Online Appendix – *Dynamic Pricing and Learning with Discounting*

Zhichao Feng

Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University

zhi-chao.feng@polyu.edu.hk

Milind Dawande, Ganesh Janakiraman, Anyan Qi

Naveen Jindal School of Management, The University of Texas at Dallas

milind@utdalla.edu, ganesh@utdallas.edu, axq140430@utdallas.edu

Appendix A Proofs of the Results in Section 2

A.1 Proof of Theorem 1

Preliminary Results

We first show some preliminary results, which are used in the proof of Theorem 1. In our analysis, we use a common quantitative measure of uncertainty known as the *KL divergence*.

Definition A.1 *(Definition 2.26 in Cover and Thomas 1999). For any probability measures Q_0 and Q_1 on a discrete sample space $\mathcal{Y}$, the KL divergence of Q_0 and Q_1 is*

$$\mathcal{K}(Q_0; Q_1) = \sum_{y \in \mathcal{Y}} Q_0(y) \log\left(\frac{Q_0(y)}{Q_1(y)}\right).$$

Broder and Rusmevichientong (2012) show the following properties of the problem class $\mathcal{C}_{LB}$ defined in the statement of Theorem 1, which are used to prove Lemmas A.2 and A.3 below.

Lemma A.1 *(Lemma EC.1.1 of Broder and Rusmevichientong 2012) For all $p \in \mathcal{P}$ and $z \in \mathcal{Z}$,*

1. $p^*(z) = \frac{1+2z}{4z}$.
2. $p^*(z_0) = 1$ *for* $z_0 = 1/2$.
3. $d(p^*(z_0); z) = 1/2$ *for all* $z \in \mathcal{Z}$.
4. $r(p^*(z); z) - r(p; z) \geq \frac{1}{3}(p^*(z) - p)^2$.
5. $|p^*(z) - p^*(z_0)| \geq \frac{1}{4}|z - z_0|$.
6. $|d(p; z) - d(p; z_0)| \leq |p^*(z_0) - p||z - z_0|$.

We recall that P_t^ψ is the random price in period t under policy ψ. For notational simplicity, we henceforth drop the superscript ψ from P_t^ψ.

Lemma A.2 *For $z_0 = 1/2$, $z \in \mathcal{Z}$, $T \geq 1$, and any policy ψ,*

$$\textit{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) \geq \frac{1}{16(z_0 - z)^2} \left[\rho^{T-1} \mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) + (1-\rho) \sum_{t=1}^{T-1} \rho^{t-1} \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z}) \right].$$

Proof of Lemma A.2: To show the lemma, we use the following Chain Rule for KL divergence (Theorem 2.5.3, Cover and Thomas 1999):

$$\mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) = \sum_{t=1}^{T} \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}),$$

where $\mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) := \sum_{\boldsymbol{y_t} \in \{0,1\}^t} Q_t^{\psi, z_0}(\boldsymbol{y_t}) \log \left(\frac{Q_t^{\psi, z_0}(y_t | \boldsymbol{y_{t-1}})}{Q_t^{\psi, z}(y_t | \boldsymbol{y_{t-1}})} \right)$ is the conditional KL divergence. Similar to Broder and Rusmevichientong (2012), we show that

$$\mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) \leq 16(z_0 - z)^2 \mathbb{E}[r(p^*(z_0); z_0) - r(P_t; z_0)].$$

Thus, we have

$$\begin{aligned}
&\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) \\
&= \sum_{t=1}^{T} \rho^{t-1} \mathbb{E}[r(p^*(z_0); z_0) - r(P_t; z_0)] \\
&\geq \sum_{t=1}^{T} \rho^{t-1} \frac{1}{16(z_0 - z)^2} \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) \\
&= \frac{1}{16(z_0 - z)^2} \left[\sum_{t=1}^{T} \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) - \sum_{t=2}^{T} (1-\rho) \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) - \right. \\
&\quad \left. \sum_{t=3}^{T} (\rho - \rho^2) \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) - \cdots - \sum_{t=T}^{T} (\rho^{T-2} - \rho^{T-1}) \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z} \Big| \boldsymbol{Y_{t-1}}) \right] \\
&= \frac{1}{16(z_0 - z)^2} \Big[\mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) - (1-\rho)(\mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) - \mathcal{K}(Q_1^{\psi, z_0}; Q_1^{\psi, z})) - \\
&\quad (\rho - \rho^2)(\mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) - \mathcal{K}(Q_2^{\psi, z_0}; Q_2^{\psi, z})) - \cdots - (\rho^{T-2} - \rho^{T-1})(\mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) - \mathcal{K}(Q_{T-1}^{\psi, z_0}; Q_{T-1}^{\psi, z})) \Big] \\
&= \frac{1}{16(z_0 - z)^2} \left[\rho^{T-1} \mathcal{K}(Q_T^{\psi, z_0}; Q_T^{\psi, z}) + (1-\rho) \sum_{t=1}^{T-1} \rho^{t-1} \mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z}) \right].
\end{aligned}$$

The third equality holds by the Chain Rule for KL-divergence. ■

Lemma A.3 *For $z_0 = 1/2$, $z \in \mathcal{Z}$, $T \geq 2$, and any policy ψ,*

$$\textit{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \textit{Regret}(z, \mathcal{C}_{LB}, T, \rho, \psi) \geq \frac{1}{6(12)^2} (z_0 - z)^2 \sum_{t=1}^{T-1} \rho^t e^{-\mathcal{K}(Q_t^{\psi, z_0}; Q_t^{\psi, z})}.$$

Proof of Lemma A.3 uses Lemma A.4 below.

Lemma A.4 *(Theorem 2.2, Tsybakov 2009) Let Q_0 and Q_1 be two probability distributions on a finite space $\mathcal{Y}$, with $Q_0(y), Q_1(y) > 0$ for all $y \in \mathcal{Y}$. Then for any function $J : \mathcal{Y} \to \{0, 1\}$,*

$$Q_0\{J = 1\} + Q_1\{J = 0\} \geq \frac{1}{2} e^{-\mathcal{K}(Q_0;Q_1)},$$

where $\mathcal{K}(Q_0; Q_1)$ denotes the KL divergence of Q_0 and Q_1.

Proof of Lemma A.3: We first define two intervals $C_{z_0} \subset \mathcal{P}$ and $C_z \subset \mathcal{P}$ by

$$C_{z_0} = \left\{ p : |p^*(z_0) - p| \leq \frac{1}{12}|z_0 - z| \right\} \text{ and } C_z = \left\{ p : |p^*(z) - p| \leq \frac{1}{12}|z_0 - z| \right\}.$$

By property 5 in Lemma A.1, i.e., $|p^*(z_0) - p^*(z)| \geq \frac{1}{4}|z_0 - z|$, C_{z_0} and C_z are disjoint. By property 4 in Lemma A.1, for each $\hat{z} \in \{z_0, z\}$, if $p \in \mathcal{P} \setminus C_{\hat{z}}$, then

$$r(p^*(\hat{z}); \hat{z}) - r(p; \hat{z}) \geq \frac{1}{3}(p - p^*(\hat{z}))^2 \geq \frac{1}{3(12)^2}(z_0 - z)^2.$$

Let $P_1, P_2, \cdots, P_T$ denote the sequence of random prices under policy ψ. Let $Pr_z\{A\}$ (resp., $Pr_{z_0}\{A\}$) denote the probability that event A occurs when the underlying parameter is z (resp., z_0). Then

$$\begin{aligned}
&\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z, \mathcal{C}_{LB}, T, \rho, \psi) \\
\geq\ & \sum_{t=1}^{T-1} \rho^t \mathbb{E}[r(p^*(z_0); z_0) - r(P_{t+1}; z_0)] + \sum_{t=1}^{T-1} \rho^t \mathbb{E}[r(p^*(z); z) - r(P_{t+1}; z)] \\
\geq\ & \frac{1}{3(12)^2}(z_0 - z)^2 \sum_{t=1}^{T-1} \rho^t \left(Pr_{z_0}\{P_{t+1} \notin C_{z_0}\} + Pr_z\{P_{t+1} \notin C_z\} \right) \\
\geq\ & \frac{1}{3(12)^2}(z_0 - z)^2 \sum_{t=1}^{T-1} \rho^t \left(Pr_{z_0}\{P_{t+1} \in C_z\} + Pr_z\{P_{t+1} \notin C_z\} \right) \\
\geq\ & \frac{1}{6(12)^2}(z_0 - z)^2 \sum_{t=1}^{T-1} \rho^t e^{-\mathcal{K}(Q_t^{\psi,z_0}; Q_t^{\psi,z})}.
\end{aligned}$$

