\\
&+\mathbb{E}_{\hat{\beta}}|u_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|\\
\leqslant&L(|D_k(p_k^*)-\bar{D}_k(p_k^*)|+W_1(\beta_k,\hat{\beta}_k))\mathbb{I}_{k\neq N}+\mathbb{E}_{\hat{\beta}}|u_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|.
\end{aligned}
$$

The last inequality is by Lemma 8. We now have

$$
\begin{aligned}
|u_k(y)-\tilde{v}_k(y)|\leqslant&|R_k(p_k^*)-\bar{R}_k(p_k^*)|+L|D_k(p_k^*)-\bar{D}_k(p_k^*)|+LW_1(\beta_k,\hat{\beta}_k)\\
&+\mathbb{E}_{\hat{\beta}}|u_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)-\tilde{v}_{k+1}(y-\bar{D}_k(p_k^*)-\hat{\beta}_k)|.
\end{aligned}
$$

The lemma is then obtained by induction. □

Proof of Lemma 3. Let A be the event that $|\eta(p)\cdot(\hat{\theta}_t^k-\theta^k)|\leqslant\Delta_{tk}(p)$ for all $t\in[T],k\in[m+1],p\in\mathcal{P}$. By the proof of Lemma 4, $\mathbb{P}(A)\geqslant 1-\delta$.

We first show that under event A, $\overline{D_t^k}(p)$ and $\overline{R_t^k}(p)=p\overline{D_t^k}(p)$ satisfy the assumption in Lemma 9. For any $p\in[\underline{p},\overline{p}]$, since $\overline{D_t^k}(p)\geqslant D_t^k(p)$ and $\overline{D_t^k}(\overline{p})\leqslant D_t^k(\overline{p})\leqslant D_t^k(p)$, then continuity of $\overline{D_t^k}(\cdot)$ implies that there exists $p'\geqslant p$ such that $\overline{D_t^k}(p')=D_t^k(p)$ and thus $\overline{R_t^k}(p')=p'\overline{D_t^k}(p')=p'D_t^k(p)\geqslant pD_t^k(p)=R_t^k(p)$.

Note that the decisions $(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot))$ given by Algorithm 1 are from the solutions of dynamic program (1) with estimated $\overline{D_t^k}$ and $\hat{\epsilon}_t^k$, $k \in [m+1]$. Lemma 9 implies that $\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \hat{\epsilon}) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|\bar{D}, \hat{\epsilon}) \mid A] \leqslant 0$. It follows that

$$\begin{aligned}
&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A] \\
=&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \hat{\epsilon}) \mid A] \\
&+\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \hat{\epsilon}) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|\bar{D}, \hat{\epsilon}) \mid A] \\
&+\mathbb{E}[g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|\bar{D}, \hat{\epsilon}) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A] \\
\leqslant&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \hat{\epsilon}) \mid A] \\
&+\mathbb{E}[g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|\bar{D}, \hat{\epsilon})] - \mathbb{E}[g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A].
\end{aligned}$$

By Lemma 10,

$$\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \hat{\epsilon}) \mid A] \leqslant L \sum_{k=1}^{m+1} \mathbb{E}[W_1(\mu_t^k, \hat{\mu}_t^k) \mid A]$$

By Lemma 11,

$$\begin{aligned}
&\mathbb{E}[g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|\bar{D}, \hat{\epsilon}) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A] \\
\leqslant&\mathbb{E}\left[\sum_{k=1}^{m+1} |p_{tS_{k-1}} \overline{D_t^k}(p_{tS_{k-1}}) - p_{tS_{k-1}} D_t^k(p_{tS_{k-1}})| + L\Delta_{tk}(p_{tS_{k-1}}) + LW_1(\mu_t^k, \hat{\mu}_t^k) \middle| A\right] \\
\leqslant&\mathbb{E}\left[\sum_{k=1}^{m+1} 2\bar{p}\Delta_{tk}(p_{tS_{k-1}}) + L\Delta_{tk}(p_{tS_{k-1}}) + LW_1(\mu_t^k, \hat{\mu}_t^k) \middle| A\right] \\
=&(2\bar{p} + L)\mathbb{E}\left[\sum_{k=1}^{m+1} \Delta_{tk}(p_{tS_{k-1}}) \middle| A\right] + L\sum_{k=1}^{m+1} \mathbb{E}\left[W_1(\mu_t^k, \hat{\mu}_t^k) | A\right],
\end{aligned}$$

where the expectation is taken over the randomness of $p_{tS_{k-1}}$. It is clear that $g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) \leqslant H\bar{p}\bar{d}$. Thus,

$$\begin{aligned}
&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon)] \\
\leqslant&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A] \\
&+\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid \overline{A}]\mathbb{P}(\overline{A}) \\
\leqslant&(2\bar{p} + L)\mathbb{E}\left[\sum_{k=1}^{m+1} \Delta_{tk}(p_{tS_{k-1}}) \middle| A\right] + 2L\sum_{k=1}^{m+1} \mathbb{E}\left[W_1(\mu_t^k, \hat{\mu}_t^k) | A\right] + \delta H\bar{p}\bar{d}.
\end{aligned}$$

Set $\delta = \frac{1}{TH}$. By Lemma 5, with probability 1,

$$\sum_{k=1}^{m+1}\sum_{t=1}^{T}\Delta_{tk}(p_{tS_{k-1}}) \leqslant \sum_{k=1}^{m+1} L\gamma_k\sqrt{2l\log(S_k - S_{k-1} + TL^2)}\sqrt{T} = \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right).$$

By Lemma 7,

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{k=1}^{m+1}\mathbb{E}\left[W_1(\mu_t^k, \hat{\mu}_t^k)|A\right]\\
\leqslant&\sum_{t=1}^{T}\sum_{k=1}^{m+1}\frac{C\sigma}{1-1/(TH)}\sqrt{\frac{S_k - S_{k-1}}{t}} + \sum_{t=1}^{T}\sum_{k=1}^{m+1}\frac{2L\gamma_k\sqrt{2l\log(S_k - S_{k-1} + tL^2)}}{\sqrt{t}}\\
=&\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right).
\end{aligned}$$

Combine the above inequalities, we have

$$\tilde{R}_T^I(\pi_S) \leqslant \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right).$$

□

Appendix EC.3: Proof of Theorem 2

We first prove the following technical result.

LEMMA EC.2. *Let A, B and C be positive semi-definite matrices such that $A = B + C$. If A, B are invertible, then*

$$\sup_{x\neq 0}\frac{x^\top B^{-1}x}{x^\top A^{-1}x} \leq \frac{\det A}{\det B}.$$

Proof of Lemma EC.2. Fix $x \neq 0$ and let $y = A^{-1}x$. Clearly, $y \neq 0$, $x^\top A^{-1}x = y^\top Ay$, and $x^\top B^{-1}x = y^\top AB^{-1}Ay$. Note that $AB^{-1}A = AB^{-1}(B + C) = A(I + B^{-1}C) = A + AB^{-1}C$ and $AB^{-1}C$ is positive semi-definite. By Lemma 12 of Abbasi-yadkori et al. (2011), we have $y^\top(A + AB^{-1}C)y \leq \frac{\det(A+AB^{-1}C)}{\det A}y^\top Ay$, which implies

$$x^\top B^{-1}x = y^\top AB^{-1}Ay \leq \frac{(\det A)^2\det B^{-1}}{\det A}y^\top Ay = \frac{\det A}{\det B}x^\top A^{-1}x.$$

□

In Algorithm 2, the estimated demand functions are updated much more rarely than in Algorithm 1. Nevertheless, Lemma EC.2 implies that the lengths of the confidence intervals for each epoch in Algorithms 1 and 2 are of the same order. More specifically, let τ_t be the last epoch (from epoch 1 to epoch t) in which the estimated demand function and noise distributions are updated in Algorithm 2.

Corollary EC.1. *For any $p \in \mathcal{P}$ and t,k, we have $\Delta_{\tau_t k}(p) \le \sqrt{2}\Delta_{tk}(p)$.*

Proof of Corollary EC.1. By Algorithm 2, the estimations are updated when the regularized design matrix is doubled for some k. Thus, $\det \Lambda_{tk} \le 2 \det \Lambda_{\tau_t k}$ for all k. By Lemma EC.2, $\Delta_{\tau_t k}(p) = \gamma_k \sqrt{\eta(p)^\top \Lambda_{\tau_t k}^{-1} \eta(p)} \le \gamma_k \sqrt{2\eta(p)^\top \Lambda_{tk}^{-1}\eta(p)} = \sqrt{2}\Delta_{tk}(p)$. □

The following lemma implies that the aggregated statistical error for the estimated noise distributions is still bounded by $O(\sqrt{T})$.

Lemma EC.3. *For any T, we have $\sum_{t=1}^{T} \frac{1}{\sqrt{\tau_t}} \le O(\sqrt{T})$.*

Proof of Lemma EC.3. By the update rule of Algorithm 2, $t \le 2\tau_t$. Thus, $\sum_{t=1}^{T} \frac{1}{\sqrt{\tau_t}} \le \sum_{t=1}^{T} \frac{1}{\sqrt{t/2}} = O(\sqrt{T})$. □

We are ready to prove Theorem 2.

