Appendix. E-Companion to "Dynamic Learning and Pricing with Model Misspecification"

A. A Different Regret Definition

In the literature on dynamic pricing with demand learning, it is standard to define regret relative to the clairvoyant who knows the *true* demand model. Let us refer to the clairvoyant defined in Section 3.1 as the "linear clairvoyant," and define a second clairvoyant, called the "true clairvoyant," who sets price $\tilde{p}_t = -\frac{f(x_t)}{2b}$ at each time period. Then we can define a second notion of regret, $\text{Regret}_2(T)$, in terms of the true clairvoyant:

$$\text{Regret}_2(T) = \sum_{t=1}^{T} \mathrm{E}[\tilde{p}_t D(\tilde{p}_t)] - \sum_{t=1}^{T} \mathrm{E}[p_t D(p_t)].$$

To see how $\text{Regret}(T)$ compares to $\text{Regret}_2(T)$, we can write

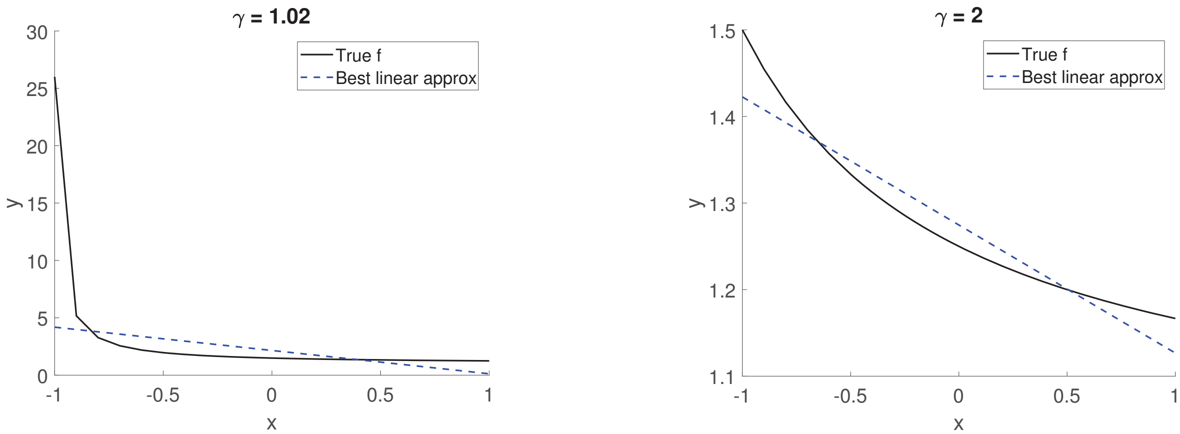
$$\begin{aligned}\text{Regret}_2(T) &= \text{Regret}(T) + \sum_{t=1}^{T} \mathrm{E}[\tilde{p}_t D(\tilde{p}_t)] - \mathrm{E}[p_t^* D(p_t^*)] \qquad \text{(EC.1)}\\ &= \text{Regret}(T) + \frac{T}{4|b|}\mathrm{E}\left[\left(f(x_t) - \mathrm{E}\left[f(x_t)\begin{bmatrix}1 & x_t^\intercal\end{bmatrix}\right]\left(\mathrm{E}\left[\begin{bmatrix}1 & x_t^\intercal \\ x_t & x_t x_t^\intercal\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1 \\ x_t\end{bmatrix}\right)^2\right]\\ &\geq \frac{T}{4|b|}\mathrm{E}\left[\left(f(x_t) - \mathrm{E}\left[f(x_t)\begin{bmatrix}1 & x_t^\intercal\end{bmatrix}\right]\left(\mathrm{E}\left[\begin{bmatrix}1 & x_t^\intercal \\ x_t & x_t x_t^\intercal\end{bmatrix}\right]\right)^{-1}\begin{bmatrix}1 \\ x_t\end{bmatrix}\right)^2\right]\end{aligned}$$

using closed form expressions for $\tilde{p}_t$ and p_t^*. This shows that the regret of any admissible pricing policy that assumes a misspecified demand model, relative to the true clairvoyant, grows linearly in T, and with the extent of model misspecification as captured by the expectation term in the second line. It reflects the fact that prices chosen by a seller who assumes a linear demand model may never converge to the optimal price $\tilde{p}_t$, because $\tilde{p}_t$ could depend nonlinearly on x_t. We have also included additional numerical experiment using $\text{Regret}_2(T)$ as the benchmark, see Appendix B.3.

Throughout the rest of this paper, we mainly focus on $\text{Regret}(T)$ rather than $\text{Regret}_2(T)$. $\text{Regret}(T)$ is a more interesting performance metric as (EC.1) shows that $\text{Regret}_2(T)$ of any admissible pricing policy affine in x_t is always $\Theta(T)$, implying that it cannot be optimized in terms of T. The term "regret" thus refers to $\text{Regret}(T)$ in the rest of this paper unless stated otherwise.

B. Additional Numerical Results

In this section we expand on the numerical results in Section 4 by investigating how our results depend on the parameter settings. Section B.1 shows how the performance of the RPS algorithm depends on the choice of demand function. Section B.2 looks at its dependence on the dimension of the feature vector m, complementing our theoretical results on the RPS algorithm's regret upper bound given in Section 3.



(a) $\gamma = 1.02$ (b) $\gamma = 2$

Figure EC.1 $f(x)$ **vs best linear approximation** $a + c'x$ **for** $\gamma = 1.02, 2$

B.1. Dependence of regret on demand function

We now investigate how the results of our simulations depend on the demand function. In the IID setting studied in Section 4, the quasilinear demand model is of the form

$$D_t(p) = \frac{1}{2(x_t + \gamma)} + 1 - 0.9p + \epsilon_t,$$

where $\gamma = 1.03$, while the closest linear approximation is

$$\hat{D}_t(p) \approx 2.05 - 0.90p - 1.76x_t$$

As γ increases, the fit of the closest linear approximation of D_t for x_t uniformly distributed between $[-1, 1]$ improves, i.e. $\mathrm{E}[(f(x_t) - a - c'x_t)^2]$ decreases. Fig. EC.1 illustrates this by comparing the function f with its best linear approximation on the interval $[-1, 1]$ for two values of γ, $\gamma = 1.02$ and 2. Since model misspecification worsens as γ decreases, we would expect that the endogeneity effect is more significant for demand models with smaller values of γ.