The last inequality holds by Lemma A.4. ■

Proof of Theorem 1

Let $z_1 = z_0 + \left(\frac{1-\rho}{\rho}\right)^{1/4}$. Then for $\rho \geq \frac{16}{17}$, we have $z_1 \in \mathcal{Z}$. Using Lemmas A.2 and A.3, we have

$$\begin{aligned}
&2(\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_1, \mathcal{C}_{LB}, T, \rho, \psi)) \\
\geq\ & \text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + (\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_1, \mathcal{C}_{LB}, T, \rho, \psi)) \\
\geq\ & \frac{1}{16}\sqrt{\frac{\rho}{1-\rho}} \left[\rho^{T-1} \mathcal{K}(Q_T^{\psi,z_0}; Q_T^{\psi,z_1}) + (1-\rho) \sum_{t=1}^{T-1} \rho^{t-1} \mathcal{K}(Q_t^{\psi,z_0}; Q_t^{\psi,z_1}) \right] + \\
& \frac{1}{6(12)^2}\sqrt{\frac{1-\rho}{\rho}} \sum_{t=1}^{T-1} \rho^t e^{-\mathcal{K}(Q_t^{\psi,z_0}; Q_t^{\psi,z_1})}
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{1}{6(12)^2}\sqrt{(1-\rho)\rho}\left[\sum_{t=1}^{T-1}\rho^{t-1}\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_1})+\sum_{t=1}^{T-1}\rho^{t-1}e^{-\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_1})}\right]\\
&= \frac{1}{6(12)^2}\sqrt{(1-\rho)\rho}\sum_{t=1}^{T-1}\rho^{t-1}\left[\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_1})+e^{-\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_1})}\right]\\
&\geq \frac{1}{6(12)^2}\sqrt{(1-\rho)\rho}\sum_{t=1}^{T-1}\rho^{t-1}\\
&= \frac{1}{6(12)^2}\sqrt{\frac{\rho}{1-\rho}}(1-\rho^{T-1}). \qquad \text{(A-1)}
\end{aligned}
$$

The second inequality holds by Lemmas A.2 and A.3. The third inequality holds by the the fact that the KL divergence is nonnegative. The last inequality holds since $x+e^{-x}\geq 1$ for all $x\geq 0$.

Let $z_2 = z_0 + \left(\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}\right)^{1/4}$. Note that

$$\lim_{T\to\infty}\lim_{\rho\to 1}\left(\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}\right)^{1/4}=0.$$

Thus, there exists $\hat{\rho}\in(\frac{16}{17},1)$ and $\hat{T}\in\mathbb{N}$ such that for all $\rho\geq\hat{\rho}$ and $T\geq\hat{T}$, $\left(\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}\right)^{1/4}\leq 1/2$ and $z_2\in\mathcal{Z}$.

Note that for $z\in\mathcal{Z}$, $\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z})$ is non-decreasing in t because

$$\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z}) = \mathcal{K}(Q_{t-1}^{\psi,z_0};Q_{t-1}^{\psi,z})+\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z}\,\big|\,\boldsymbol{Y_{t-1}})\geq\mathcal{K}(Q_{t-1}^{\psi,z_0};Q_{t-1}^{\psi,z}).$$

The equality holds by the Chain Rule for KL divergence (Theorem 2.5.3, Cover and Thomas 1999). The inequality holds since $\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z}\,\big|\,\boldsymbol{Y_{t-1}})$ is nonnegative (Theorem 2.6.3, Cover and Thomas 1999).

Using Lemmas A.2 and A.3, we have

$$
\begin{aligned}
&2(\text{Regret}(z_0,\mathcal{C}_{LB},T,\rho,\psi)+\text{Regret}(z_2,\mathcal{C}_{LB},T,\rho,\psi))\\
&\geq \text{Regret}(z_0,\mathcal{C}_{LB},T,\rho,\psi)+(\text{Regret}(z_0,\mathcal{C}_{LB},T,\rho,\psi)+\text{Regret}(z_2,\mathcal{C}_{LB},T,\rho,\psi))\\
&\geq \frac{1}{16}\sqrt{\frac{1-\rho^{T-1}}{(1-\rho)\rho^{T-2}}}\left[\rho^{T-1}\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})+(1-\rho)\sum_{t=1}^{T-1}\rho^{t-1}\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_2})\right]+\\
&\quad\frac{1}{6(12)^2}\sqrt{\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}}\sum_{t=1}^{T-1}\rho^t e^{-\mathcal{K}(Q_t^{\psi,z_0};Q_t^{\psi,z_2})}\\
&\geq \frac{1}{16}\sqrt{\frac{1-\rho^{T-1}}{(1-\rho)\rho^{T-2}}}\rho^{T-1}\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})+\frac{1}{6(12)^2}\sqrt{\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}}\sum_{t=1}^{T-1}\rho^t e^{-\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})}\\
&= \frac{1}{16}\sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}}\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})+\frac{1}{6(12)^2}\sqrt{\frac{(1-\rho)\rho^{T-2}}{1-\rho^{T-1}}}\frac{\rho(1-\rho^{T-1})}{1-\rho}e^{-\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})}\\
&\geq \frac{1}{6(12)^2}\sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}}\left[\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})+e^{-\mathcal{K}(Q_T^{\psi,z_0};Q_T^{\psi,z_2})}\right]\\
&\geq \frac{1}{6(12)^2}\sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}}. \qquad \text{(A-2)}
\end{aligned}
$$

The second inequality holds by Lemmas A.2 and A.3. The third inequality holds by the the fact that the KL divergence is nonnegative (see Theorem 2.6.3 in Cover and Thomas 1999) and $\mathcal{K}(Q_t^{\psi,z_0}; Q_t^{\psi,z_2})$ is non-decreasing in t. The last inequality holds since $x + e^{-x} \geq 1$ for all $x \geq 0$.

Combining (A-1) and (A-2), we have

$$
\begin{aligned}
&2(\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_1, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_2, \mathcal{C}_{LB}, T, \rho, \psi)) \\
&\geq (\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_1, \mathcal{C}_{LB}, T, \rho, \psi)) + (\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_2, \mathcal{C}_{LB}, T, \rho, \psi)) \\
&\geq \frac{1}{(12)^3} \left(\sqrt{\frac{\rho}{1-\rho}}(1 - \rho^{T-1}) + \sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}} \right).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\max_{z \in \{z_0, z_1, z_2\}} \text{Regret}(z, \mathcal{C}_{LB}, T, \rho, \psi) \\
&\geq \frac{\text{Regret}(z_0, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_1, \mathcal{C}_{LB}, T, \rho, \psi) + \text{Regret}(z_2, \mathcal{C}_{LB}, T, \rho, \psi)}{3} \\
&\geq K_0 \left(\sqrt{\frac{\rho}{1-\rho}}(1 - \rho^{T-1}) + \sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}} \right),
\end{aligned}
$$

where $K_0 = \frac{1}{6(12)^3}$.

Let $f(\rho, T) = K_0 \left(\sqrt{\frac{\rho}{1-\rho}}(1 - \rho^{T-1}) + \sqrt{\frac{\rho^T(1-\rho^{T-1})}{1-\rho}} \right)$. Then we have

$$
\begin{aligned}
\lim_{\rho \to 1} f(\rho, T) &= K_0 \sqrt{T-1} = \Omega(\sqrt{T}), \text{ and} \\
\lim_{T \to \infty} f(\rho, T) &= K_0 \sqrt{\frac{\rho}{1-\rho}} = \Omega(\sqrt{\frac{1}{1-\rho}}).
\end{aligned}
$$

■

A.2 Proof of Theorem 2

Lemmas A.5 and A.6 below are used in the proof of Theorem 2. Similar to the proof of Lemma 3.7 in Broder and Rusmevichientong 2012), it is straightforward to obtain Lemma A.5 using the tail inequality for MLE based on IID Samples in Theore

Lemma A.5 ***(Mean-Squared Errors for MLE Based on IID Samples, Borovkov 1998)*** *For any $\tau \geq 1$, there exists a constant K_{mle} depending only on the exploration prices $\bar{\boldsymbol{p}}$ and the problem class $\mathcal{C}$ such that*

$$
\mathbb{E}[\|\boldsymbol{Z}(\tau) - \boldsymbol{z}\|]^2 \leq \frac{K_{mle}}{\tau}.
$$

Lemma A.6 is reproduced verbatim from Corollary 2.4 of Broder and Rusmevichientong (2012).