Proof of Theorem 2. We only need to prove that the learning regret of π_S satisfies

$$\sup_{I \in \mathcal{I}} \tilde{R}_T^I(\pi_S) \leqslant \tilde{O}\Big(\sum_{k=1}^{m+1} \sqrt{S_k - S_{k-1}}\sqrt{T}\Big).$$

For ease of exposition, we consider the following variant of Assumption 2(d): there exists $\delta_0 > 0$ and $i \in [H]$ such that for all $t \in [T]$, $D_{ti}(\underline{p}) + \epsilon_{ti} \ge \delta_0$ almost surely. One can apply concentration inequalities to reduce Assumption 2(d) to this variant.

After solving each DP, the number of epochs used to reduce the inventory level is upper bounded by $H\bar{d}/\delta_0$. As the number of DP computations is bounded by $O(\log(T))$. The learning regret due to the inventory clearance process is upper bounded by $O(H\log(T))$. In the following, we only consider epochs whose initial inventory is no more than the target order-up-to level.

Let A be the event that $|\eta(p)\cdot(\hat{\theta}^k_{\tau_t}-\theta^k)|\leqslant \Delta_{\tau_t k}(p)$ for all $t\in[T], k\in[m+1], p\in\mathcal{P}$. By the same argument of the proof of Lemma 3, we have

$$\begin{aligned}
&\mathbb{E}[g_t(y^S, p_1^S(\cdot),...,p_H^S(\cdot)|D,\epsilon) - g_t(y_t, p_{t1}(\cdot),...,p_{tH}(\cdot)|D,\epsilon)]\\
\leqslant&(2\bar{p}+L)\mathbb{E}\left[\sum_{k=1}^{m+1}\Delta_{\tau_t k}(p_{tS_{k-1}})\middle| A\right] + 2L\sum_{k=1}^{m+1}\mathbb{E}\left[W_1(\mu^k_{\tau_t},\hat{\mu}^k_{\tau_t})|A\right] + \delta H\bar{p}\bar{d}.
\end{aligned}$$

Set $\delta=\frac{1}{TH}$. By Lemma 5 and Corollary EC.1, with probability 1,

$$\sum_{t=1}^T \Delta_{\tau_t k}(p_{tS_{k-1}}) \leqslant \sqrt{2}\sum_{t=1}^T \Delta_{tk}(p_{tS_{k-1}}) = \tilde{O}\left(\sqrt{S_k - S_{k-1}}\sqrt{T}\right).$$

By Lemmas 7 and EC.3,

$$\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\left[W_1(\mu^k_{\tau_t},\hat{\mu}^k_{\tau_t})|A\right]\\
\leqslant&\sum_{t=1}^T \frac{C\sigma}{1-1/(TH)}\sqrt{\frac{S_k-S_{k-1}}{\tau_t}} + \sum_{t=1}^T \frac{2L\gamma_k\sqrt{2l\log(S_k-S_{k-1}+tL^2)}}{\sqrt{\tau_t}}\\
=&\tilde{O}\left(\sqrt{S_k-S_{k-1}}\sqrt{T}\right).
\end{aligned}$$

Combine the above inequalities and note that $H\le O(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T})$, we have

$$\tilde{R}^I_T(\pi_S)\leqslant \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right).$$

□

Appendix EC.4: Results for Different Inventory Cost-Accounting Schemes

In this section, we extend our results to the settings where inventory costs are accounted for more frequently (instead of being accounted for every epoch in the original model). We first consider a cost-accounting scheme where inventory cost is accounted for every period, and then consider a cost-accounting scheme where inventory cost is accounted for in the period before the price change.

EC.4.1. Account for Inventory Cost Every Period

We now consider the case where the inventory cost is accounted for every period. Specifically, we assume that at the end of period i of epoch t, $i\in[H], t\in[T]$, a holding and backlogging cost $G(\cdot)$ depending on the current inventory level is incurred. We still impose Assumption 1 and a slightly modified Assumption 2 defined below.

Assumption EC.1. *Assumption 2 holds except that Assumption 2(e) is replaced by the following: function $G(\cdot)$ is L/H-Lipschitz continuous.*

This assumption requires that the per-period inventory cost scales with the length of a period, which is $O(\frac{1}{H})$ (if considering the length of an epoch is fixed). This is especially relatable under the simplest setting of linear holding and backlog costs. If this assumption does not hold, we discuss its impact on regret after presenting Theorem EC.1 in this section.

Under this cost-accounting scheme, one must estimate the inventory level at the end of each period to calculate the inventory cost, which requires estimating the accumulated demand $\sum_{j=S_{k-1}}^{i} D_{tj}(p)$ and noise $\sum_{j=S_{k-1}}^{i} \epsilon_{tj}$ for each period i within two consecutive price change points S_{k-1} and S_k. This leads to the following modifications of our learning algorithm design.

Algorithm Design. The algorithm design is similar to that in Section 3. The main difference is that we need to estimate the accumulated demand and noise for each period. Fix the price change plan $S = (S_0, S_1, \dots, S_{m+1})$. Let $\{d_{si}\}_{s\in[t-1], i\in[H]}$ be the observed historical demand at the beginning of epoch t. For $k \in [m+1]$ and $j = S_{k-1}, \dots, S_k - 1$, we use superscript k,j to denote the corresponding aggregated quantity from period S_{k-1} to j during an epoch (e.g., $\theta^{k,j} = \sum_{i=S_{k-1}}^{j} \theta_i$ and $d_t^{k,j} = \sum_{i=S_{k-1}}^{j} d_{ti}$ for $t \in [T]$). Using regularized least-squares estimation, we construct estimates $\hat{\theta}_t^{k,j}$ of the aggregated demand parameters $\theta^{k,j}$ as

$$\hat{\theta}_t^{k,j} = \operatorname*{arg\,min}_{\theta\in\mathbb{R}^l} \sum_{s=1}^{t-1} \left(d_s^{k,j} - \eta(p_{sS_{k-1}})^\top \theta\right)^2 + \frac{j - S_{k-1} + 1}{2} \|\theta\|_2^2$$

for $k \in [m+1], j = S_{k-1}, \dots, S_k - 1$.

Define $\Lambda_{tkj} = (j - S_{k-1} + 1)I + \sum_{s=1}^{t-1} \eta(p_{sS_{k-1}})\eta(p_{sS_{k-1}})^\top$, where I is an l-by-l identity matrix. For every $p \in \mathcal{P}$, $k \in [m+1]$, and $j = S_{k-1}, \dots, S_k - 1$, define $\Delta_{tkj}(p)$ as

$$\Delta_{tkj}(p) = \gamma_{k,j} \|\eta(p)\|_{\Lambda_{tkj}^{-1}}.$$

Parameters $\gamma_{k,j}$ will be determined later. We construct an upper confidence bound (UCB) of the aggregated demand $D_t^{k,j}(p)$ as

$$\overline{D_t^{k,j}}(p) = \min\left\{\eta(p)^\top \hat{\theta}_t^{k,j} + \Delta_{tkj}(p), D_t^{k,j}(\overline{p}) + (j - S_{k-1} + 1)L^2 \|\overline{p} - p\|_2\right\}, \tag{EC.2}$$

Let $\mathcal{H}_{<t}^{k,j} = \{s \in [t-1] : \Delta_{skj}(p_{sS_{k-1}}) < \alpha_{k,j}/\sqrt{t}\}$, where $\alpha_{k,j}$ is determined later. Define the estimated distribution of the aggregated noise $\epsilon_t^{k,j}$ as

$$\hat{\mu}_t^{k,j} = \frac{1}{|\mathcal{H}_{<t}^{k,j}|} \sum_{s \in \mathcal{H}_{<t}^{k,j}} \mathbb{I}[d_s^{k,j} - \eta(p_{sS_{k-1}})^\top \hat{\theta}_s^{k,j}], \tag{EC.3}$$

where $\mathbb{I}[x]$ is the point mass at x.

It is convenient to define the following dynamic program.

$$u_k(y) = \max_{p \in \mathcal{P}} \mathbb{E}[R_k(p) - \sum_{j=S_{k-1}}^{S_k - 1} G(y - D^{k,j}(p) - \beta^{k,j}) + u_{k+1}(y - D^{k,S_k-1}(p) - \beta^{k,S_k-1})],\ k = 1, \dots, N,$$

$$u_{N+1}(\cdot) = 0, \tag{EC.4}$$

where $N = m+1$, $\mathcal{P} \subseteq \mathbb{R}_+$ is the feasible set of a generic control p, $\beta^{k,j}$ are generic independent random variables, $R_k, D^{k,j}$ are generic univariate functions, and the expectation is taken with respective to $\beta^{k,j}$.

The learning algorithm is similar to Algorithm 2 except that the dynamic program we solve is (EC.4) with $R_k(p) = p\overline{D_t^{k,S_k-1}}(p)$, $D^{k,j}(p) = \overline{D_t^{k,j}}(p)$, $\beta^{k,j} \sim \hat{\mu}_t^{k,j}$, and $\gamma_{k,j} = \sqrt{j - S_{k-1} + 1}\left[\sigma\sqrt{l \log((1+TL^2)TH^2)} + L\right]$, $\alpha_{k,j} = 2L\gamma_{k,j}\sqrt{2l\log(S_k - S_{k-1} + TL^2)}$. We still refer to this algorithm as Algorithm 2 for simplicity.

We need an additional technical assumption.