We ran the RPS and one-stage regression algorithms for $\gamma = 1.02, 1.03, 1.05, 1.1, 1.25, 1.5, 2.0$, keeping the price and parameter bounds the same as in the IID case numerical example with $\gamma = 1.03$. Table EC.1, which gives the estimates of the parameter b at the end of 5000 time periods averaged over 50 iterations, shows that for all γ, the RPS algorithm produces unbiased estimates of the parameter b. The one-stage regression algorithm estimates, on the other hand, are biased for smaller values of γ. As γ increases, the one-stage regression estimates of b improve. This is consistent with the observation that the endogeneity effect becomes more significant as γ decreases; the RPS algorithm, which corrects for endogeneity, produces unbiased parameter estimates for all γ, while the one-stage regression algorithm, which does not correct for endogeneity, only accurately estimates the parameters when the endogeneity effect becomes insignificant. Fig. EC.2 plots the average cumulative regret (over 50 iterations) of the RPS and one-stage regression algorithms at the end of 5000 time periods for the different values of γ. The RPS algorithm outperforms the one-stage regression algorithm for $\gamma < 2.0$, and the improvement of RPS relative to one-stage generally increases as γ decreases and the endogeneity effect increases. However, for $\gamma = 2.0$, one-stage regression outperforms RPS algorithm; In the absence of endogeneity, parameters can be estimated more efficiently using a one-stage rather than a two-stage regression, and RPS loses its competitive edge.

Table EC.1 Estimates of parameter b in Linear Demand Example

$\gamma =$	1.02	1.03	1.05	1.10	1.25	1.05	2.00s
RPS algo.	-0.94	-0.90	-0.91	-0.92	-0.90	-0.91	-0.90
One-stage reg.	-0.50	-0.50	-0.50	-0.53	-0.66	-0.77	-0.86

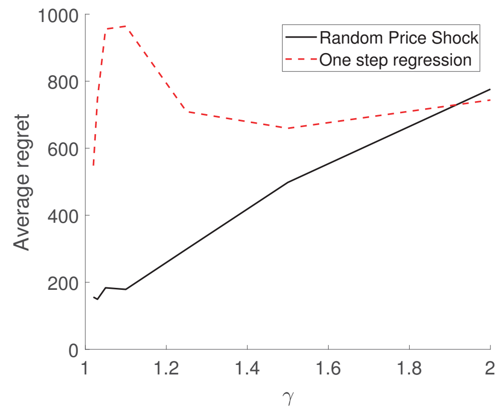


Figure EC.2 Average regret over 50 iterations of RPS vs one-stage regression algorithms as γ is varied

B.2. Dependence of regret on feature vector dimension m

We conducted numerical experiments in an attempt to investigate the dependence of the results on m. For simplicity, we looked at a number of different settings without any model misspecification, with $T = 5000$ and m varying from 1 to 1001. Unfortunately, almost none of these settings yielded a clear regret trend, and showed the regret seesawing with increasing m. One possible explanation is that the asymptotic dependence of the results on m only becomes detectable for larger values of m, which would be computationally infeasible to test.

However, for one of the settings tested, a clear regret trend was observed. Below, we report the results from this numerical experiment. The demand function is given by

$$D_t(p) = 2 - 0.7p + c^T x_t + \epsilon_t.$$

For each m, the feature vectors x_t are drawn IID from the distribution $[-1,1]^m$, and c is a vector of length m with the first entry set to 0.9 and all other entries set to 0. Note that $\|c\|_1$ is constant for all m, and thus so is $\bar{c}$, on which our regret bound depends (see Eq (EC.8) for the full statement of the IID regret bound in terms of all parameters). We set $c_{\max}$ to $c + [0.5, 0.5, \ldots, 0.5]$ and $c_{\min}$ to $c - [0.5, 0.5, \ldots, 0.5]$, and let the noise ϵ_t be normally distributed with mean 0 and variance 0.3. The price across all periods t is lower bounded by \$1.75 and upper bounded by \$8.25.

Fig. EC.3 plots the regrets of the RPS algorithm for $m = 1, 3, 5, 11, 51, 101, 201, 501, 1001$, averaged over 10 iterations each. We can see that the regret of RPS is increasing with m, and that the growth of the regret with m appears to be $O((m+1)T)$, in accordance with our regret upper bound. This numerical example thus supports the idea that the regret of the RPS algorithm does indeed depend on m, and that there is a gap in terms of m between our lower and upper bounds.

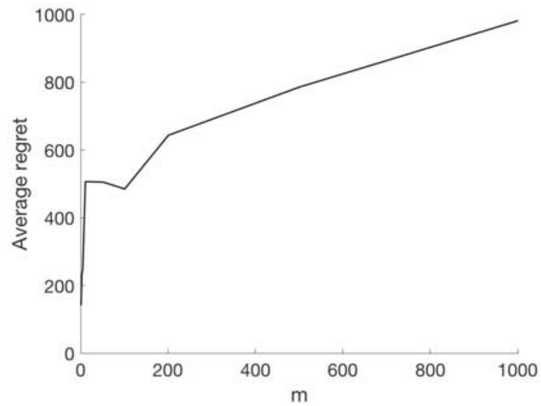


Figure EC.3 **Average regret over 10 iterations of the RPS algorithm as m is increased from 1 to 1001.**

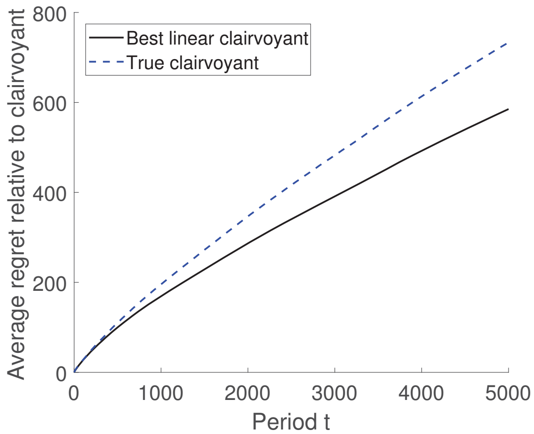
B.3. Regret relative to different clairvoyants