Lemma A.6 *For any problem class $\mathcal{C} = (\mathcal{P}, \mathcal{Z}, d)$ satisfying Assumption 1 and for any $\boldsymbol{z}, \hat{\boldsymbol{z}} \in \mathcal{Z}$,*

$$r(p^*(\boldsymbol{z}); \boldsymbol{z}) - r(p^*(\hat{\boldsymbol{z}}); \boldsymbol{z}) \leq c_r L^2 \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|^2.$$

Proof of Theorem 2: First, we show an upper bound on the regret incurred during the exploration phase. Recall from Assumption 1 that the revenue function is twice differentiable. In addition, the pricing interval $\mathcal{P}$ is compact; thus, there exists a constant K_1 depending only on the problem class $\mathcal{C}$ such that $r(p^*(\boldsymbol{z}); \boldsymbol{z}) - r(p; \boldsymbol{z}) \leq K_1$ for all $p \in \mathcal{P}$ and $\boldsymbol{z} \in \mathcal{Z}$. Thus, the regret incurred during the exploration phase satisfies

$$\sum_{s=1}^{\tau} \sum_{l=1}^{k} \rho^{(s-1)k+l-1} \mathbb{E}[r(p^*(\boldsymbol{z}); \boldsymbol{z}) - r(\bar{p}_l; \boldsymbol{z})] \leq \frac{1-\rho^{k\tau}}{1-\rho} K_1. \tag{A-3}$$

Next, we show an upper bound on the regret incurred during the exploitation phase. During the exploitation phase, we use price $p^*(\boldsymbol{Z}(\tau))$ and we offer this price for all $T - k\tau$ periods. It follows from Lemmas A.5 and A.6 that

$$\mathbb{E}\left[r(p^*(\boldsymbol{z}); \boldsymbol{z}) - r(p^*(\boldsymbol{Z}(\tau)); \boldsymbol{z})\right] \leq c_r L^2 \mathbb{E}\left[\|\boldsymbol{z} - \boldsymbol{Z}(\tau)\|^2\right] \leq c_r L^2 \frac{K_{mle}}{\tau}.$$

Thus, the regret incurred during the exploitation phase satisfies

$$\sum_{t=k\tau+1}^{T} \rho^{t-1} \mathbb{E}[r(p^*(\boldsymbol{z}); \boldsymbol{z}) - r(p^*(\boldsymbol{Z}(\tau)); \boldsymbol{z})] \leq \frac{\rho^{k\tau} - \rho^T}{(1-\rho)\tau} c_r L^2 K_{mle}. \tag{A-4}$$

Let $K_2 = c_r L^2 K_{mle}$. Combining (A-3) and (A-4), the cumulative regret under policy $\hat{\psi}$ satisfies

$$\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho, \hat{\psi}) \leq K_1 \frac{1-\rho^{k\tau}}{1-\rho} + K_2 \frac{\rho^{k\tau} - \rho^T}{(1-\rho)\tau}.$$

Let $g(\rho, T) = K_1 \frac{1-\rho^{k\tau}}{1-\rho} + K_2 \frac{\rho^{k\tau} - \rho^T}{(1-\rho)\tau}$, where $\tau = \left[\sqrt{\frac{1-\rho^T}{1-\rho}}\right]$. Then we have $\lim_{\rho \to 1} \tau = [\sqrt{T}]$ and

$$\lim_{\rho \to 1} g(\rho, T) = \lim_{\rho \to 1} \left(K_1 \frac{1-\rho^{k[\sqrt{T}]}}{1-\rho} + K_2 \frac{\rho^{k[\sqrt{T}]} - \rho^T}{(1-\rho)[\sqrt{T}]} \right) = K_1 k[\sqrt{T}] + K_2 \left(\frac{T}{[\sqrt{T}]} - k \right) = \mathcal{O}(\sqrt{T}).$$

Note that $\lim_{T \to \infty} \tau = [\sqrt{1/(1-\rho)}]$. Thus, we have

$$\lim_{T \to \infty} g(\rho, T) = K_1 \frac{1-\rho^{k[\sqrt{1/(1-\rho)}]}}{1-\rho} + K_2 \frac{\rho^{k[\sqrt{1/(1-\rho)}]}}{(1-\rho)[\sqrt{1/(1-\rho)}]} = \mathcal{O}\left(\sqrt{\frac{1}{1-\rho}}\right).$$

■

A.3 Proof of Theorem 3

Proof of Theorem 3: Note that the MLE-CYCLE policy $\check{\psi}$ operates in cycles. Broder and Rusmevichientong (2012) show that the regret incurred in each cycle is bounded from above by a constant, denoted by K_3. Note

that cycle h starts in period $\frac{h^2+h(2k-1)-2k+2}{2}$ and the total number of cycles is no more than $\lfloor\sqrt{2T}\rfloor$ for $T \geq 2$. Thus, we have

$$\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho, \check{\psi}) \leq \sum_{h=1}^{\lfloor\sqrt{2T}\rfloor} \rho^{[h^2+h(2k-1)-2k]/2} K_3 \leq \left(1 + \sum_{h=1}^{\lfloor\sqrt{2T}\rfloor} \rho^{h^2/2}\right) K_3.$$

and

$$\begin{aligned}\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho, \check{\psi}) \leq K_3 + \sum_{h=1}^{\lfloor\sqrt{2T}\rfloor} \rho^{h^2/2} K_3 \leq K_3 + K_3 \int_0^{\sqrt{2T}} \rho^{h^2/2} dh =& K_3 + K_3 \int_0^{\sqrt{2T}} e^{\log(\rho) h^2/2} dh \\ =& K_3 + K_3 \sqrt{\frac{1}{2}} \int_0^T e^{\log(\rho)x} \sqrt{\frac{1}{x}} dx.\end{aligned}$$

The last equality holds by letting $x = h^2/2$. When $T \to \infty$,

$$\begin{aligned}\lim_{T\to\infty} K_3 \left(1 + \sqrt{\frac{1}{2}} \int_0^T e^{\log(\rho)x} \sqrt{\frac{1}{x}} dx\right) &= K_3 \left(1 + \sqrt{\frac{1}{2}} \sqrt{\frac{1}{-\log\rho}} \int_0^\infty e^{-w} w^{-1/2} dw\right) \\ &= K_3 \left(1 + \sqrt{\frac{\pi}{2}} \sqrt{\frac{1}{-\log\rho}}\right) \\ &= \mathcal{O}\left(\sqrt{\frac{1}{-\log\rho}}\right) = \mathcal{O}\left(\sqrt{\frac{1}{1-\rho}}\right).\end{aligned}$$

The first equality holds by letting $w = -\log(\rho)x$. The second equality holds since $\int_0^\infty e^{-w} w^{-1/2} dw = \Gamma(1/2) = \sqrt{\pi}$. When $\rho \to 1$,

$$\lim_{\rho\to 1} \left(1 + \sum_{h=1}^{\lfloor\sqrt{2T}\rfloor} \rho^{h^2/2}\right) K_3 \leq K_3(\sqrt{2T} + 1) = \mathcal{O}\left(\sqrt{T}\right).$$

■

Appendix B Proofs of the Results in Section 3

B.1 Proof of Theorem 4

Let p_t denote the random price in period t under policy π and $\mathcal{F}_t = \sum_{s=1}^t \begin{bmatrix} 1 & p_s \\ p_s & p_s^2 \end{bmatrix}$ denote the Fisher information matrix. It is straightforward that Lemma A.7 holds using Lemma 1 in Keskin and Zeevi (2014), which is used to show Theorem 4.

Lemma A.7 *There exist positive constants μ_0 and μ_1 such that*

$$\sup_{\theta\in\Theta} \left\{\sum_{t=2}^T \rho^{t-1} \mathbb{E}(p_t - \varphi(\theta))^2\right\} \geq \sum_{t=2}^T \rho^{t-1} \frac{\mu_0}{\mu_1 + \sup_{\theta\in\Theta}\{C(\theta)\mathbb{E}[\mathcal{F}_{t-1}]C(\theta)^\intercal\}}, \tag{A-5}$$

where $C(\cdot)$ is a 1×2 matrix function on Θ such that $C(\theta) = [-\varphi(\theta)\ 1]$.