Assumption EC.2. *For all $t \in [T]$ and $y \in \mathcal{Y}$, the profit-to-go function $u_1(y)$ defined in* (EC.4) *with $R_k(p) = p\overline{D_t^{k,S_k-1}}(p)$, $D^{k,j}(p) = \overline{D_t^{k,j}}(p)$, $\beta^{k,j} \sim \hat{\mu}_t^{k,j}$ is no less than that defined with $R_k(p) = pD_t^{k,S_k-1}(p)$, $D^{k,j}(p) = D_t^{k,j}(p)$, $\beta^{k,j} \sim \hat{\mu}_t^{k,j}$.*

Assumption EC.2 means that it is more profitable if the demand is increased while the price remains the same for each feasible price. This assumption trivially holds when inventory is accounted only at the end of each epoch (Lemma 9) or at the end of the period $S_k - 1$ (Lemma EC.17). This is because one can always increase the price to match aggregated demand during periods S_{k-1} to $S_k - 1$ while collecting more revenue. This argument does not hold when inventory cost is accounted for every period as one cannot find a price p' to match all the accumulated demands $D^{k,j}(p) = \overline{D_t^{k,j}}(p'),\ j = S_{k-1}, \ldots, S_k - 1$. Nevertheless, we can show that Assumption EC.2 holds under a mild cost parameter assumption.

LEMMA EC.4. *Assumption EC.2 holds if* $G(x) = h^+x^+ + h^-x^-$ *with* $h^- \leqslant h^+ + \underline{p}$ *and* $h^+, h^- \geqslant 0$.

Proof. This is clear because under the same price, more demand means more profit when $h^- \leqslant h^+ + \underline{p}$. $\square$

We can now state the main result of this section.

THEOREM EC.1. *Suppose Assumptions 1, EC.1, and EC.2 hold.*

(a) (Lower bound) There is a constant $c > 0$ *such that*

$$\inf_{\pi \in \Pi^S} \sup_{I \in \mathcal{I}} R_T^I(\pi) \geq \frac{(H - m - 1)T}{32} + c \sum_{k=1}^{m+1} \sqrt{S_k - S_{k-1}} \sqrt{T};$$

(b) (Upper bound)

$$\sup_{I \in \mathcal{I}} R_T^I(\pi_S) \leq O((H - m - 1)T) + \tilde{O}\left(\sum_{k=1}^{m+1} \sqrt{S_k - S_{k-1}} \sqrt{T}\right),$$

where π_S *is the policy of Algorithm 2.*

Theorem EC.1 implies that the worst-case regret of Algorithm 2 matches the lower bound. We point out that when Assumption EC.1 does not hold, especially if the per period holding and backlogging costs are of order $O(1)$ instead of $O(\frac{1}{H})$, the regret upper bound would be scaled with a factor of H, i.e., $O((H - m - 1)HT) + \tilde{O}\left(H \sum_{k=1}^{m+1} \sqrt{S_k - S_{k-1}} \sqrt{T}\right)$. It seems that this upper bound is hard to improve in this case. To gain some insights, consider the case that price can be changed every period, i.e., $S_k - S_{k-1} = 1$ for any k, and the regret upper bound we have is $\tilde{O}\left(H^2 \sqrt{T}\right)$ (while the lower bound is $\Omega(H\sqrt{T})$). Because there are H demand functions and noise distributions to be learned and the estimation error will be carried over during H periods, the upper bound is very likely to scale with H^2 no matter what algorithm we use. Thus, we conjecture that the lower bound is not tight when the unit holding and backlogging costs are of order $O(1)$. However, it is rather challenging to improve the lower bound because one needs to construct a hard problem instance where inventory cost plays a significant role and has to investigate the difference between two value-to-go functions for any two pricing and inventory policies. We thus leave this as a future research direction.

The proof of the lower bound is the same as that in Section 5. The proof of the upper bound is similar to that of Theorem 2. By the decomposition $R_T^I(\pi_S) = G_T^I(S) + \tilde{R}_T^I(\pi_S)$, we have

$$\sup_{I\in\mathcal{I}} R_T^I(\pi_S) \leqslant \sup_{I\in\mathcal{I}} G_T^I(S) + \sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S). \tag{EC.5}$$

Note that the first term $\sup_{I\in\mathcal{I}} G_T^I(S)$ is the profit gap due to less pricing flexibility under the full-information setting and the second term $\sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S)$ is the profit gap due to lack of knowledge of the demand-price function as well as the noise distribution.

Upper Bound for $\sup_{I\in\mathcal{I}} G_T^I(S)$

LEMMA EC.5. *It holds that*

$$\sup_{I\in\mathcal{I}} G_T^I(S) \leqslant (\overline{p}\overline{d} + 2L\overline{d})(H-m-1)T.$$

Proof: Fix a problem instance I. Consider the policy π constructed in the proof of Lemma 2. The profit gap in epoch t between the optimal policy and π under the same realization of random noises is

$$\begin{aligned}
&\sum_{i\neq S_k, k=0,1,\dots,m} [p_i^* D_i(p_i^*) - p_i^\pi D_i(p_i^\pi)] - \sum_{j=1}^H \left[G\left(y^* - \sum_{i=1}^j D_i(p_i^*) - \sum_{i=1}^j \epsilon_{ti}\right) - G\left(y^* - \sum_{i=1}^j D_i(p_i^\pi) - \sum_{i=1}^j \epsilon_{ti}\right)\right] \\
\leqslant &\sum_{i\neq S_k, k=0,1,\dots,m} \overline{p}(\overline{d} - \epsilon_{ti}) + \sum_{j=1}^H \frac{L}{H} \left| \sum_{\substack{i\neq S_k, k=0,\dots,m \\ 1\leq i\leq j}} D_i(p_i^*) - D_i(p_i^\pi) \right| \\
\leqslant &\overline{p}\overline{d}(H-m-1) - \sum_{i\neq S_k, k=0,1,\dots,m} \overline{p}\epsilon_{ti} + 2L\overline{d}(H-m-1),
\end{aligned}$$

where the first inequality is from the upper bounds of demand and price and the Lipschitz continuity of the function $G(\cdot)$, and the second inequality is from the upper bound of demand. The statement is obtained by taking a summation over t and taking the expectation of the noises. $\square$

Upper Bound for $\sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S)$

Lemma EC.6. $\sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S) \leqslant \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right)$.

To prove Lemma EC.6, we need to establish similar technical results as in Sections 6.2.1 and 6.2.2. As the proofs are similar, we only present the statements.

Lemma EC.7. *Fix arbitrary failure probability* $\delta \in (0,1)$. *Suppose* $\gamma_{k,j} \geqslant \sqrt{j - S_{k-1}+1}\left[\sigma\sqrt{l\log\left(\frac{1+TL^2}{\delta/H}\right)} + L\right]$. *Then with probability* $1-\delta$ *uniformly over* $t\in[T]$, $k\in[m+1]$, *and* $j = S_{k-1},\dots,S_k - 1$, *it holds that* $\overline{D_t^{k,j}}(p) \geqslant D_t^{k,j}(p)$ *and* $\overline{D_t^{k,j}}(p) - D_t^{k,j}(p) \leqslant 2\Delta_{tkj}(p)$ *for any* $p\in\mathcal{P}$.

Lemma EC.8. *For any* $t\in[T]$, $k\in[m+1]$, *and* $j = S_{k-1},\dots,S_k-1$, *it holds that*

$$\sum_{s=1}^{t}\Delta_{skj}(p_{sS_{k-1}}) \leqslant L\gamma_{k,j}\sqrt{2l\log(S_k - S_{k-1}+tL^2)}\sqrt{t}.$$

Lemma EC.9. *If* $\alpha_{k,j} \geqslant 2L\gamma_{k,j}\sqrt{2l\log(S_k - S_{k-1} + tL^2)}$, *then* $|\mathcal{H}_{<t}^{k,j}| \geqslant t/2$.

Lemma EC.10. *Let* A *be the event that* $|\eta(p)\cdot(\hat{\theta}_t^{k,j} - \theta^{k,j})| \leqslant \Delta_{tkj}(p)$ *for all* $t\in[T], k\in[m+1], p\in\mathcal{P}$ *and* $j = S_{k-1},\dots,S_k-1$. *There exists an absolute constant* C *such that for any* $t\in[T], k\in[m+1]$ *and* $j = S_{k-1},\dots,S_k-1$, *we have*

$$\mathbb{E}[W_1(\mu_t^{k,j},\hat{\mu}_t^{k,j})|A] \leqslant \frac{C\sigma}{1-\delta}\sqrt{\frac{j-S_{k-1}+1}{t}} + \frac{\alpha_{k,j}}{\sqrt{t}}.$$

The following are error propagation results for the dynamic program (EC.4).