The above numerical experiments benchmark the performance of the RPS algorithm against the linear clairvoyant, who bases pricing decisions on the closest linear approximation of the true quasilinear demand model. Here we present additional numerical experiments benchmarking the performance of RPS against the true clairvoyant, who has full knowledge of the true quasilinear demand model, and sets price $\tilde{p}_t = -\frac{f(x)}{2b}$ at each time period. Fig. EC.4a plots the results of repeating the IID setting experiments from Section 4.1; it plots the average regret of the RPS algorithm relative to both clairvoyants over 200 iterations and 5000 time periods. Similarly, Fig. EC.4b plots the results of repeating the price ladder setting experiments from Section 4.1, and Fig. EC.4c plots the results of repeating the non IID experiments from Section 4.1.

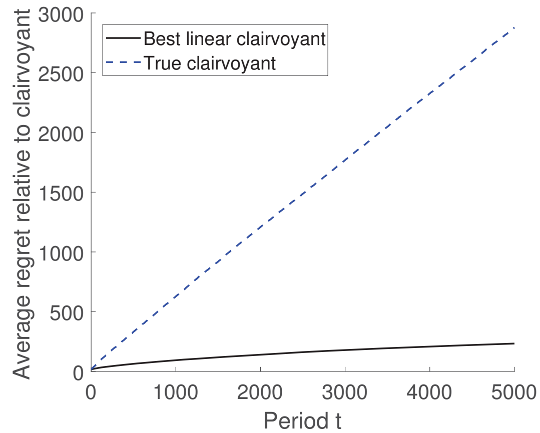
Fig. EC.4a confirms the result that the regret of RPS relative to the true clairvoyant grows linearly with T in the IID setting. On the other hand, Fig. EC.4b shows that, depending on the function f and the distribution of the feature vectors, the regret of RPS relative to the true clairvoyant need not grow linearly with T in the non IID setting. We can also observe from Figures EC.4a - EC.4c that the difference in revenue earned by the true clairvoyant and the revenue earned by the linear clairvoyant can vary considerably depending on the demand model and parameters; In the IID and price ladder settings, the extent of model misspecification is extremely large, while in the non IID setting, the linear clairvoyant achieves nearly as much revenue as the true clairvoyant. One way the retailer could try to improve the fit of her demand model in the first two cases is by including higher order terms of x_t in the feature vector and performing polynomial regression; however we note that she faces a tradeoff in doing so: The regret bound stated in Theorem 1, shows that the regret of RPS is $O((m+1)\sqrt{T})$, i.e. including more terms of x_t in the feature vector could decrease the regret from model misspecification, but increase the regret due to parameter estimation errors.

C. Appendix: Proofs for Theoretical Analysis

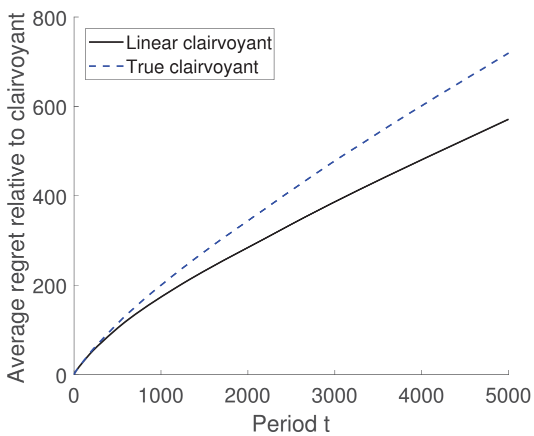
Notation. The following notations will be used in this section. We define $e := (a, c^\mathsf{T})^\mathsf{T}$ and $e_t := (\hat{a}_t, \hat{c}_t^\mathsf{T})^\mathsf{T}$. Let $\tilde{x} := (1, x^\mathsf{T})^\mathsf{T}$, $M := \mathrm{E}[\tilde{x}\tilde{x}^\mathsf{T}]$ and $M_t := \frac{1}{t-1}\sum_{j=1}^{t-1} \tilde{x}_j \tilde{x}_j^\mathsf{T}$.



(a) IID setting – average regret



(b) Price ladder IID setting – average regret



(c) Non IID setting – average regret

Figure EC.4 Average regret over 200 iterations of RPS algorithm relative to two different clairvoyants in IID and Price ladder IID settings

C.1. Proof of Proposition 1

Proof of Proposition 1. Consider price $p_t' = -\frac{\alpha+\gamma^\top x_t}{2\beta}$, where α, β, γ are measurable with respect to history $\mathcal{H}_{t-1}$. Since $p_t^* = -\frac{a+c^\top x_t}{2b}$, we have

$$\begin{aligned}
\mathrm{E}\left[p_t^* D(p_t^*) - p_t' D(p_t') \mid \mathcal{H}_{t-1}\right] =& \mathrm{E}\left[p_t^*(bp_t^* + f(x_t)) - p_t'(bp_t' + f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& \mathrm{E}\left[p_t^*(bp_t^* + a + c^\top x_t) - p_t'(bp_t' + a + c^\top x_t) - (p_t^* - p_t')(a + c^\top x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& \mathrm{E}\left[p_t^*(bp_t^* - 2bp_t^*) - p_t'(bp_t' - 2bp_t^*) \mid \mathcal{H}_{t-1}\right] - \mathrm{E}\left[(p_t^* - p_t')(a + c^\top x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right] \\
=& -b\mathrm{E}\left[(p_t^* - p_t')^2 \mid \mathcal{H}_{t-1}\right] - \mathrm{E}\left[(p_t^* - p_t')(a + c^\top x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right].
\end{aligned}$$