Proof of Theorem 4: By the definition of $C(\theta)$, we have $C(\theta)\mathbb{E}[\mathcal{F}_{t-1}]C(\theta)^\intercal = \sum_{s=1}^{t-1}\mathbb{E}(p_s-\varphi(\theta))^2$. Thus, inequality (A-5) in Lemma A.7 is equivalent to the following:

$$\sup_{\theta\in\Theta}\left\{\sum_{t=2}^{T}\rho^{t-1}\mathbb{E}(p_t-\varphi(\theta))^2\right\} \geq \sum_{t=2}^{T}\rho^{t-1}\frac{\mu_0}{\mu_1+\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\mathbb{E}(p_s-\varphi(\theta))^2\}}, \tag{A-6}$$

By the definition of regret, we have

$$\begin{aligned}
\Delta^\pi(T,\rho) &= \sup_{\theta\in\Theta}\left\{\sum_{t=1}^{T}\rho^{t-1}\mathbb{E}(r_\theta^*-r_\theta(p_t))\right\}\\
&= \sup_{\theta\in\Theta}\left\{-\beta\sum_{t=1}^{T}\rho^{t-1}\mathbb{E}(p_t-\varphi(\theta))^2\right\}\\
&\geq |b_{\max}|\sup_{\theta\in\Theta}\left\{\sum_{t=1}^{T}\rho^{t-1}\mathbb{E}(p_t-\varphi(\theta))^2\right\}.
\end{aligned}$$

The second equality holds since $r_\theta^*-r_\theta(p_t)=\varphi(\theta)(\alpha+\beta\varphi(\theta))-p_t(\alpha+\beta p_t)$ and we replace α with $-2\beta\varphi(\theta)$. Using (A-6), we have

$$\begin{aligned}
\Delta^\pi(T,\rho) &\geq b_{\max}^2\mu_0\sum_{t=2}^{T}\rho^{t-1}\frac{1}{\mu_1|b_{\max}|+|b_{\max}|\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\mathbb{E}(p_s-\varphi(\theta))^2\}}\\
&\geq K_{10}\sum_{t=2}^{T}\rho^{t-1}\frac{1}{K_{11}|b_{\max}|\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\mathbb{E}(p_s-\varphi(\theta))^2\}}\\
&= K_{10}\sum_{t=2}^{T}\frac{\rho^{2t-3}}{K_{11}|b_{\max}|\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\rho^{t-2}\mathbb{E}(p_s-\varphi(\theta))^2\}}\\
&\geq K_{10}\sum_{t=2}^{T}\frac{\rho^{2t-3}}{K_{11}|b_{\max}|\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\rho^{s-1}\mathbb{E}(p_s-\varphi(\theta))^2\}}\\
&\geq K_{10}\sum_{t=2}^{T}\frac{\rho^{2t-3}}{K_{11}\Delta^\pi(t-1,\rho)}\\
&\geq K_{10}\sum_{t=2}^{T}\frac{\rho^{2t-3}}{K_{11}\Delta^\pi(T,\rho)} = \frac{K_{10}}{K_{11}\Delta^\pi(T,\rho)}\frac{\rho(1-\rho^{2T-2})}{1-\rho^2},
\end{aligned}$$

where $K_{10}=\mu_0 b_{\max}^2$ and $K_{11}=1+\frac{4\mu_1}{(u-l)^2}\geq 1+\frac{\mu_1}{\sup_{\theta\in\Theta}\{\mathbb{E}(p_1-\varphi(\theta))^2\}}\geq 1+\frac{\mu_1}{\sup_{\theta\in\Theta}\{\sum_{s=1}^{t-1}\mathbb{E}(p_s-\varphi(\theta))^2\}}$. The last inequality holds since $\Delta^\pi(t,\rho)$ increases in t.

Then, we have $\Delta^\pi(T,\rho)\geq\sqrt{\frac{K_{10}}{K_{11}}\frac{\rho(1-\rho^{2T-2})}{1-\rho^2}}$. Thus, $\Delta^\pi(T,\rho)\geq K_4\sqrt{\frac{\rho(1-\rho^{2T-2})}{1-\rho^2}}$, where $K_4=\sqrt{K_{10}/K_{11}}$.

When $T\to\infty$,

$$\lim_{T\to\infty}\Delta^\pi(T,\rho)\geq\lim_{T\to\infty}K_4\sqrt{\frac{\rho(1-\rho^{2T-2})}{1-\rho^2}}=K_4\sqrt{\frac{\rho}{1-\rho^2}}=\Omega\left(\sqrt{\frac{1}{1-\rho}}\right).$$

When $\rho\to 1$,

$$\lim_{\rho\to 1}\Delta^\pi(T,\rho)\geq\lim_{\rho\to 1}K_4\sqrt{\frac{\rho(1-\rho^{2T-2})}{1-\rho^2}}=K_4\sqrt{T-1}=\Omega(\sqrt{T}).$$

■

B.2 Proof of Theorem 5

We show Theorem 5 using Lemma A.8 below, which is reproduced verbatim from Lemma 3 of Keskin and Zeevi (2014).

Lemma A.8 *There exist finite positive constants λ and γ such that, under any pricing policy π,*

$$\mathbb{P}_\theta^\pi\{\|\hat{\theta}_t - \theta\| > \delta, J_t \geq m\} \leq \gamma t \exp(-\lambda(\delta \wedge \delta^2)m) \tag{A-7}$$

for all $\delta, m > 0$, and $t \geq 2$.

Proof of Theorem 5: Using Lemma A.8 and condition (i), Keskin and Zeevi (2014) show that, there exists a constant K_{12} such that

$$\mathbb{E}(\varphi(\theta) - p_{t+1})^2 \leq \frac{K_{12}\log t}{\lambda\kappa_0\sqrt{t}} + 2\mathbb{E}(\varphi(\vartheta_t) - p_{t+1})^2 \text{ for all } t \geq N,$$

where N satisfies $kN\exp(-\frac{1}{2}\lambda\kappa_0\sqrt{N}) \leq 1$. For $N \geq \exp$, we have

$$\begin{aligned}
&\sum_{t=N}^{T-1}\rho^t\mathbb{E}(\varphi(\theta) - p_{t+1})^2 \\
&\leq K_{12}\sum_{t=N}^{T-1}\rho^t\frac{\log t}{\lambda\kappa_0\sqrt{t}} + 2\sum_{t=N}^{T-1}\rho^t\mathbb{E}(\varphi(\vartheta_t) - p_{t+1})^2 \\
&\leq K_{12}\sum_{t=N}^{N^2-1}\rho^t\frac{\log t}{\lambda\kappa_0\sqrt{t}} + K_{12}\sum_{t=N^2}^{T-1}\rho^t\frac{\log t}{\lambda\kappa_0\sqrt{t}} + 2\left(\kappa_1 N + \kappa_2\sum_{s=\kappa_3 N^2}^{T-1}\rho^s s^{-1/2}\right) \\
&\leq 4K_{12}\log N\frac{N}{\lambda\kappa_0} + K_{12}\frac{2\log N}{\lambda\kappa_0 N}\frac{\rho^{N^2}(1-\rho^{T-N^2})}{1-\rho} + 2\kappa_1 N + \frac{2\kappa_2}{\sqrt{\kappa_3}N}\frac{\rho^{\kappa_3 N^2}(1-\rho^{T-\kappa_3 N^2})}{1-\rho} \\
&\leq K_{13}N\log N + K_{14}\frac{\log N}{N}\frac{1-\rho^{T-N^2}}{1-\rho},
\end{aligned}$$

where $K_{13} = \frac{4K_{12}}{\lambda\kappa_0} + 2\kappa_1$ and $K_{14} = \frac{2K_{12}}{\lambda\kappa_0} + \frac{2\kappa_2}{\sqrt{\kappa_3}}$. The third inequality holds since $\log t/\sqrt{t}$ decreases in t for $t \geq \exp^2$. Since $\beta \in [b_{\min}, b_{\max}]$, we have

$$\begin{aligned}
\Delta^\pi(T,\rho) &= \sup_{\theta\in\Theta}\left\{-\beta\sum_{t=0}^{T-1}\rho^t\mathbb{E}(p_{t+1} - \varphi(\theta))^2\right\} \\
&\leq |b_{\min}|\sup_{\theta\in\Theta}\left\{\sum_{t=0}^{N-1}\rho^t\mathbb{E}(p_{t+1} - \varphi(\theta))^2 + \sum_{t=N}^{T-1}\rho^t\mathbb{E}(p_{t+1} - \varphi(\theta))^2\right\} \\
&\leq |b_{\min}|\left\{\frac{1-\rho^N}{1-\rho}(u-l)^2 + K_{13}N\log N + K_{14}\frac{\log N}{N}\frac{1-\rho^{T-N^2}}{1-\rho}\right\} \\
&= K_5\frac{1-\rho^N}{1-\rho} + K_6 N\log N + K_7\frac{\log N}{N}\frac{1-\rho^{T-N^2}}{1-\rho},
\end{aligned}$$

where $K_5 = |b_{\min}|(u-l)^2$, $K_6 = |b_{\min}|K_{13}$, and $K_7 = |b_{\min}|K_{14}$. Let $h(\rho,T) = K_5\frac{1-\rho^N}{1-\rho} + K_6 N\log N + K_7\frac{\log N}{N}\frac{1-\rho^{T-N^2}}{1-\rho}$ and $N = \left\lfloor\sqrt{\frac{1-\rho^T}{1-\rho}}\right\rfloor$. When $\rho\to 1$, we have $N = \lfloor\sqrt{T}\rfloor$ and

$$\lim_{\rho\to 1} h(\rho,T) = K_5\lfloor\sqrt{T}\rfloor + K_6\lfloor\sqrt{T}\rfloor\log\lfloor\sqrt{T}\rfloor + K_7\log\lfloor\sqrt{T}\rfloor\frac{(T-\lfloor\sqrt{T}\rfloor^2)}{\lfloor\sqrt{T}\rfloor} = \mathcal{O}(\sqrt{T}\log T).$$