Lemma EC.11. *For any two initial states* y, y' *and* $k\in[N+1]$, *we have*

$$|u_k(y) - u_k(y')| \leqslant \frac{L(H+1-S_{k-1})}{H}|y-y'|.$$

For any pricing policies $p_1(\cdot),...,p_N(\cdot)$, we can define value-to-go functions following these policies by

$$v_k(y) = \mathbb{E}[R_k(p_k(y)) - \sum_{j=S_{k-1}}^{S_k-1} G(y - D^{k,j}(p_k(y)) - \beta^{k,j}) + v_{k+1}(y - D^{k,S_k-1}(p_k(y)) - \beta^{k,S_k-1})],\ k=1,...,N,$$

$$v_{N+1}(\cdot) = 0. \tag{EC.6}$$

LEMMA EC.12. *Let $p_1^*(\cdot),...p_N^*(\cdot)$ be optimal policies obtained from the dynamic program* (EC.4). *Let $\hat{\beta}^{k,j}, k \in [N], j = S_{k-1}, \ldots, S_k - 1$ be independent random variables. Denote $\hat{v}_k(y), k \in [N+1]$ the value-to-go functions defined in* (EC.6) *with noises $\hat{\beta}^{k,j}$ and following $p_1^*(\cdot),...,p_N^*(\cdot)$. Then for any state y and $k \in [N+1]$,*

$$|u_k(y) - \hat{v}_k(y)| \leqslant \sum_{i=k}^{N} \sum_{j=S_{i-1}}^{S_i - 1} \frac{L}{H} W_1(\beta^{i,j}, \hat{\beta}^{i,j}) + \sum_{i=k}^{N} \frac{L(H+1-S_i)}{H} W_1(\beta^{i,S_i-1}, \hat{\beta}^{i,S_i-1}).$$

LEMMA EC.13. *Let $p_1^*(\cdot),...p_N^*(\cdot)$ be optimal policies obtained from the dynamic program* (EC.4). *For $k \in [N]$ and $j = S_{k-1}, \ldots, S_k - 1$, let $\bar{D}^{k,j}, \bar{R}_k$ be functions of $p \in \mathcal{P}$. Let $\hat{\beta}^{k,j}, k \in [N], j = S_{k-1}, \ldots, S_k - 1$ be independent random variables. Let $\tilde{v}_k(y), k \in [N+1]$ be value-to-go functions defined in* (EC.6) *with $\bar{D}^{k,j}, \bar{R}_k, \hat{\beta}^{k,j}$ and following $p_1^*(\cdot),...p_N^*(\cdot)$. Then*

$$\begin{aligned}|u_k(y) - \tilde{v}_k(y)| \leqslant &\mathbb{E}\Bigg[\sum_{i=k}^{N} |R_i(p_i^*) - \bar{R}_i(p_i^*)| + \frac{L}{H} \sum_{i=k}^{N} \sum_{j=S_{i-1}}^{S_i-1} \left(|D^{i,j}(p_i^*) - \bar{D}^{i,j}(p_i^*)| + W_1(\beta^{i,j}, \hat{\beta}^{i,j})\right) \\ &+ \sum_{i=k}^{N} \frac{L(H+1-S_i)}{H} \left(|D^{i,S_i-1}(p_i^*) - \bar{D}^{i,S_i-1}(p_i^*)| + W_1(\beta^{i,S_i-1}, \hat{\beta}^{i,S_i-1})\right)\Bigg],\end{aligned}$$

where $p_i^* = p_i^*(y_i)$, $y_{i+1} = y_i - \bar{D}^{i,S_i-1}(p_i^*) - \hat{\beta}_i$ with $y_k = y$, and the expectation is taken for the randomness of p_i^*.

Proof of Lemma EC.6 Similar to the proof of Theorem 2, we only consider epochs whose initial inventory is no more than the target order-up-to level. Let A be the event that $|\eta(p) \cdot (\hat{\theta}_t^{k,j} - \theta^{k,j})| \leqslant \Delta_{tkj}(p)$ for all $t \in [T], k \in [m+1], j = S_{k-1}, \ldots, S_k - 1, p \in \mathcal{P}$. Again, $\mathbb{P}(A) \geqslant 1 - \delta$. Similar to the proof of Theorem 2, we have

$$\begin{aligned}&\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon)] \\ \leqslant &\mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid A] \\ &+ \mathbb{E}[g_t(y^S, p_1^S(\cdot), ..., p_H^S(\cdot)|D, \epsilon) - g_t(y_t, p_{t1}(\cdot), ..., p_{tH}(\cdot)|D, \epsilon) \mid \overline{A}]\mathbb{P}(\overline{A}) \\ \leqslant &\mathbb{E}\left[O(1) \sum_{k=1}^{m+1} \Delta_{\tau_t k, S_k - 1}(p_{tS_{k-1}}) + \frac{L}{H} \sum_{k=1}^{m+1} \sum_{j=S_{k-1}}^{S_k - 1} \Delta_{\tau_t kj}(p_{tS_{k-1}}) \middle| A\right] \\ &+ O(1) \sum_{k=1}^{m+1} \mathbb{E}[W_1(\mu_{\tau_t}^{k,S_k-1}, \hat{\mu}_{\tau_t}^{k,S_k-1}) \mid A] + 2 \sum_{k=1}^{m+1} \sum_{j=S_{k-1}}^{S_k-1} \frac{L}{H} \mathbb{E}[W_1(\mu_{\tau_t}^{k,j}, \hat{\mu}_{\tau_t}^{k,j}) \mid A] + \delta H \overline{p} \bar{d}.\end{aligned}$$

Set $\delta = \frac{1}{TH}$. For $k \in [m+1], j = S_{k-1}, \ldots, S_k - 1$, with probability 1,

$$\sum_{t=1}^{T} \Delta_{\tau_t kj}(p_{tS_{k-1}}) = \tilde{O}\left(\sqrt{j - S_{k-1} + 1}\sqrt{T}\right).$$

Hence, $\sum_{t=1}^{T}\sum_{k=1}^{m+1} \Delta_{\tau_t k, S_k - 1}(p_{tS_{k-1}}) = \tilde{O}(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T})$ and

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{L}{H} \sum_{k=1}^{m+1} \sum_{j=S_{k-1}}^{S_k - 1} \Delta_{\tau_t kj}(p_{tS_{k-1}}) = \tilde{O}\left(\frac{1}{H}\sum_{k=1}^{m+1}\sum_{j=S_{k-1}}^{S_k-1}\sqrt{j - S_{k-1} + 1}\sqrt{T}\right) \\
=&\tilde{O}\left(\frac{1}{H}\sum_{k=1}^{m+1}(S_k - S_{k-1})^{\frac{3}{2}}\sqrt{T}\right) = \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right),
\end{aligned}$$

where the last equality is from the fact that $\sum_{k=1}^{m+1}(S_k - S_{k-1}) = H$. Similarly, we have

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{k=1}^{m+1} \mathbb{E}[W_1(\mu_{\tau_t}^{k,S_k-1}, \hat{\mu}_{\tau_t}^{k,S_k-1}) \mid A] \\
\leqslant &\sum_{k=1}^{m+1}\sum_{t=1}^{T} \frac{C\sigma}{1 - 1/(TH)}\sqrt{\frac{S_k - S_{k-1}}{\tau_t}} + \sum_{k=1}^{m+1}\sum_{t=1}^{T} \frac{2L\gamma_{k,S_k-1}\sqrt{2l\log(S_k - S_{k-1} + TL^2)}}{\sqrt{\tau_t}} \\
=&\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right),
\end{aligned}$$

and

$$\frac{L}{H}\sum_{t=1}^{T}\sum_{k=1}^{m+1}\sum_{j=S_{k-1}}^{S_k-1} \mathbb{E}[W_1(\mu_{\tau_t}^{k,j}, \hat{\mu}_{\tau_t}^{k,j}) \mid A] = \tilde{O}\left(\frac{1}{H}\sum_{k=1}^{m+1}(S_k - S_{k-1})^{\frac{3}{2}}\sqrt{T}\right) = \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right).$$

Combining the above inequalities, we have

$$\tilde{R}_T^I(\pi_S) \leqslant \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k - S_{k-1}}\sqrt{T}\right).$$

□

EC.4.2. Numerical Comparison of Different Cost-Accounting Schemes

In this section, we numerically compare the optimal expected profit during one epoch under two inventory cost-accounting schemes. For a given inventory level y, we denote $u_1^{PE}(y)$ as the optimal expected profit during one epoch under the cost-accounting scheme where inventory cost is accounted only at the end of each epoch (per-epoch cost-accounting scheme), and $u_1^{PP}(y)$ as the optimal expected profit during one epoch under the cost-accounting scheme where inventory cost is accounted every period (per-period cost-accounting scheme). The inventory cost $G(x) = h^+x^+ + h^-x^-$ is incurred at the end of the epoch for the per-epoch cost-accounting scheme and the inventory cost $G(x) = \frac{h^+}{H}x^+ + \frac{h^-}{H}x^-$ is incurred at the end of each period for the per-period cost-accounting scheme. We use similar demand functions and noise distributions as well as parameter settings as in Section 8. Specifically, we consider linear demand functions $D_{ti}(p) = 30 - 20p$ for any $i = 1, \dots, H$. The range of prices is $\mathcal{P} = [0, 1]$. We set $h^+ = 0.05$ and $h^- = 1$. The random noises ϵ_{ti} are generated by the normal distribution with mean zero and a variance of four. Price can be changed at every period. Figure EC.1 shows the curves of average profits $u_1^{PE}(y)/H$ and $u_1^{PP}(y)/H$ for $H \in \{10, 100\}$.