To finish the proof, we shall prove that $\mathrm{E}\left[(p_t^* - p_t')(a + c^\top x_t - f(x_t)) \mid \mathcal{H}_{t-1}\right] = 0$. By definition, a, c is the optimal solution of the following least squares problem

$$\min_{a', c'} \mathrm{E}\left[(f(x_t) - (a' + c'^\top x_t))^2\right].$$

By first order conditions, we have

$$\mathrm{E}\left[a+c^{\mathsf{T}}x_t-f(x_t)\right]=0,\quad \mathrm{E}\left[x_t\left(a+c^{\mathsf{T}}x_t-f(x_t)\right)\right]=0.$$

Since x_t is independent of the history $\mathcal{H}_{t-1}$, we have

$$\mathrm{E}\left[a+c^{\mathsf{T}}x_t-f(x_t)\mid\mathcal{H}_{t-1}\right]=0,\quad \mathrm{E}\left[x_t\left(a+c^{\mathsf{T}}x_t-f(x_t)\right)\mid\mathcal{H}_{t-1}\right]=0.$$

Therefore,

$$\begin{aligned}\mathrm{E}\left[(p_t^*-p_t')(a+c^{\mathsf{T}}x_t-f(x_t))\mid\mathcal{H}_{t-1}\right]=&\mathrm{E}\left[(-\frac{a+c^{\mathsf{T}}x_t}{2b}+\frac{\alpha+\gamma^{\mathsf{T}}x_t}{2\beta})(a+c^{\mathsf{T}}x_t-f(x_t))\mid\mathcal{H}_{t-1}\right]\\=&\mathrm{E}\left[(-\frac{a}{2b}+\frac{\alpha}{2\beta})\mathrm{E}\left[(a+c^{\mathsf{T}}x_t-f(x_t))\mid\mathcal{H}_{t-1}\right]\right]\\&+\mathrm{E}\left[(-\frac{c^{\mathsf{T}}}{2b}+\frac{\gamma^{\mathsf{T}}}{2\beta})\mathrm{E}\left[x_t(a+c^{\mathsf{T}}x_t-f(x_t))\mid\mathcal{H}_{t-1}\right]\right]=0.\quad\square\end{aligned}$$

which implies that $\mathrm{E}\left[(p_t^*-p_t')(a+c^{\mathsf{T}}x_t-f(x_t))\mid\mathcal{H}_{t-1}\right]=0$. Then, applying the law of total expectation, we prove the theorem.

C.2. Proof of Theorem 1

Proof. Recall that the expected regret over the selling horizon is defined as

$$\text{Expected Regret}(T)=\sum_{t=1}^{T}\mathrm{E}[p_t^*D(p_t^*)]-\sum_{t=1}^{T}\mathrm{E}[p_tD(p_t)]. \tag{EC.2}$$

First, let Q be a positive definite matrix such that $M=Q^2$ (Q must exist since M is positive definite). Then, let us define the event A_t as follows:

$$A_t=\left\{M_t\text{ is invertible and }\|QM_t^{-1}Q\|_2\le 2\right\}.$$

We can write the regret as

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)]=&\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t]+\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)|A_t^C]\cdot\mathbb{P}[A_t^C]\\\le&\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t]+\frac{(\bar{a}+\bar{c})^2}{2\underline{b}}\mathbb{P}[A_t^C]\\\le&\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t]\\&+\frac{(\bar{a}+\bar{c})^2}{\underline{b}}2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2+8(m+1)}\right)\end{aligned}$$

where the second inequality follows from the definition of p_t^* and our assumptions on the boundedness of the true parameters a,b,c, and the final inequality follows by bounding $\mathbb{P}(A_t^C)$ by Lemma EC.2, where $V:=\mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^{\mathsf{T}}Q^{-1}-I)^2]$. Since the second addend in the final line is $O(e^{-t})$, it is left to show that the first addend is $O(\sqrt{1/t})$.

We decompose it as follows:

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t]=&\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)-p_{g,t}^uD_t(p_{g,t}^u)|A_t]\cdot\mathbb{P}[A_t]\\&+\sum_{t=1}^{T}\mathrm{E}[p_{g,t}^uD_t(p_{g,t}^u)-p_{g,t}D_t(p_{g,t})|A_t]\cdot\mathbb{P}[A_t]\\&+\sum_{t=1}^{T}\mathrm{E}[p_{g,t}D_t(p_{g,t})-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t].\end{aligned}$$

Since $p_{g,t}=\mathrm{Proj}(p_{g,t}^u,[\underline{p}_t+\delta_t,\bar{p}_t-\delta_t])$ and the optimal price $\tilde{p}_t\in[\underline{p}_t,\bar{p}_t]$, we have

$$\sum_{t=1}^{T}\mathrm{E}[p_{g,t}^uD_t(p_{g,t}^u)-p_{g,t}D_t(p_{g,t})|A_t]\cdot\mathbb{P}[A_t]\le\sum_{t=1}^{T}\bar{b}\delta_t^2\cdot\mathbb{P}[A_t]\le\sum_{t=1}^{T}\frac{\delta^2\bar{b}}{4}\frac{1}{\sqrt{t}}\le\frac{\bar{b}\delta^2\sqrt{T}}{2}.$$

In addition, $p_t=p_{g,t}+\Delta p_t$, where Δp_t is generated independently from $p_{g,t},x_t$ and the history $\mathcal{H}_{t-1}$ with variance $\delta_t^2=\frac{\delta^2}{4\sqrt{t}}$. So

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}[p_{g,t}D_t(p_{g,t})-p_tD_t(p_t)|A_t]\cdot\mathbb{P}[A_t]&=\sum_{t=1}^{T}\mathrm{E}[p_{g,t}D_t(p_{g,t})-(p_{g,t}+\Delta p_t)D_t(p_{g,t}+\Delta p_t)|A_t]\cdot\mathbb{P}[A_t]\\&=\sum_{t=1}^{T}\mathrm{E}[\Delta p_t(-2bp_{g,t}-f(x_t))-b(\Delta p_t)^2|A_t]\cdot\mathbb{P}[A_t]\\&=\sum_{t=1}^{T}\mathrm{E}[-b(\Delta p_t)^2|A_t]\cdot\mathbb{P}[A_t]\\&\le\sum_{t=1}^{T}-\frac{b\delta^2}{4}\frac{1}{\sqrt{t}}\le\frac{\bar{b}\delta^2\sqrt{T}}{2}.\end{aligned}$$