When $T\to\infty$, we have $N = \left\lfloor\sqrt{\frac{1}{1-\rho}}\right\rfloor$ and

$$\begin{aligned}\lim_{T\to\infty} h(\rho,T) &= K_5\frac{1-\rho^N}{1-\rho} + K_6\left\lfloor\sqrt{\frac{1}{1-\rho}}\right\rfloor\log\left\lfloor\sqrt{\frac{1}{1-\rho}}\right\rfloor + K_7\frac{1/(1-\rho)}{\left\lfloor\sqrt{1/(1-\rho)}\right\rfloor}\log\left\lfloor\sqrt{\frac{1}{1-\rho}}\right\rfloor \\ &= \mathcal{O}\left(\sqrt{\frac{1}{1-\rho}}\log\left(\frac{1}{1-\rho}\right)\right).\end{aligned}$$

■

B.3 Proof of Theorem 6

We first show Lemma A.9, which will be used to show Theorem 6.

Lemma A.9 *Let $\hat{\pi}$ denote our policy. Then there exist positive constants K_{15}, $\check{T}\in\mathbb{N}$, and $\check{\rho}\in[0,1)$ such that, under policy $\hat{\pi}$, for all $T\geq\check{T}$ and $\rho\geq\check{\rho}$, we have $\mathbb{E}(\varphi(\theta)-\varphi(\vartheta_t))^2 \leq K_{15}\frac{\eta}{\tau}$ for $t\geq 2c_2\tau$.*

Proof of Lemma A.9: For $t\geq 2c_2\tau$, we have $J_t \geq \sum_{s=1}^{2c_2\tau}(p_s-\bar{p}_t)^2 \geq \frac{c_2\tau}{2}(\tilde{p}_1-\tilde{p}_2)^2 = k_1\tau$, where $k_1 = \frac{c_2}{2}(\tilde{p}_1-\tilde{p}_2)^2$. By the mean value theorem, we have $|\varphi(\theta)-\varphi(\vartheta_t)| \leq \sqrt{2k_2}\|\theta-\vartheta_t\|$, where $k_2 = \max_{j\in\{1,2\}}\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\}$. By monotonicity of expectation, we have

$$\begin{aligned}&\mathbb{E}(\varphi(\theta)-\varphi(\vartheta_t))^2 \\ &\leq 2k_2\mathbb{E}\|\theta-\vartheta_t\|^2 \\ &\leq 2k_2\mathbb{E}\|\theta-\hat{\theta}_t\|^2 \\ &= 2k_2\int_0^\infty \mathbb{P}(\|\theta-\hat{\theta}_t\|^2 > x, J_t\geq k_1\tau)dx \\ &\leq \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\int_{\frac{2\eta}{\lambda k_1\tau}}^\infty \gamma t\exp(-\lambda(\sqrt{x}\wedge x)k_1\tau)dx \\ &\leq \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\left[\int_{\frac{2\eta}{\lambda k_1\tau}}^1 \gamma t\exp(-\lambda x k_1\tau)dx + \int_1^\infty \gamma t\exp(-\lambda\sqrt{x}k_1\tau)dx\right] \\ &= \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\left[\frac{\gamma t\exp(-2\eta)}{\lambda k_1\tau} + \frac{\gamma t\exp(-\lambda k_1\tau)}{\lambda k_1\tau} + \frac{2\gamma t\exp(-\lambda k_1\tau)}{\lambda^2 k_1^2\tau^2}\right] \\ &\leq \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\left[\frac{\gamma T\exp(-2\eta)}{\lambda k_1\tau} + \frac{\gamma T\exp(-\lambda k_1\tau)}{\lambda k_1\tau} + \frac{2\gamma T\exp(-\lambda k_1\tau)}{\lambda^2 k_1^2\tau^2}\right].\end{aligned}$$

The third inequality holds by Lemma A.8. Note that

$$\lim_{\rho\to 1} T\exp(-2\eta) = T^{-1} < \log T = \lim_{\rho\to 1}\eta \text{ for } T\geq 2.$$

$$\lim_{T\to\infty}\lim_{\rho\to 1} T\exp(-\lambda k_1\tau) = \lim_{T\to\infty} T\exp\left(-\lambda k_1\left[\sqrt{T}\right]\right) = 0 < \lim_{T\to\infty}\log T = \lim_{T\to\infty}\lim_{\rho\to 1}\eta,$$
$$\lim_{T\to\infty}\lim_{\rho\to 1} T\exp(-\lambda k_1\tau)/\tau = \lim_{T\to\infty} T\exp\left(-\lambda k_1\left[\sqrt{T}\right]\right)/\left(\left[\sqrt{T}\right]\right) = 0 < \lim_{T\to\infty}\log T = \lim_{T\to\infty}\lim_{\rho\to 1}\eta.$$

Thus, there exists $\check{T}\in\mathbb{N}$ and $\check{\rho}\in[0,1)$ such that for all $T\geq\check{T}$ and $\rho\geq\check{\rho}$, we have

$$\begin{aligned}
&\mathbb{E}(\varphi(\theta)-\varphi(\vartheta_t))^2\\
&\leq \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\left[\frac{\gamma T\exp(-2\eta)}{\lambda k_1\tau} + \frac{\gamma T\exp(-\lambda k_1\tau)}{\lambda k_1\tau} + \frac{2\gamma T\exp(-\lambda k_1\tau)}{\lambda^2 k_1^2\tau^2}\right]\\
&\leq \frac{4k_2\eta}{\lambda k_1\tau} + 2k_2\left[\frac{\gamma\eta}{\lambda k_1\tau} + \frac{\gamma\eta}{\lambda k_1\tau} + \frac{2\gamma\eta}{\lambda^2 k_1^2\tau}\right]\\
&\leq K_{15}\frac{\eta}{\tau},
\end{aligned}$$

where $K_{15} = \frac{4k_2}{\lambda k_1} + \frac{4k_2\gamma}{\lambda k_1} + \frac{4k_2\gamma}{\lambda^2 k_1^2}$. ■

Proof of Theorem 6: Under policy $\hat{\pi}$, we have

$$\sum_{t=2c_2\tau}^{T-1}\rho^t\mathbb{E}(\varphi(\theta)-p_{t+1})^2 = \sum_{t=2c_2\tau}^{T-1}\rho^t\mathbb{E}(\varphi(\theta)-\varphi(\vartheta_t))^2 \leq K_{15}\sum_{t=2c_2\tau}^{T-1}\rho^t\frac{\eta}{\tau} = K_{15}\frac{\eta}{\tau}\frac{\rho^{2c_2\tau}(1-\rho^{T-2c_2\tau})}{1-\rho}.$$