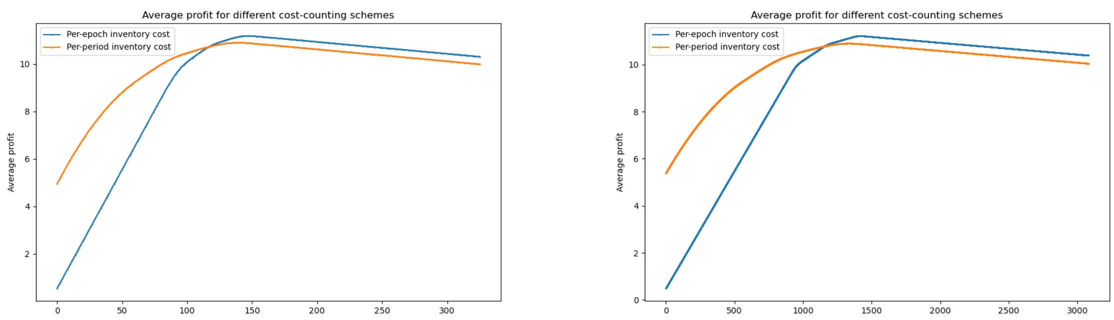


Figure EC.1 $H = 10$ **(left) and** $H = 100$ **(right).**

One can see from Figure EC.1 that the two curves are very close to each other near their peaks. Let $y^{PE} \in \arg\max_{y \geqslant 0} u_1^{PE}(y)$ and $y^{PP} \in \arg\max_{y \geqslant 0} u_1^{PP}(y)$. Table EC.1 shows that the relative gap of the two curves $u_1^{PE}(y)$ and $u_1^{PP}(y)$ at their peaks y^{PE} and y^{PP} are quite small for both small $H = 10$ and large $H = 100$. This numerically shows that our regret bounds are robust to the inventory cost-accounting schemes.

H	$\frac{u_1^{PE}(y^{PP}) - u_1^{PP}(y^{PP})}{u_1^{PP}(y^{PP})}$	$\frac{u_1^{PE}(y^{PE}) - u_1^{PP}(y^{PE})}{u_1^{PE}(y^{PE})}$
10	2.68%	2.69%
100	3.02%	2.08%

Table EC.1 Relative gap at the peaks

EC.4.3. Account for Inventory Cost Right Before a Price Change

We now consider the case where the inventory cost is accounted for right before a price change. Specifically, fix a price change scheme $S=(S_0,S_1,\dots,S_{m+1})\in\mathbb{Z}^{m+2}$ such that $1=S_0<S_1<S_2<\cdots<S_m<S_{m+1}=H+1$. We assume that at the end of period S_k-1 of epoch t, for $k\in[m+1]$ and $t\in[T]$, a holding and backlogging cost $G_k(\cdot)$, depending on the current inventory level, is incurred. We continue to impose Assumption 1 and a slightly modified Assumption 2, defined below.

ASSUMPTION EC.3. *Assumption 2 holds except that Assumption 2(e) is replaced by the following: for any $k\in[m+1]$, $G_k(\cdot)$ is $L(S_k-S_{k-1})/H$-Lipschitz continuous.*

This assumption requires that the inventory cost function at the end of a price interval scales with the length of the period(s) during which the price is in effect.

THEOREM EC.2. *Suppose Assumptions 1 and EC.3 hold.*

(a) (Lower bound) There is a constant $c>0$ such that

$$\inf_{\pi\in\Pi^S}\sup_{I\in\mathcal{I}} R_T^I(\pi)\geq\frac{(H-m-1)T}{32}+c\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T};$$

(b) (Upper bound)

$$\sup_{I\in\mathcal{I}} R_T^I(\pi_S)\leq\ O((H-m-1)T)+\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right),$$

where π_S is the policy of Algorithm 2.

Note that in Algorithm 2, the dynamic program solved is slightly modified to include the inventory cost G_k (see equation (EC.8)). The proof of the lower bound is the same as that in Section 5. The proof of the upper bound is similar to that in Section 6. By the decomposition $R_T^I(\pi_S)=G_T^I(S)+\tilde{R}_T^I(\pi_S)$, we have

$$\sup_{I\in\mathcal{I}} R_T^I(\pi_S)\leqslant\sup_{I\in\mathcal{I}} G_T^I(S)+\sup_{I\in\mathcal{I}}\tilde{R}_T^I(\pi_S). \tag{EC.7}$$

Note that the first term $\sup_{I\in\mathcal{I}} G_T^I(S)$ is the profit gap due to less pricing flexibility under the full-information setting and the second term $\sup_{I\in\mathcal{I}}\tilde{R}_T^I(\pi_S)$ is the profit gap due to lack of knowledge of the demand-price function as well as the noise distribution.

Upper Bound for $\sup_{I\in\mathcal{I}} G_T^I(S)$

Lemma EC.14. *It holds that*

$$\sup_{I\in\mathcal{I}} G_T^I(S) \leqslant (\bar{p}\bar{d} + 2L\bar{d})(H-m-1)T.$$

Proof: Fix a problem instance I. Consider the policy π constructed in the proof of Lemma 2. The profit gap in epoch t between the optimal policy and π under the same realization of random noises is

$$\begin{aligned}
&\sum_{i\neq S_k, k=0,1,\dots,m} [p_i^* D_i(p_i^*) - p_i^\pi D_i(p_i^\pi)] - \sum_{k=1}^{m+1}\left[G_k\left(y^* - \sum_{i=1}^{S_k-1} D_i(p_i^*) - \sum_{i=1}^{S_k-1}\epsilon_{ti}\right) - G_k\left(y^* - \sum_{i=1}^{S_k-1} D_i(p_i^\pi) - \sum_{i=1}^{S_k-1}\epsilon_{ti}\right)\right]\\
\leqslant &\sum_{i\neq S_k, k=0,1,\dots,m} \bar{p}(\bar{d}-\epsilon_{ti}) + \sum_{k=1}^{m+1}\frac{L(S_k-S_{k-1})}{H}\left|\sum_{\substack{i\neq S_j, j=0,\dots,m\\ 1\leq i\leq S_k-1}} D_i(p_i^*) - D_i(p_i^\pi)\right|\\
\leqslant &\bar{p}\bar{d}(H-m-1) - \sum_{i\neq S_k, k=0,1,\dots,m}\bar{p}\epsilon_{ti} + 2L\bar{d}(H-m-1),
\end{aligned}$$

where the first inequality is from the upper bounds of demand and price and the Lipschitz continuity of the function $G_k(\cdot)$, and the second inequality is from the upper bound of demand. The statement is obtained by taking a summation over t and taking the expectation of the noises. □

Upper Bound for $\sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S)$

Lemma EC.15. $\sup_{I\in\mathcal{I}} \tilde{R}_T^I(\pi_S) \leqslant \tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right)$.

To prove Lemma EC.15, we need to establish similar technical results as in Section 6.2.2. Results on statistical error bound in Section 6.2.1 still hold in this setting.

Error Propagation Through Dynamic Program. This subsection provides general stability results for the following generic dynamic program,

$$\begin{aligned}
u_k(y) &= \max_{p\in\mathcal{P}} \mathbb{E}[R_k(p) - G_k(y - D_k(p) - \beta_k) + u_{k+1}(y - D_k(p) - \beta_k)],\ k = 1,\dots,N,\\
u_{N+1}(\cdot) &= 0,
\end{aligned} \tag{EC.8}$$

where $\mathcal{P}\subseteq\mathbb{R}_+$ is the feasible set of a generic control p, $\{\beta_1,\dots,\beta_N\}$ are generic independent random variables, R_k, D_k, G_k are generic univariate functions, and the expectation is taken with respect to $\beta_1,\dots,\beta_N$. Because all of the following lemmas can be proved with similar arguments in the proofs of the corresponding lemmas in Section 6.2.2, we omit their proofs.

LEMMA EC.16. *If $G_k(\cdot)$ is L_k-Lipschitz continuous, $k \in [N]$, then for any two initial states y, y' and $k \in [N+1]$, we have*

$$|u_k(y) - u_k(y')| \leqslant L_{k:N}|y - y'|,$$

where $L_{k:N} = \sum_{i=k}^{N} L_i$ for $k = 1, ..., N$.

LEMMA EC.17. *Let $\bar{R}_k(\cdot), \bar{D}_k(\cdot)$ be functions of $p \in \mathcal{P}$ such that for any $p \in \mathcal{P}$ there exists p' with $\bar{D}_k(p') = D_k(p)$ and $\bar{R}_k(p') \geqslant R_k(p)$, $k \in [N]$. Let $\bar{u}_k(y)$ be defined in dynamic program* (EC.8) *with $\bar{R}_k, \bar{D}_k$. Then for any $k \in [N+1]$,*

$$\bar{u}_k(y) \geqslant u_k(y).$$

For any pricing policies $p_1(\cdot), ..., p_N(\cdot)$, we can define value-to-go functions following these policies by

$$\begin{aligned} &v_k(y) = \mathbb{E}[R_k(p_k(y)) - G_k(y - D_k(p_k(y)) - \beta_k) + v_{k+1}(y - D_k(p_k(y)) - \beta_k)], \ k = 1, ..., N, \\ &v_{N+1}(\cdot) = 0. \end{aligned} \tag{EC.9}$$

LEMMA EC.18. *Suppose $G_k(\cdot)$ is L_k-Lipschitz continuous, $k \in [N]$. Let $p_1^*(\cdot), ...p_N^*(\cdot)$ be optimal policies obtained from dynamic program* (EC.8)*. Let $\hat{\beta}_k, k \in [N]$ be independent random variables. Denote $\hat{v}_k(y), k \in [N+1]$ the value-to-go functions defined in* (EC.9) *with noises $\hat{\beta}_k, k \in [N]$ and following $p_1^*(\cdot), ...p_N^*(\cdot)$. Then for any state y and $k \in [N+1]$,*

$$|u_k(y) - \hat{v}_k(y)| \leqslant \sum_{i=k}^{N} (L_i + L_{i:N}) W_1(\beta_i, \hat{\beta}_i).$$

LEMMA EC.19. *Suppose $G_k(\cdot)$ is L_k-Lipschitz continuous, $k \in [N]$. Let $p_1^*(\cdot), ...p_N^*(\cdot)$ be optimal policies obtained from dynamic program* (EC.8)*. Let $\bar{D}_k, \bar{R}_k$ be functions of $p \in \mathcal{P}$ and $\hat{\beta}_k, k \in [N]$ be independent noises. Let $\tilde{v}_k(y), k \in [N+1]$ be value-to-go functions defined in* (EC.9) *with $\bar{D}_k, \bar{R}_k, \hat{\beta}_k, k \in [N]$ and following $p_1^*(\cdot), ...p_N^*(\cdot)$. Then*

$$|u_k(y) - \tilde{v}_k(y)| \leqslant \mathbb{E}\left[\sum_{i=k}^{N} |R_i(p_i^*) - \bar{R}_i(p_i^*)| + L_{i:N}|D_i(p_i^*) - \bar{D}_i(p_i^*)| + L_{i:N} W_1(\beta_i, \hat{\beta}_i)\right],$$

where $p_i^* = p_i^*(y_i)$, $y_{i+1} = y_i - \bar{D}_i(p_i^*) - \hat{\beta}_i$ with $y_k = y$, and the expectation is taken for the randomness of p_i^*.