To finish the proof, we want to show that $\mathrm{E}[p_t^*D_t(p_t^*)-p_{g,t}^uD_t(p_{g,t}^u)|A_t]\cdot\mathbb{P}[A_t]=O(1/\sqrt{t})$. In the proof of Proposition 1, we show that

$$\mathrm{E}[p_t^*D_t(p_t^*)-p_t'D_t(p_t')|\mathcal{H}_{t-1}]=-b\mathrm{E}[(p_t^*-p_t')^2|\mathcal{H}_{t-1}].$$

for any $p_t'=-\frac{\alpha+\gamma^\intercal x_t}{2\beta}$ with α,β,γ measurable with respect to the history $\mathcal{H}_{t-1}$.

Since the event A_t depends on the history $\mathcal{H}_{t-1}$ and is independent of x_t, this gives

$$\mathrm{E}[p_t^*D_t(p_t^*)-p_{g,t}^uD_t(p_{g,t}^u)|A_t]\cdot\mathbb{P}[A_t]=-b\mathrm{E}[(p_t^*-p_{g,t}^u)^2|A_t]\cdot\mathbb{P}[A_t],$$

where $p_{g,t}^u=-\frac{\hat{a}_t+\hat{c}_t^\intercal x_t}{2\hat{b}_t}$ is the greedy price given the estimates $\hat{a}_t,\hat{b}_t,\hat{c}_t$, and $p_t^*=-\frac{a+c^\intercal x_t}{2b}$ is the optimal price of the following linear model

$$D_t(p)=a+bp+c^\intercal x_t+\nu_t,\quad\forall p\in[\underline{p}_t,\bar{p}_t],$$

with $\nu_t=f(x_t)-a-c^\intercal x_t+\epsilon_t$.

By the definition of $p_{t,g}^u$ and p_t^*, we have

$$\begin{aligned}\mathrm{E}[(p_{t,g}^u-p_t^*)^2|A_t]\cdot\mathbb{P}[A_t]&=\mathrm{E}\left[\left(\frac{a+c^\intercal x_t}{2b}-\frac{\hat{a}_t+\hat{c}_t^\intercal x_t}{2\hat{b}_t}\right)^2|A_t\right]\cdot\mathbb{P}[A_t]\\&\le2\mathrm{E}\left[\left(\frac{a+c^\intercal x_t}{2b}-\frac{a+c^\intercal x_t}{2\hat{b}_t}\right)^2|A_t\right]\cdot\mathbb{P}[A_t]+2\mathrm{E}\left[\left(\frac{a+c^\intercal x_t}{2\hat{b}_t}-\frac{\hat{a}_t+\hat{c}_t^\intercal x_t}{2\hat{b}_t}\right)^2|A_t\right]\cdot\mathbb{P}[A_t]\\&\le(\bar{a}+\bar{c})^2\mathrm{E}\left[\left(\frac{1}{b}-\frac{1}{\hat{b}_t}\right)^2|A_t\right]\cdot\mathbb{P}[A_t]+\frac{1}{\underline{b}^2}\mathrm{E}\left[\left((a+c^\intercal x_t)-(\hat{a}_t+\hat{c}_t^\intercal x_t)\right)^2|A_t\right]\cdot\mathbb{P}[A_t]\end{aligned}$$

where the second line follows from the inequality $(x+y)^2 \le 2x^2 + 2y^2$, and the third line follows from the fact that the true parameters a, c satisfy $\|a\| \le \bar{a}$ and $\|c\|_1 \le \bar{c}$, as well as from the fact that $\hat{b}_t \in [-\bar{b}, -\underline{b}]$.

Now, for demand parameter b', let h be the function $h(b') = \frac{1}{b'}$. The gradient of h, denoted by ∇h, is given by $\nabla h(b') = -\frac{1}{b'^2}$, and we have $|\nabla h(b')|^2 = \frac{1}{b'^4} \le \frac{1}{\underline{b}^4}$. Then by the Mean Value Theorem, we have

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{b} - \frac{1}{\hat{b}_t}\right)^2 |A_t\right] \cdot \mathbb{P}[A_t] &\le \frac{1}{\underline{b}^4}\mathrm{E}[(b-\hat{b}_t)^2|A_t] \cdot \mathbb{P}[A_t] \\ &\le \frac{1}{\underline{b}^4}\mathrm{E}[(b-\hat{b}_t)^2]. \qquad \text{(EC.3)}\end{aligned}$$

By Lemma EC.1, we immediately have $\mathrm{E}[(\hat{b}_t - b)^2] = \mathrm{O}(1/\sqrt{t})$. Now we will bound the error in the estimates of a and c, namely $\mathrm{E}[\left((e-e_t)^\intercal \tilde{x}_t\right)^2 |A_t]$. Note that e_t is measurable with history $\mathcal{H}_{t-1}$ and $\tilde{x}_t$ is independent of $\mathcal{H}_{t-1}$, so

$$\begin{aligned}\mathrm{E}[\left((e-e_t)^\intercal \tilde{x}_t\right)^2 |A_t] =& \mathrm{E}\left[(e-e_t)^\intercal \mathrm{E}[\tilde{x}_t\tilde{x}_t^\intercal|A_t, \mathcal{H}_{t-1}](e-e_t)\right] \\ =& \mathrm{E}\left[(e-e_t)^\intercal M(e-e_t)|A_t\right] = \mathrm{E}\left[\|e-e_t\|_M^2|A_t\right],\end{aligned}$$

where $\|y\|_A := \sqrt{y^\intercal A y}$ for any positive definite matrix A.