Then, we have

$$\begin{aligned}
\Delta^{\hat{\pi}}(T,\rho) &\leq |b_{\min}|\sup_{\theta\in\Theta}\left\{\sum_{t=0}^{2c_2\tau-1}\rho^t\mathbb{E}(p_{t+1}-\varphi(\theta))^2 + \sum_{t=2c_2\tau}^{T-1}\rho^t\mathbb{E}(p_{t+1}-\varphi(\theta))^2\right\}\\
&\leq |b_{\min}|(u-l)^2\frac{1-\rho^{2c_2\tau}}{1-\rho} + |b_{\min}|K_{15}\frac{\eta}{\tau}\frac{\rho^{2c_2\tau}(1-\rho^{T-2c_2\tau})}{1-\rho}\\
&= K_8\frac{1-\rho^{2c_2\tau}}{1-\rho} + K_9\frac{\eta}{\tau}\frac{\rho^{2c_2\tau}-\rho^T}{1-\rho},
\end{aligned}$$

where $K_8 = |b_{\min}|(u-l)^2$ and $K_9 = |b_{\min}|K_{15}$. When $\rho\to 1$, we have $\tau = \left[\sqrt{T}\right]$, $\eta = \log T$, and

$$\lim_{\rho\to 1}\Delta^{\hat{\pi}}(T,\rho) \leq 2c_2K_8\tau + K_9\frac{\eta}{\tau}(T-2c_2\tau) = \mathcal{O}(\sqrt{T}\log T).$$

When $T\to\infty$, we have $\tau = \left[\sqrt{\frac{1}{1-\rho}}\right]$, $\eta = \log\left(\frac{1}{1-\rho}\right)$, and

$$\lim_{T\to\infty}\Delta^{\hat{\pi}}(T,\rho) \leq K_8\frac{1-\rho^{2c_2\tau}}{1-\rho} + K_9\frac{\eta}{\tau}\frac{1}{1-\rho} = \mathcal{O}\left(\sqrt{\frac{1}{1-\rho}}\log\left(\frac{1}{1-\rho}\right)\right).$$

■

Proof of Lemma 1: Using Lemma A.8 and condition (i), Keskin and Zeevi (2014) show that, there exists a constant K_{12} such that

$$\mathbb{E}(\varphi(\theta)-p_{t+1})^2 \leq \frac{K_{12}\log t}{\lambda\kappa_0\sqrt{t}} + 2\mathbb{E}(\varphi(\vartheta_t)-p_{t+1})^2 \text{ for all } t\geq N,$$

where N is a constant which satisfies $kN\exp(-\frac{1}{2}\lambda\kappa_0\sqrt{N})\leq 1$. Using condition (ii), we have

$$\sum_{t=N}^{T-1}\mathbb{E}(\varphi(\theta)-p_{t+1})^2$$

$$
\begin{aligned}
&\leq K_{12} \sum_{t=N}^{T-1} \frac{\log t}{\lambda \kappa_0 \sqrt{t}} + 2 \sum_{t=N}^{T-1} \mathbb{E}(\varphi(\vartheta_t) - p_{t+1})^2 \\
&\leq \frac{2K_{12}}{\lambda \kappa_0} \sqrt{T} \log T + 2\kappa_1 \sqrt{T}.
\end{aligned}
$$

Since $\beta \in [b_{\min}, b_{\max}]$, we have

$$
\begin{aligned}
\Delta^{\pi}(T, \rho) = & \sup_{\theta \in \Theta} \left\{ -\beta \sum_{t=0}^{T-1} \rho^t \mathbb{E}(p_{t+1} - \varphi(\theta))^2 \right\} \\
&\leq |b_{\min}| \sup_{\theta \in \Theta} \left\{ \sum_{t=0}^{N-1} \mathbb{E}(p_{t+1} - \varphi(\theta))^2 + \sum_{t=N}^{T-1} \mathbb{E}(p_{t+1} - \varphi(\theta))^2 \right\} \\
&\leq |b_{\min}| \left\{ N(u-l)^2 + \frac{2K_{12}}{\lambda \kappa_0} \sqrt{T} \log T + 2\kappa_1 \sqrt{T} \right\} \\
&\leq K_{16} \sqrt{T} \log T,
\end{aligned}
$$

where $K_{16} = |b_{\min}|(N(u-l)^2 + \frac{2K_{12}}{\lambda \kappa_0} + 2\kappa_1)$. ■

Appendix C Proofs of the Results in Section 4

In this section, we consider the setting in Section 4 where the effective discount rate per decision period is $\rho(T) = (\rho_0)^{1/T}$ for $\rho_0 \in (0, 1)$. We show that for the models in BR and KZ, the regret under any policy is $\Omega(\sqrt{T})$ (Propositions A.1 and A.4). For the model in BR, we show that the regret under our policy as well that under the MLE-CYCLE policy in BR is $\mathcal{O}(\sqrt{T})$ (Propositions A.2 and A.3). For the model of KZ, we show that the regret is $\mathcal{O}(\log T \sqrt{T})$ under three policies – namely, the two variants of the greedy Iterated-Least-Squares policy in KZ and a different policy that we propose (Propositions A.5 and A.6).

Proposition A.1 *Consider the problem class $\mathcal{C}_{LB}$ defined in Theorem 1. For any policy ψ and $\rho(T) = (\rho_0)^{1/T}$, there exists a parameter $z \in \mathcal{Z}$, such that*

$$
\textit{Regret}(z, \mathcal{C}_{LB}, T, \rho(T), \psi) = \Omega(\sqrt{T}).
$$

Proof of Proposition A.1: Recall from Theorem 1 that there exists a parameter $z \in \mathcal{Z}$ such that

$$
\text{Regret}(z, \mathcal{C}_{LB}, T, \rho, \psi) \geq K_0 \left(\sqrt{\frac{\rho}{1-\rho}} (1 - \rho^{T-1}) + \sqrt{\frac{\rho^T (1 - \rho^{T-1})}{1-\rho}} \right).
$$

For $T \geq 2$ and $\rho(T) = (\rho_0)^{1/T}$, we have

$$
\begin{aligned}
&\text{Regret}(z, \mathcal{C}_{LB}, T, \rho(T), \psi) \\
&\geq K_0 \left(\sqrt{\frac{(\rho_0)^{1/T}}{1 - (\rho_0)^{1/T}} \left(1 - (\rho_0)^{(1-1/T)}\right)} + \sqrt{\frac{\rho_0 \left(1 - (\rho_0)^{(1-1/T)}\right)}{1 - (\rho_0)^{1/T}}} \right)
\end{aligned}
$$

$$
\begin{aligned}
&\geq K_0 \left(\sqrt{(\rho_0)^{1/2}} \left(1 - (\rho_0)^{(1-1/2)}\right) \sqrt{\frac{1}{1-(\rho_0)^{1/T}}} + \sqrt{\rho_0 \left(1 - (\rho_0)^{(1-1/2)}\right)} \sqrt{\frac{1}{1-(\rho_0)^{1/T}}} \right) \\
&= K_0 \left(\sqrt{(\rho_0)^{1/2}} \left(1 - (\rho_0)^{(1-1/2)}\right) + \sqrt{\rho_0 \left(1 - (\rho_0)^{(1-1/2)}\right)} \right) \sqrt{\frac{1}{1-(\rho_0)^{1/T}}} \\
&= \Omega(\sqrt{T}).
\end{aligned}
$$

■

Proposition A.2 *For any problem class $\mathcal{C}$ satisfying Assumptions 1 and 2 with corresponding exploration prices $\bar{\boldsymbol{p}} \in \mathcal{P}^k$ and $\rho(T) = (\rho_0)^{1/T}$, our policy $\hat{\psi}$ (defined in Section 2.2) satisfies*

$$
\textit{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho(T), \hat{\psi}) = \mathcal{O}\left(\sqrt{T}\right).
$$

Proof of Proposition A.2: Recall from Theorem 2, our policy $\hat{\psi}$ satisfies

$$
\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho, \hat{\psi}) \leq K_1 \frac{1-\rho^{k\tau}}{1-\rho} + K_2 \frac{\rho^{k\tau} - \rho^T}{(1-\rho)\tau},
$$

where $\tau = \left\lfloor \sqrt{\frac{1-\rho^T}{1-\rho}} \right\rfloor$. For $\rho(T) = (\rho_0)^{1/T}$, we have

$$
\begin{aligned}
\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho(T), \hat{\psi}) &\leq K_1 \frac{1-(\rho_0)^{k\tau/T}}{1-(\rho_0)^{1/T}} + K_2 \frac{(\rho_0)^{k\tau/T} - \rho_0}{\left(1-(\rho_0)^{1/T}\right)\tau} \\
&= \mathcal{O}\left(\frac{1-(\rho_0)^{-2k\sqrt{1-\rho_0}\sqrt{1-(\rho_0)^{1/T}}/\log \rho_0}}{1-(\rho_0)^{1/T}} \right) + \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \right) \\
&= \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \right) = \mathcal{O}\left(\sqrt{T}\right).
\end{aligned}
$$