Proof of Lemma EC.15 Similar to the proof of Theorem 2, we only consider epochs whose initial inventory is no more than the target order-up-to level. Let A be the event that $|\eta(p)\cdot(\hat{\theta}_t^k-\theta^k)|\leqslant \Delta_{tk}(p)$ for all $t\in[T], k\in[m+1], p\in\mathcal{P}$. By the proof of Lemma 4, $\mathbb{P}(A)\geqslant 1-\delta$.

Denote $L_k=L(S_k-S_{k-1})/H, k\in[m+1]$. Similar to the proof of Theorem 2, we have

$$\begin{aligned}
&\mathbb{E}[g_t(y^S,p_1^S(\cdot),...,p_H^S(\cdot)|D,\epsilon)-g_t(y_t,p_{t1}(\cdot),...,p_{tH}(\cdot)|D,\epsilon)]\\
\leqslant&\mathbb{E}[g_t(y^S,p_1^S(\cdot),...,p_H^S(\cdot)|D,\epsilon)-g_t(y_t,p_{t1}(\cdot),...,p_{tH}(\cdot)|D,\epsilon)\mid A]\\
&+\mathbb{E}[g_t(y^S,p_1^S(\cdot),...,p_H^S(\cdot)|D,\epsilon)-g_t(y_t,p_{t1}(\cdot),...,p_{tH}(\cdot)|D,\epsilon)\mid \overline{A}]\mathbb{P}(\overline{A})\\
\leqslant&\mathbb{E}\left[\sum_{k=1}^{m+1}(2\bar{p}+L_{k:m+1})\Delta_{\tau_t k}(p_{tS_{k-1}})\middle|A\right]+\sum_{k=1}^{m+1}(L_k+2L_{k:m+1})\mathbb{E}[W_1(\mu_{\tau_t}^k,\hat{\mu}_{\tau_t}^k)\mid A]+\delta H\bar{p}\bar{d}.
\end{aligned}$$

Set $\delta=\frac{1}{TH}$. By Lemma 5 and Corollary EC.1, with probability 1,

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{k=1}^{m+1}(2\bar{p}+L_{k:m+1})\Delta_{\tau_t k}(p_{tS_{k-1}})=&\sum_{k=1}^{m+1}(2\bar{p}+L_{k:m+1})\sum_{t=1}^{T}\Delta_{\tau_t k}(p_{tS_{k-1}})\\
\leq&\sqrt{2}\sum_{k=1}^{m+1}(2\bar{p}+L_{k:m+1})L\gamma_k\sqrt{2l\log(S_k-S_{k-1}+TL^2)}\sqrt{T}\\
=&\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right),
\end{aligned}$$

where the last equality is from $L_{k:m+1}=\sum_{j=k}^{m+1}L_j=\sum_{j=k}^{m+1}L(S_j-S_{j-1})/H=L(S_{m+1}-S_{k-1})/H\leq L=O(1)$.

By Lemma 7 and Lemma EC.3,

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{k=1}^{m+1}(L_k+2L_{k:m+1})\mathbb{E}\left[W_1(\mu_{\tau_t}^k,\hat{\mu}_{\tau_t}^k)|A\right]\\
\leqslant&\sum_{k=1}^{m+1}(L_k+2L_{k:m+1})\sum_{t=1}^{T}\frac{C\sigma}{1-1/(TH)}\sqrt{\frac{S_k-S_{k-1}}{\tau_t}}+\sum_{k=1}^{m+1}(L_k+2L_{k:m+1})\sum_{t=1}^{T}\frac{2L\gamma_k\sqrt{2l\log(S_k-S_{k-1}+TL^2)}}{\sqrt{\tau_t}}\\
=&\tilde{O}\left(\sum_{k=1}^{m+1}(L_k+2L_{k:m+1})\sqrt{S_k-S_{k-1}}\sqrt{T}\right)=\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right).
\end{aligned}$$

Combine the above inequalities, we have

$$\tilde{R}_T^I(\pi_S)\leqslant\tilde{O}\left(\sum_{k=1}^{m+1}\sqrt{S_k-S_{k-1}}\sqrt{T}\right).$$

□

Appendix EC.5: Explore-Then-Exploit Algorithm for Numerical Studies

This section presents the explore-then-exploit heuristic used for performance comparison in our numerical experiment. Please see Algorithm 3.

Algorithm 3: Explore-Then-Exploit for Numerical Studies

Set $\gamma_k = \sqrt{S_k - S_{k-1}}\left[\sigma\sqrt{l\log\left((1+TL^2)TH(m+1)\right)} + L\right]$ and $\alpha_k = 2L\gamma_k\sqrt{2l\log(S_k - S_{k-1} + TL^2)}$, $k \in [m+1]$; the exploration length $T_0 = \lceil T^{1/2}\rceil$ or $\lceil T^{2/3}\rceil$; set the order-up-to level in exploration phase by $y_0 = 0$

for *epoch* $t = 1, \ldots, T_0$ **do**

  Raise the inventory to y_0

  for *period* $i = 1, \ldots, H$ **do**

    if $i = S_{k-1}$ *for some* $k \in [m+1]$ **then**

      Randomly sample the price following the uniform distribution over $[\underline{p}, \overline{p}]$;

    else

      Implement the price of period $i-1$

Calculate $\overline{D^k_{T_0+1}}(p)$ according to (4) for all $p \in \mathcal{P}$, $k \in [m+1]$

Construct empirical distribution $\hat{\mu}^k_{T_0+1}$ according to (5), $k \in [m+1]$

Solve the dynamic program (1) with $N = m+1$, $R_k(p) = p\overline{D^k_{T_0+1}}(p)$, $D_k(p) = \overline{D^k_{T_0+1}}(p)$, $\beta_k \sim \hat{\mu}^k_{T_0+1}$, $k \in [m+1]$, and G being the inventory cost function

Let $p_k(\cdot),\ k \in [m+1]$ be the pricing policies from solving the dynamic program and $y_{T_0+1} \in \arg\max_{y \in \mathcal{Y}} u_1(y)$

for *epoch* $t = T_0+1, \ldots, T$ **do**

  Raise the inventory to y_{T_0+1} and implement pricing policies $p_k(\cdot),\ k \in [m+1]$ under the price change scheme S

Appendix EC.6: More Numerical Experiments

In this section, we test our algorithm for epoch-wise nonstationary demand functions and show that our algorithm is still efficient when the estimated demand functions (4) are constructed without the knowledge of $D_{ti}(\bar{p})$.

EC.6.1. Epoch-Wise Nonstationary Demand

Online decision-making in a nonstationary environment is challenging as one has to update the estimates of the changing parameters (or even functions) frequently to make good decisions. Although our algorithm is not designed for this setting, we still test its performance when the demand functions are nonstationary over the epochs.

We consider demand functions $D_{ti}(p) = a_i + b_i(t)p,\ i \in [H]$, where the price sensitivity changes over epochs. The price range is $[0,1]$. Following Besbes et al. (2015a), we define the total variation of functions $D_{ti}(\cdot), t \in [T]$ by

$$V = \sum_{t=2}^{T} \max_{p\in[0,1]} |D_{ti}(p) - D_{t+1,i}(p)| = \sum_{t=2}^{T} |b_i(t+1) - b_i(t)|.$$

Following Cheung et al. (2023), we consider

$$b_i(t) = -\left(5 + \lambda \sin\left(\frac{V\pi t}{2T}\right)\right), t = 1, ..., T$$

One can see that the total variation of the above sequence is approximately λV. We set $H = 20, m = 10,\ a_i = 10, i \in [H]$, and $b_i(t)$ follows the above formula with $\lambda = 4, T = 2^{12}$. Figure EC.2 shows the average profit of Algorithm 2 for different total variations of demand functions $V = \log_2(T), T^{0.2}, T^{0.5}$. Not surprisingly, for all the cases, our algorithm cannot converge to the oracle solution that knows the demand functions and noise distributions.