By the definition of Algorithm 1, assuming that M_t is invertible, $e_t - e$ can be written as

$$e_t - e = \mathrm{Proj}\left(M_t^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\epsilon_s)}{t-1}\right). \qquad \text{(EC.4)}$$

Then we have

$$\begin{aligned}\mathrm{E}[\left((e-e_t)^\intercal \tilde{x}_t\right)^2 |] &= \mathrm{E}[\|e_t - e\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] \\ &\le \mathrm{E}[\|e_t - e\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] + \mathrm{E}[4(e^\intercal\tilde{x}_t)^2 + 4(e_t^\intercal\tilde{x}_t)^2] \cdot \mathbb{P}[A_t^C] \\ &\le \mathrm{E}[\|e_t - e\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] + 16\bar{b}^2 p_{\max}^2 \mathbb{P}[A_t^C] \\ &\le \mathrm{E}[\|e_t - e\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] \qquad \text{(EC.5)} \\ &\quad + 16\bar{b}^2 p_{\max}^2 \cdot 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right). \qquad \text{(EC.6)}\end{aligned}$$

The third line follows from the assumption that the true parameter $e \in E$. In the last step, we bound $\mathbb{P}(A_t^C)$ by Lemma EC.2, where $V := \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^\intercal Q^{-1} - I)^2]$. Since Eq (EC.6) is $O(e^{-t})$, it is left to show that Eq (EC.5) is $O(\sqrt{1/t})$.

We write Eq (EC.5) as

$$\begin{aligned}\mathrm{E}[\|e_t - e\|_M^2 \mid A_t] \cdot \mathbb{P}[A_t] &\le \mathrm{E}\left[\|QM_t^{-1}QQ^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\nu_s)}{t-1}\|^2 \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[\|QM_t^{-1}Q\|_2^2 \cdot \|Q^{-1}\|_2^2 \cdot \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s(p_s(b-\hat{b}_t)+\nu_s)}{t-1}\|^2 \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[4 \cdot \frac{1}{\lambda_{\min}(M)} \cdot 2\left(\|\frac{\sum_{s=1}^{t-1}\tilde{x}_s p_s(b-\hat{b}_t)}{t-1}\|^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right) \mid A_t\right]\mathbb{P}[A_t] \\ &\le \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left(\|\frac{\sum_{s=1}^{t-1}\tilde{x}_s p_s(b-\hat{b}_t)}{t-1}\|^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right)\right]\end{aligned}$$

$$\leq \mathrm{E}\left[\frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2(b-\hat{b}_t)^2 + \|\frac{\sum_{s=1}^{t-1}\tilde{x}_s\nu_s}{t-1}\|^2\right)\right]$$
$$= \frac{8}{\lambda_{\min}(M)}\left((m+1)p_{\max}^2\mathrm{E}[(b-\hat{b}_t)^2] + \frac{1}{(t-1)^2}\mathrm{E}[\|\sum_{s=1}^{t-1}\tilde{x}_s\nu_s\|^2]\right). \tag{EC.7}$$

The first inequality holds by Eq (EC.4) and the assumption that the true parameter $e \in E$. The second inequality holds from the submultiplicative property of the spectral norm. By the definition of Q, we have $\|Q^{-1}\|_2 = 1/\sqrt{\lambda_{\min}(M)}$. The third inequality uses the definition of event A_t and the fact $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$. The fourth inequality simply uses the definition of conditional expectation. The fifth inequality uses the assumptions that $\|x_t\|_\infty \leq 1$ and $p_j \leq p_{\max}$.

It has already been established using Lemma EC.1 that $\mathrm{E}[(b-\hat{b}_t)^2]$ is $O(1/\sqrt{t})$, so the first term of Eq (EC.7) is $O(1/\sqrt{t})$. For the second term, note that $(\tilde{x}_s, \nu_s)$ is independent of $(\tilde{x}_{s'}, \nu_{s'})$ for $s \neq s'$. Furthermore, by the first order condition of the least squares estimator, we have $\mathrm{E}[\nu_t] = 0$ and $\mathrm{E}[x_t\nu_t] = 0$. So for each s, $\mathrm{E}[\tilde{x}_s\nu_s \mid \mathcal{H}_{s-1}] = \mathrm{E}[\tilde{x}_s\nu_s] = 0$. Thus,

$$\frac{1}{(t-1)^2}\mathrm{E}[\|\sum_{s=1}^{t-1}\tilde{x}_s\nu_s\|^2] = \frac{1}{(t-1)^2}\sum_{s=1}^{t-1}\mathrm{E}[\|\tilde{x}_s\nu_s\|^2]$$
$$= \frac{1}{(t-1)^2}\sum_{s=1}^{t-1}\mathrm{E}[\|\tilde{x}_s(f(x_t) - a - c^\mathsf{T}x_t + \epsilon_t)\|^2]$$
$$\leq \frac{(m+1)}{t-1}3(\bar{f}^2 + 4\bar{b}^2p_{\max}^2 + \sigma^2),$$

where the last step uses the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and $\|\tilde{x}_s\|^2 \leq m+1$. Therefore, by Eq (EC.7), $\mathrm{E}[\|e-e_t\|_M^2] \leq O(1/\sqrt{t}) + O((m+1)/t) = O(1/\sqrt{t})$ as desired.

Dependence on $m, \underline{b}, \bar{b}$ and other parameters By combining constant factors, the expected regret of RPS algorithm over N periods can be bounded by

$$O\left(\frac{\bar{b}^2(p_{\max}^2+1)}{\underline{b}^4}\frac{(\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2)}{\delta^2}(1+p_{\max}^2\frac{m+1}{\lambda_{\min}(M)})\sqrt{T}\right) + O((m+1)\log T), \tag{EC.8}$$

where the pre-factor in the first big O notation only contains an absolute constant. □

C.3. Proof of Theorem 2

Proof. We will prove that the lower bound of regret is $\Omega(\sqrt{T})$ even if *the model is correctly specified*. Suppose there is no model misspecification, i.e. the demand function is given by

$$D_t(p) = a + bp + c^\mathsf{T}x_t + \epsilon_t.$$

We assume feature vector x_t is i.i.d. and sampled uniformly from $[-1/2, 1/2]^m$, and demand noise ϵ_t is i.i.d. normal with variance 1. By the first order condition, the optimal price that any non-anticipating pricing policy can charge at period t is $p_t^* = (a + c^\mathsf{T}x_t)/(-2b)$.