■

Proposition A.3 *For any problem class $\mathcal{C}$ satisfying Assumptions 1 and 2 with corresponding exploration prices $\bar{\boldsymbol{p}} \in \mathcal{P}^k$ and $\rho(T) = (\rho_0)^{1/T}$, the MLE-CYCLE policy $\check{\psi}$ in BR satisfies*

$$
\textit{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho(T), \check{\psi}) = \mathcal{O}\left(\sqrt{T}\right).
$$

Proof of Proposition A.3: Recall from Theorem 3, the MLE-CYCLE policy $\check{\psi}$ satisfies

$$
\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho, \hat{\psi}) \leq K_3 \left(1 + \sum_{h=1}^{\lfloor \sqrt{2T} \rfloor} \rho^{h^2/2} \right).
$$

For $\rho(T) = (\rho_0)^{1/T}$, we have

$$
\text{Regret}(\boldsymbol{z}, \mathcal{C}, T, \rho(T), \hat{\psi}) \leq K_3 + \sum_{h=1}^{\lfloor \sqrt{2T} \rfloor} (\rho_0)^{h^2/(2T)} K_3
$$

$$
\begin{aligned}
&\leq K_3 + K_3 \int_0^{\sqrt{2T}} (\rho_0)^{h^2/(2T)} dh \\
&= K_3 + K_3 \int_0^{\sqrt{2T}} e^{\log(\rho_0) h^2/(2T)} dh \\
&= K_3 + K_3 \sqrt{\frac{1}{2}} \int_0^{T} e^{\log(\rho_0) x/T} \sqrt{\frac{1}{x}} dx.
\end{aligned}
$$

The last equality holds by letting $x = h^2/2$. When $T \to \infty$,

$$
\begin{aligned}
\lim_{T\to\infty} K_3 \left(1 + \sqrt{\frac{1}{2}} \int_0^{T} e^{\log(\rho_0) x/T} \sqrt{\frac{1}{x}} dx\right) &= \lim_{T\to\infty} K_3 \left(1 + \sqrt{\frac{1}{2}} \sqrt{\frac{1}{-\log(\rho_0)}} \sqrt{T} \int_0^{-\log(\rho_0)} e^{-w} w^{-1/2} dw\right) \\
&\leq K_3 \left(1 + \sqrt{\frac{\pi}{2}} \sqrt{\frac{1}{-\log(\rho_0)}} \sqrt{T}\right) \\
&= \mathcal{O}\left(\sqrt{T}\right).
\end{aligned}
$$

The first equality holds by letting $w = -\log(\rho_0) x/T$. The inequality holds since

$$
\int_0^{-\log(\rho_0)} e^{-w} w^{-1/2} dw < \int_0^{\infty} e^{-w} w^{-1/2} dw = \Gamma(1/2) = \sqrt{\pi}.
$$

■

Next, we consider the model in KZ with discounting where the discount factor $\rho(T) = (\rho_0)^{1/T}$.

Proposition A.4 *For any policy π and $\rho(T) = (\rho_0)^{1/T}$, we have*

$$
\Delta^{\pi}(T, \rho(T)) = \Omega\left(\sqrt{T}\right).
$$

Proof of Proposition A.4: Recall from Theorem 4, we have

$$
\Delta^{\pi}(T, \rho) \geq K_4 \sqrt{\frac{\rho(1 - \rho^{2T-2})}{1 - \rho^2}} \quad \text{for any policy } \pi, \rho \in [0, 1), \text{ and } T \geq 3.
$$

For $\rho(T) = (\rho_0)^{1/T}$, we have

$$
\begin{aligned}
&\Delta^{\pi}(T, \rho(T)) \\
&\geq K_4 \sqrt{\frac{(\rho_0)^{1/T}(1 - (\rho_0)^{(2T-2)/T})}{1 - (\rho_0)^{2/T}}} \\
&\geq K_4 \sqrt{\rho_0(1 - \rho_0)} \sqrt{\frac{1}{1 - (\rho_0)^{2/T}}} \\
&= \Omega(\sqrt{T}).
\end{aligned}
$$

■

Proposition A.5 *Let π be a pricing policy that satisfies the conditions in Theorem 5. Then, we have*

$$
\Delta^{\pi}(T, \rho(T)) = \mathcal{O}\left(\sqrt{T} \log T\right).
$$

Proof of Proposition A.5: Recall from Theorem 5 that the regret under policy π satisfies

$$\Delta^{\pi}(T,\rho) \leq K_5 \frac{1-\rho^N}{1-\rho} + K_6 N \log N + K_7 \frac{\log N}{N} \frac{1-\rho^{T-N^2}}{1-\rho}, \text{ for } N = \left\lfloor \sqrt{\frac{1-\rho^T}{1-\rho}} \right\rfloor.$$

For $\rho(T) = \rho_0^{1/T}$, we have

$$\begin{aligned}
&\Delta^{\pi}(T,\rho(T)) \\
&\leq K_5 \frac{1-(\rho_0)^{N/T}}{1-(\rho_0)^{1/T}} + K_6 N \log N + K_7 \frac{\log N}{N} \frac{1-(\rho_0)^{(T-N^2)/T}}{1-(\rho_0)^{1/T}} \\
&= \mathcal{O}\left(\frac{1-(\rho_0)^{-2\sqrt{1-\rho_0}\sqrt{1-(\rho_0)^{1/T}}/\log\rho_0}}{1-(\rho_0)^{1/T}} \right) + \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \log\left(\frac{1}{1-(\rho_0)^{1/T}} \right) \right) + \\
&\quad \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \log\left(\frac{1}{1-(\rho_0)^{1/T}} \right) \right) \\
&= \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \log\left(\frac{1}{1-(\rho_0)^{1/T}} \right) \right) = \mathcal{O}\left(\log T \sqrt{T} \right).
\end{aligned}$$

■

Proposition A.6 *Our policy $\hat{\pi}$ in Section 3.2.3 satisfies*

$$\Delta^{\hat{\pi}}(T,\rho(T)) = \mathcal{O}\left(\sqrt{T} \log T \right).$$

Proof of Proposition A.6: Recall from Theorem 6 that our policy $\hat{\pi}$ satisfies

$$\Delta^{\hat{\pi}}(T,\rho) \leq K_8 \frac{1-\rho^{2c_2\tau}}{1-\rho} + K_9 \frac{\eta}{\tau} \frac{\rho^{2c_2\tau} - \rho^T}{1-\rho},$$

where $\tau = \left\lfloor \sqrt{\frac{1-\rho^T}{1-\rho}} \right\rfloor$ and $\eta = \log\left(\frac{1-\rho^T}{1-\rho} \right)$. For $\rho(T) = (\rho_0)^{1/T}$, we have

$$\begin{aligned}
&\Delta^{\hat{\pi}}(T,\rho(T)) \\
&\leq K_8 \frac{1-(\rho_0)^{2c_2\tau/T}}{1-(\rho_0)^{1/T}} + K_9 \frac{\eta}{\tau} \frac{(\rho_0)^{2c_2\tau/T} - \rho_0}{1-(\rho_0)^{1/T}} \\
&= O\left(\frac{1-(\rho_0)^{-4c_2\sqrt{1-\rho_0}\sqrt{1-(\rho_0)^{1/T}}/\log\rho_0}}{1-(\rho_0)^{1/T}} \right) + \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \log\left(\frac{1-\rho_0}{1-(\rho_0)^{1/T}} \right) \right) \\
&= \mathcal{O}\left(\frac{1}{\sqrt{1-(\rho_0)^{1/T}}} \log\left(\frac{1}{1-(\rho_0)^{1/T}} \right) \right) = \mathcal{O}\left(\log T \sqrt{T} \right).
\end{aligned}$$

■

Appendix D Number of Exploration Periods as a Function of the Discount Factor

Table A.1 below shows the number of exploration periods as a function of the discount factor ρ for our policies, for policy MLE-CYCLE in BR, and for policy ILS-d in KZ.

Table A.1: The number of exploration periods as a function of ρ under our policy, policy MLE-CYCLE in BR, and policy ILS-d in KZ, for $T = 40,000$.