If the demand functions change across epochs, then as illustrated in the numerical results, our learning algorithm and regret guarantees cannot be directly applied, but they can be a very useful starting point. One possible solution is to adopt the restarting procedure analyzed in Besbes et al. (2015b). The restarting procedure takes both an online learning algorithm $\mathscr{A}$ and a batch size Δ_T (which is a function of the planning horizon T and the variation budget) as input. For every Δ_T

epochs, the procedure restarts the algorithm $\mathscr{A}$. Besbes et al. (2015b) proves that when either the gradient feedback or bandit feedback is available, the restarting structure guarantees tight regret. Our problem in a nonstatonary environment will be very different and probably much more complex than the setting studied in Besbes et al. (2015b), due to the required analyses of the DP recursions. But we believe the restarting procedure is one promising direction to try, and our proposed learning algorithm can be an excellent candidate for $\mathscr{A}$. Completely solving the nonstationary setting will require some modifications and different analyses, therefore it is beyond the scope of this paper and requires a separate paper. This is an interesting and highly relevant future research topic to explore. For related discussions, see Chen and Shi (2019).

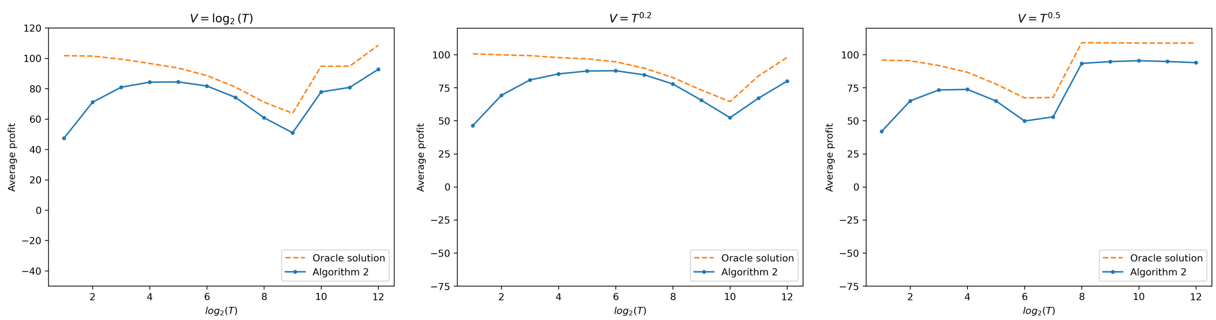


Figure EC.2 Average profit of Algorithm 2 for different total variations of demand functions

EC.6.2. Algorithm 2 with Different UCB Constructions

In the design of our algorithm, we assume $D_{ti}(\bar{p})$ is known (see Assumption 2(f)) when constructing the estimated demand functions (4). In this section, we propose two alternative constructions of the estimated demand functions without this assumption. The first alternative is

$$\overline{D_t^k}(p) = \min\left\{\eta(p)^\top \hat{\theta}_t^k + \Delta_{tk}(p), \sum_{i=S_{k-1}}^{S_k-1} \underline{d}_i + (S_k - S_{k-1})L^2 \|\bar{p} - p\|_2\right\}, \tag{EC.10}$$

where $\underline{d}_i$ is a lower bound of the demand in period i, $i \in [H]$. In other words, we replace $D_{ti}(\bar{p})$ with a lower bound of the demand. The second alternative is

$$\overline{D_t^k}(p) = \eta(p)^\top \hat{\theta}_t^k + \Delta_{tk}(p). \tag{EC.11}$$

This alternative simply drops the term involving the lower bound of demand.

We first test the impact of the tightness of lower bounds $\underline{d}_i$ on the average profit of our algorithm. We consider a lower bound $\underline{d}_i = D_{ti}(\bar{p}) - \lambda\sigma$, where σ is the standard deviation of the random noise and the multiplier $\lambda \in [0, \infty)$ measures the tightness of the lower bound (the smaller the better). From the property of subgaussian random variables, the realized demand is greater than this lower bound with a probability larger than $1 - \exp(-\lambda^2/2)$ (e.g., the probability is 0.864 when $\lambda = 2$ and is 0.988 when $\lambda = 3$). Set $T = 2^{12}, H = 5, m = 3, D_{ti}(p) = 30 - 5p, i \in [H]$. Figure EC.3 shows the average profit of Algorithm 2 using the first UCB alternative with different levels of tightness of the lower bound. When the lower bound becomes looser (i.e., when the multiplier λ becomes large), the average profit becomes smaller and does not change when it is too loose. This is because when the lower bound becomes looser, the second term of (EC.10) would be smaller than the first term for some price points, and this makes the estimated demand function inaccurate and thus results in profit loss. When the lower bound is too loose, the estimated demand function would be determined by the second term of (EC.10) for most price points, and thus the policy of Algorithm 2 and the average profit would not change. To obtain a tight lower bound, in our algorithm, one can first implement $\bar{p}$ in the first $O(\log(T))$ epochs to estimate $D_{ti}(\bar{p})$. The modified algorithm would have the same regret bound as that of Algorithm 2.

We then compare the performance of Algorithm 2 with different UCB constructions. In addition to the above two alternatives, we consider Algorithm 2 with an additional demand cap term $\sum_{i=S_{k-1}}^{S_k-1} \bar{d}_i$ in the UCB construction (4), where $\bar{d}_i$ is an upper bound of the demand in period i, $i \in [H]$. We set $H = 5, m = 3, D_{ti}(p) = 30 - 20p, i \in [H]$. Figure EC.4 compares the average profit of Algorithm 2 using UCB with a demand cap, Algorithm 2, and Algorithm 2 using the second UCB alternative. The demand cap we use is $D_t^k(\underline{p})$ for illustration purposes. All the algorithms converge to the oracle solution. Moreover, incorporating a demand cap in the UCB improves the average profit at the early stage of the time horizon because it makes the demand function estimation more accurate when the first term in UCB (4) is large due to inadequate data collected at the early stage. Algorithm 2 converges faster than Algorithm 2 with UCB (EC.11) due to the same logic.

Appendix EC.7: Figures in Section 8

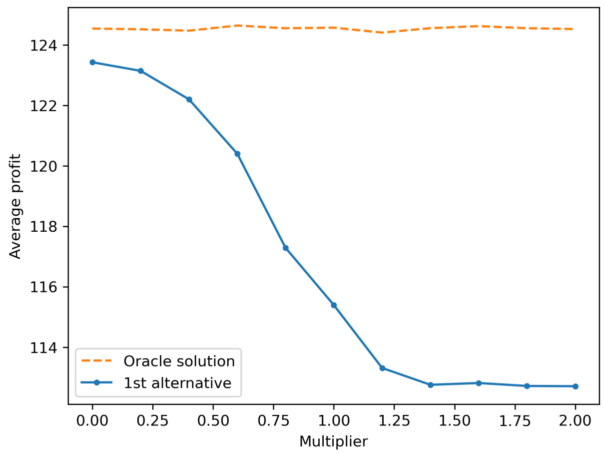


Figure EC.3 Average profit of Algorithm 2 with UCB construction (EC.10)

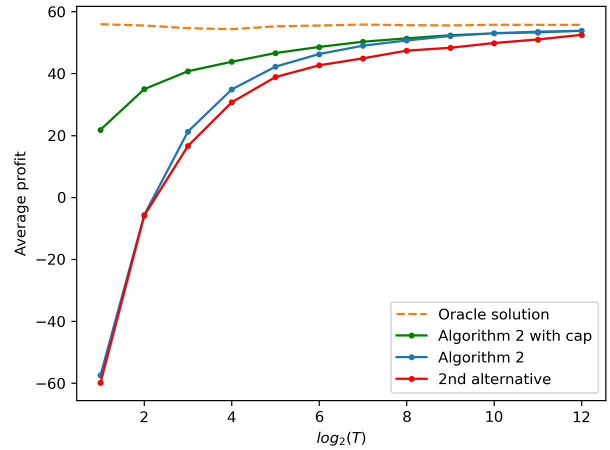
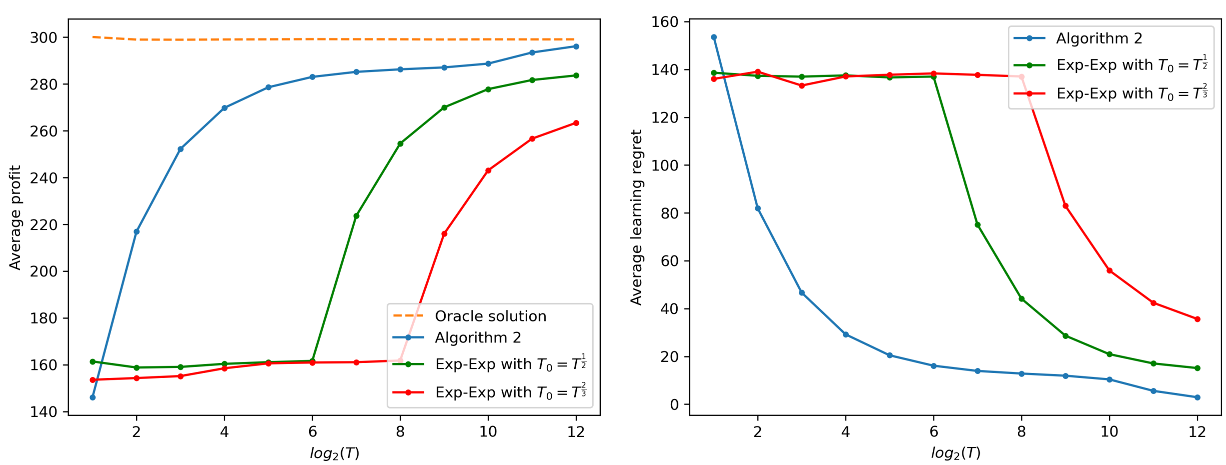