By Lemma EC.3, we can assume without loss of generality that the seller uses a linear pricing strategy π at period t given by $p_t = S_t + (U_t)^\mathsf{T}x_t$, where S_t and U_t are measurable with respect to the history

$\mathcal{H}_{t-1} = \sigma(x_1, \epsilon_1, \ldots, x_{t-1}, \epsilon_{t-1})$. Denote the regret incurred by the seller at the end of T periods as Regret(T). By Proposition 1, we have

$$\begin{aligned}\text{Regret}(T) &= -b\mathrm{E}[(p_t - p_t^*)^2] \\ &= -b\mathrm{E}[(S_t + (U_t)^\intercal x_t - S^* - (U^*)^\intercal x_t)^2] \\ &= -b\left\{\mathrm{E}[(S_t - S^*)^2] + \sum_{k=1}^m \mathrm{E}[(U_{t,k}x_{t,k} - U_k^* x_{t,k})^2]\right\} \\ &= -b\left\{\mathrm{E}[(S_t - S^*)^2] + \frac{1}{12}\sum_{k=1}^m \mathrm{E}[(U_{t,k} - U_k^*)^2]\right\},\end{aligned} \tag{EC.9}$$

where $S^* = -a/(2b)$, $U^* = -c/(2b)$, the third line follows since $\mathrm{E}[x_t] = 0$ and $x_t = (x_{t,1}, \ldots, x_{t,m})$ has independent entries for our particular choice of x_t, and the last line is because each entry of x_t has variance $\frac{1}{12}$.

Now we use the Van Trees inequality (Gill and Levit 1995), a Bayesian version of the Crámer-Rao inequality, to lower bound the regret of any admissible policy. The proof below is a generalization of the proof of Theorem 1 in Keskin and Zeevi (2014). Suppose the parameters $\theta = (a, b, c)$ belong to compact sets $\Theta = A \times B \times C$, where $A = [-\bar{a}, \bar{a}]$, $B = [-\bar{b}, -\underline{b}]$, $C = \{c' \in \mathbb{R}^m : \sum_{k=1}^m |c'_k| \le \bar{c}\}$. We can construct a prior distribution on Θ with density function λ which is positive on the interior and 0 on the boundary of Θ. We finish the proof by showing for any pricing policy that

$$\mathrm{E}_\lambda\left[\text{Regret}_\theta(T)\right] = \Omega(\sqrt{T}),$$

where $\text{Regret}_\theta(T)$ is the regret associated with a particular (unknown) parameter θ, and $\mathrm{E}_\lambda[\cdot]$ is the expectation operator on parameter θ under distribution λ. The above result immediately implies that there exists some parameter θ with regret $\Omega(\sqrt{T})$ for any pricing policy, namely

$$\max_{\theta \in \Theta}\{\text{Regret}_\theta(T)\} \ge \mathrm{E}_\lambda\left[\text{Regret}_\theta(T)\right] = \Omega(\sqrt{T}).$$

Let $f_t(H_t \mid \theta)$ be the joint probability density function of history $H_t = (x_1, p_1, D_1, \cdots, x_t, p_t, D_t)$ under parameter θ and a particular pricing policy $p_s = \pi(H_{s-1}, x_s)$. By our assumption that x_t is uniform and ϵ_t is normal, we have

$$f_t(H_t \mid \theta) = \Pi_{j=1}^t \phi(D_j - a - bp_j - c^\intercal x_j),$$

where ϕ is the density function of the standard normal distribution. The Fisher information matrix of θ given history H_t is

$$\mathcal{I}_t(\theta) = \mathrm{E}_\theta\left[\nabla_\theta \log f_t(H_t \mid \theta) \cdot (\nabla_\theta \log f_t(H_t \mid \theta))^\intercal\right] = \mathrm{E}_\theta\left[\sum_{j=1}^t \begin{bmatrix} 1 & x_j^\intercal & p_j \\ x_j & x_j x_j^\intercal & p_j x_j^\intercal \\ p_j & p_j x_j & p_j^2 \end{bmatrix}\right]. \tag{EC.10}$$

Define function $g(\theta) = [a/(2b), 1, c/(2b)]^\intercal$ and function $S(\theta) = -a/(2b) = S^*$. Applying the multivariate Van Trees inequality to S_t, which is an estimate of $S(\theta)$ based on history H_{t-1}, gives

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t - S(\theta))^2]] \ge \frac{\mathrm{E}_\lambda[g(\theta)^\intercal \nabla S(\theta)]^2}{\mathrm{E}_\lambda[g(\theta)^\intercal \mathcal{I}_{t-1}(\theta) g(\theta)] + \tilde{I}(\lambda)}, \tag{EC.11}$$

where $\tilde{\mathcal{I}}(\lambda)$ is the Fisher information of θ given prior λ. We have

$$g(\theta)^\mathsf{T}\cdot(\nabla S(\theta)) = \left[\frac{a}{2b},\,1,\,\frac{c}{2b}\right]^\mathsf{T}\cdot\left[-\frac{1}{2b},\,\frac{a}{2b^2},\,0\right] = \frac{a}{4b^2}.$$

By Eq (EC.10) and $p_j^* = -(a+c^\mathsf{T}x_j)/(2b)$, one can show that

$$g(\theta)^\mathsf{T}\mathcal{I}_{t-1}(\theta)g(\theta) = \mathrm{E}_\theta\left[\sum_{j=1}^{t-1}(p_j-p_j^*)^2\right] \le \mathrm{E}_\theta\left[\sum_{j=1}^{T}(p_j-p_j^*)^2 = \mathrm{Regret}_\theta(T)\right].$$

Substituting the equations above into Eq (EC.11), we get

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(S_t-S(\theta))^2]] \ge \frac{\left(\mathrm{E}_\lambda[\frac{a}{4b^2}]\right)^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)]+\tilde{I}(\lambda)}. \tag{EC.12}$$