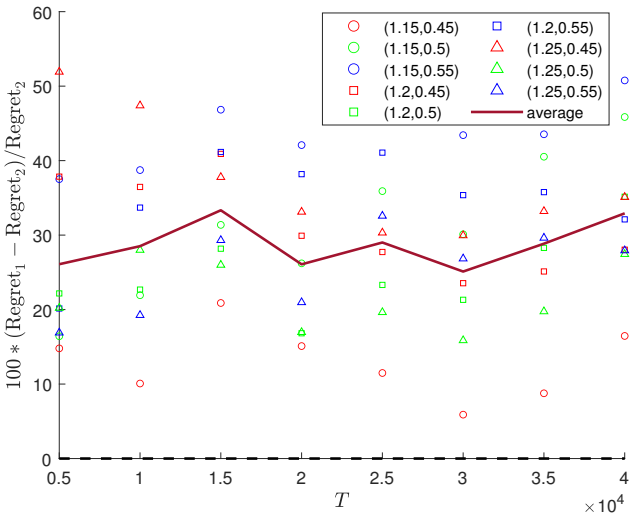
$\log_{10}\left(\frac{1}{1-\rho}\right)$	ρ	our policy	MLE-CYCLE	ILS-d
1	0.9	6	562	399
2	0.99	20	562	399
3	0.999	64	562	399
4	0.9999	198	562	399
5	0.99999	364	562	399
6	0.999999	396	562	399

We do not report the number of exploration periods for CILS in KZ because there is no clear boundary between exploration and exploitation under the CILS policy, thus making it difficult to determine the exact length of exploration. We now elaborate. Recall that under the CILS policy, in each period t, we first compute the difference between the greedy ILS price $\varphi(\vartheta_{t-1})$ in period t and the average price $\bar{p}_{t-1}$ in the first $t-1$ periods, denoted by $\delta_t = \varphi(\vartheta_{t-1}) - \bar{p}_{t-1}$. Then, for a positive constant c_1, the CILS policy charges $\bar{p}_{t-1} + \text{sgn}(\delta_t)c_1 t^{-1/4}$ if $|\delta_t| < c_1 t^{-1/4}$ and $\varphi(\vartheta_{t-1})$ otherwise. When $|\delta_t| \geq c_1 t^{-1/4}$, the CILS policy uses the greedy price $\varphi(\vartheta_{t-1})$ for exploitation. However, when $|\delta_t| < c_1 t^{-1/4}$, it is unclear whether the price $\bar{p}_{t-1} + \text{sgn}(\delta_t)c_1 t^{-1/4}$ is used purely exploration or exploitation. On the one hand, we exploit the average price $\bar{p}_{t-1}$, which is dynamically updated as time t increases and is sufficiently close to the greedy price when t is large. On the other hand, while we use a price deviation ($\text{sgn}(\delta_t)c_1 t^{-1/4}$) from the average price $\bar{p}_{t-1}$ for exploration, this deviation decreases with time t and is relatively small, so that the deviation from the greedy or "exploitation" price is not too much. That is, the CILS policy focuses more on exploitation and less on exploration as time t increases. When t is sufficiently large, the offered price can be very close to the greedy price. Therefore, there is no clear or simple answer to whether the price is used for exploration or exploitation. Alternatively, one can say that the price is used for both exploration and exploitation, and balances the tradeoff between the two. Since there is no clear definition for exploration periods in CILS, we do not report that number for the CILS policy in Table A.1.

Appendix E Additional Numerical Experience

We show the robustness of the superior performance of our policy when ρ is sufficient close to 1. In particular, for $\rho = 0.999999$, we numerically examine the behavior of the regret under different policies with respect to the time horizon T by varying T from 5000 to 40000, in increments of 5000, in the settings of both BR (see Section 2.3) and KZ (see Section 3.3). Figure A.1 (resp., Figure A.2) plots the relative difference between the average regret under policy MLE-CYCLE in BR and that under our policy for the linear (resp., logit) model.

Figure A.1: The relative percentage difference between the average regret under policy MLE-CYCLE and that under our policy for a linear model ($\rho = 0.999999$).



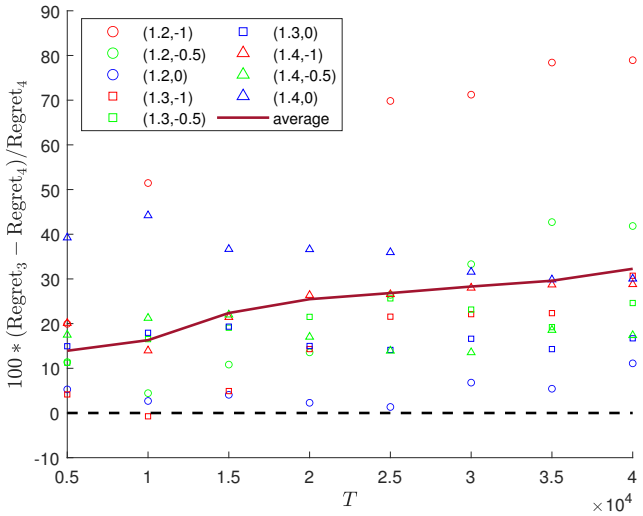
Note: Regret_1 is the average regret under MLE-CYCLE and Regret_2 is the average regret under our policy.

Figure A.3 (resp., Figure A.4) plots the relative difference between the average regret under policy ILS-d (resp., CILS) in KZ and that under our policy. We also provide a companion table (Table A.2) to Figures A.1, A.2, and A.3 to show the number of exploration periods as a function of T for our policy, for policy MLE-CYCLE in BR, and for policy ILS-d in KZ. As seen in Table A.2, the exploration length of our policy is similar to that of ILS-d. As seen in Figures A.1, A.2, A.3, and A.4, our policy consistently performs better when ρ is sufficient close to 1.

Table A.2: The number of exploration periods as a function of T under our policy, policy MLE-CYCLE in BR, and policy ILS-d in KZ, for $\rho = 0.999999$.

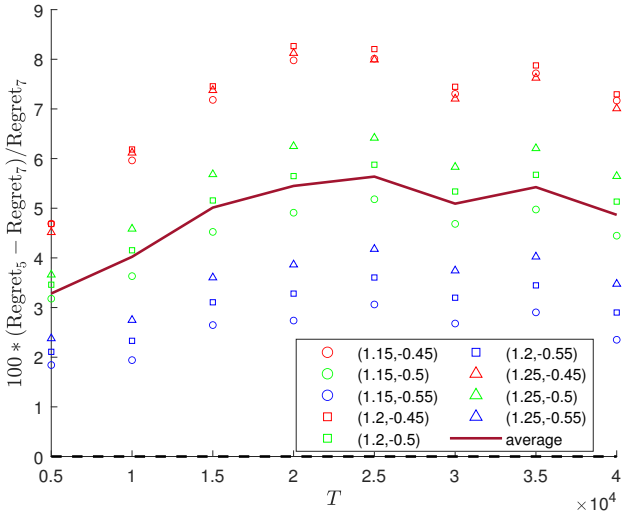
T	our policy	MLE-CYCLE	ILS-d
5000	142	196	140
10000	200	278	199
15000	244	342	244
20000	282	396	282
25000	314	444	316
30000	344	486	346
35000	370	526	374
40000	396	562	399

Figure A.2: The relative percentage difference between the average regret under policy MLE-CYCLE and that under our policy for a logit model ($\rho = 0.999999$).



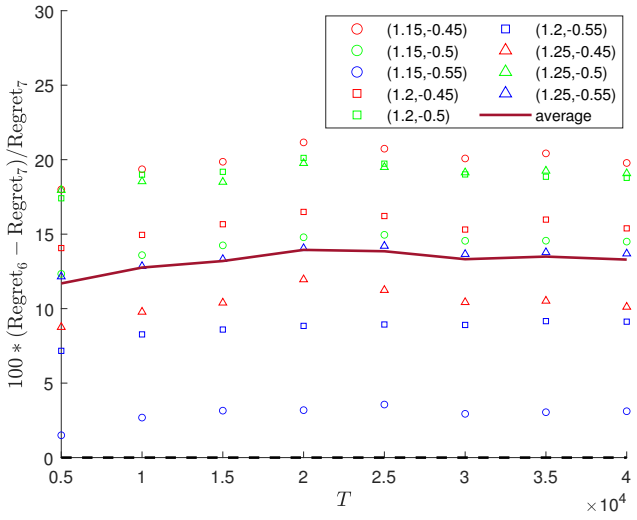
Note: Regret_3 is the average regret under MLE-CYCLE and Regret_4 is the average regret under our policy.

Figure A.3: The relative percentage difference between the average regret under policy ILS-d and that under our policy ($\rho = 0.999999$).



Note: Regret_5 is the average regret under ILS-d and Regret_7 is the average regret under our policy.

Figure A.4: The relative percentage difference between the average regret under policy CILS and that under our policy ($\rho = 0.999999$).



Note: Regret_6 is the average regret under CILS and Regret_7 is the average regret under our policy.