Figure EC.4 Average profit of Algorithm 2 with different UCB constructions



(a) Average profit

(b) Average learning regret

Figure EC.5 Average profit (left) and learning regret (right) of our algorithm and the baseline explore-then-exploit algorithm for different time horizons with $H = 20, m = 5$

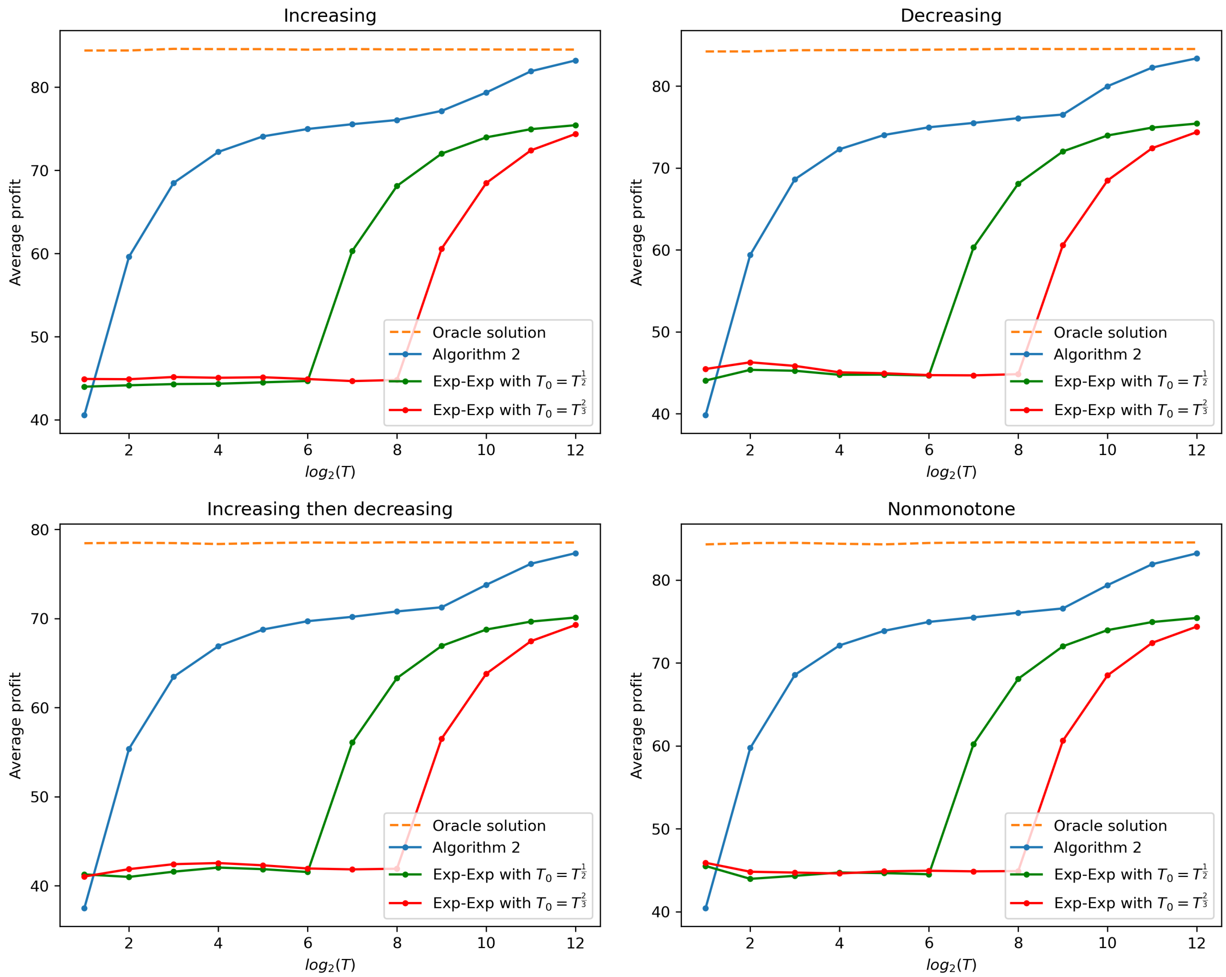


Figure EC.6 Average profit when demand parameters $\{a_i\}_{i\in[H]}$ are period-wise nonstationary. Increasing case $(a_i, b_i) = (19+i, -5), i \in [H]$; decreasing case $(a_i, b_i) = (25-i, -5), i \in [H]$; increasing-then-decreasing case $(a_1, \ldots, a_H) = (20, 21, 22, 21, 20)$ and $(b_1, \ldots, b_H) = (-5, \ldots, -5)$; and non-monotone case $(a_1, \ldots, a_H) = (20, 25, 20, 25, 20)$ and $(b_1, \ldots, b_H) = (-5, \ldots, -5)$. $H = 5, m = 3$.

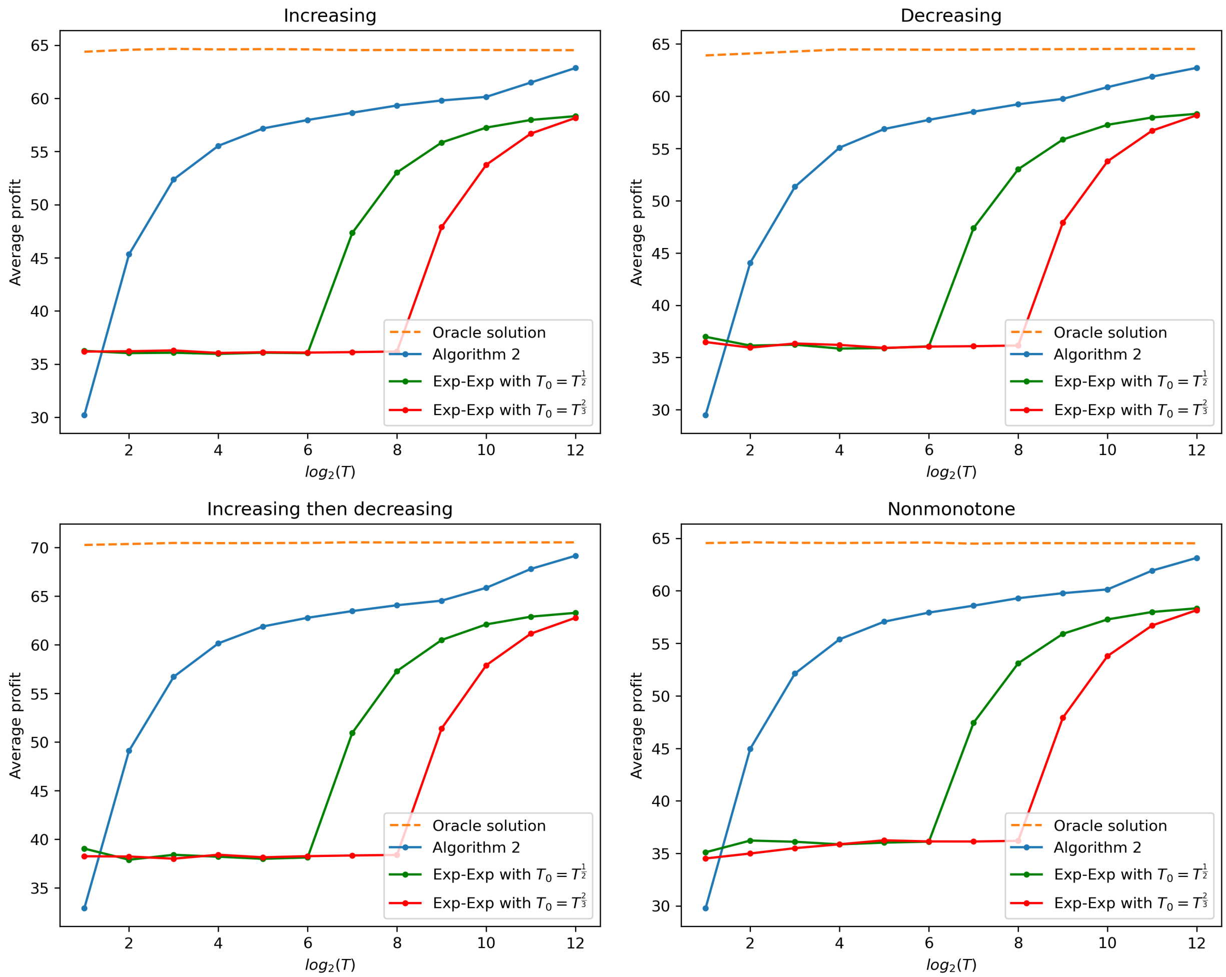


Figure EC.7 Average profit when price sensitivity parameters $\{b_i\}_{i\in[H]}$ are period-wise nonstationary. Increasing case $(a_i, b_i) = (20, -4-i), i \in [H]$; decreasing case $(a_i, b_i) = (20, -10+i), i \in [H]$; increasing-then-decreasing case $(a_1, \ldots, a_H) = (20, \ldots, 20)$ and $(b_1, \ldots, b_H) = (-5, -6, -7, -6, -5)$; and non-monotone case $(a_1, \ldots, a_H) = (20, \ldots, 20)$ and $(b_1, \ldots, b_H) = (-5, -10, -5, -10, -5)$. $H = 5, m = 3$.