Similarly, for each $k=1,\dots,m$, by letting $U_k(\theta)=U_k^*=-c_k/(2b)$ and applying Van Trees inequality, we get

$$\mathrm{E}_\lambda[\mathrm{E}_\theta[(U_{t,k}-U_k(\theta))^2]] \ge \frac{(\mathrm{E}_\lambda[\frac{c_k}{4b^2}])^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)]+\tilde{I}(\lambda)}. \tag{EC.13}$$

Combining (EC.9), (EC.12), (EC.13), and summing over $t=1,\dots,T$, we have

$$\mathrm{E}_\lambda\left[\mathrm{Regret}_\theta(T)\right] \ge \sum_{t=1}^{T}\underline{b}\left\{\frac{\left(\mathrm{E}_\lambda[\frac{a}{4b^2}]\right)^2+\frac{1}{12}\sum_{k=1}^m\mathrm{E}_\lambda[\frac{c_k}{4b^2}])^2}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)]+\tilde{I}(\lambda)}\right\} = \frac{\Omega(mT)}{\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)]+\tilde{I}(\lambda)}.$$

Note that $\tilde{I}(\lambda)$ is a constant independent of T. Consequently, we have

$$\mathrm{E}_\lambda[\mathrm{Regret}_\theta(T)] \ge \sqrt{\Omega(T)} - \frac{\tilde{I}(\lambda)}{2} = \Omega(\sqrt{T}). \quad \square$$

C.4. Proof of Theorem 3.

Proof. In the following, let $p_{t,u}^* := -\frac{a+c'x_t}{2b}$. We can decompose the regret into the loss due to imperfect knowledge of the true demand model, and the loss due to price experimentation, namely

$$\begin{aligned}\mathrm{Regret}(T) &= \sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)]-\mathrm{E}[p_tD_t(p_t)]\\ &= \sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)]-\mathrm{E}[p_{g,t}D_t(p_{g,t})]+\mathrm{E}[p_{g,t}D_t(p_{g,t})]-\mathrm{E}[p_tD_t(p_t)].\end{aligned}$$

The loss from price experimentation is upper bounded by

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}[p_{g,t}D_t(p_{g,t})]-\mathrm{E}[p_tD_t(p_t)] &= -b\sum_{t=1}^{T}\mathrm{E}[\Delta p_t^2]\\ &= -b\sum_{t=1}^{T}\mathrm{E}[(q_{i_t}-q_{i_t-1})(q_{i_t+1}-q_{i_t})t^{-1/3}]\\ &\le 3\bar{b}\bar{\delta}^2T^{2/3}.\end{aligned}$$

where the last line uses the assumption that $q_i - q_{i-1} \le \bar{\delta}$ for $i=1,\dots,N-1$.

The loss from parameter estimation is upper bounded by

$$\sum_{t=1}^{T}\mathrm{E}[p_t^*D_t(p_t^*)]-\mathrm{E}[p_{g,t}D_t(p_{g,t})] = \mathrm{E}[(p_{t,u}^*-(p_t^*-p_{t,u}^*)D_t(p_{t,u}^*-(p_t^*-p_{t,u}^*)]-\mathrm{E}[p_{g,t}D_t(p_{g,t})] \tag{EC.14}$$

$$\le K\mathrm{E}[|p_{t,u}^*-p_t^*+p_{t,u}^*-p_{t,g}|] \tag{EC.15}$$

$$\begin{aligned}&\le K(\mathrm{E}[|p_{t,u}^*-p_t^*|]+\mathrm{E}[|p_{t,u}^*-p_{t,g}|])\\ &\le 2K\mathrm{E}[|p_{t,u}^*-p_{t,g}|].\end{aligned}$$

The first line, (EC.14), follows from the fact that

$$\mathrm{E}[p_t^* D_t(p_t^*)] = \mathrm{E}[(p_{t,u}^* + (p_t^* - p_{t,u}^*) D_t(p_{t,u}^* + (p_t^* - p_{t,u}^*)] = \mathrm{E}[(p_{t,u}^* - (p_t^* - p_{t,u}^*) D_t(p_{t,u}^* - (p_t^* - p_{t,u}^*)],$$

by the symmetry of the function $p \mapsto \mathrm{E}[pD_t(p)]$ around its maximizer $p = p_{t,u}^*$. The second line, (EC.15), follows from the mean value theorem since $\mathrm{E}[pD_t(p)]$ is a differentiable function of p. By the mean value theorem, we have, for any $p_1, p_2 \in \{q_1, \ldots, q_N\}$, that

$$|\mathrm{E}[p_1 D_t(p_1)] - \mathrm{E}[p_2 D_t(p_2)]| \leq \max_{p \in \{q_1, \ldots, q_M\}} |\frac{dpD(p)}{dp}| \leq 2|b| p_{\max} + \bar{f},$$

thus (EC.15) follows by setting $K = 2|b|p_{\max} + \bar{f}$. Finally, the third line follows from the triangle inequality, and the last line follows from the fact that $|p_{t,u}^* - p_t^*| \leq |p_{t,u}^* - p_{t,g}|$ since $p_t^* = \arg\min_{q \in \{q_1, \ldots, q_N\}} |p_{t,u}^* - q|$.

It remains to bound $\mathrm{E}[|p_{t,u}^* - p_{t,g}|]$. Since $\mathrm{E}[|p_{t,u}^* - p_{t,g}|] \leq \sqrt{\mathrm{E}[|p_{t,u}^* - p_{t,g}|^2]}$, we can then bound $\mathrm{E}[|p_{t,u}^* - p_{t,g}|^2]$ using the same argument made in the proof of Theorem 1, giving an upper bound of

$$\frac{8}{\lambda_{\min}(M)} (m+1) p_{\max}^2 \mathrm{E}[(b - \hat{b}_t)^2] + O(\frac{m+1}{t-1}).$$

Lemma EC.1 can be applied to bound the term $\mathrm{E}[(b - \hat{b}_t)^2]$. Then, using the identity $\sqrt{x+y+z} \leq \sqrt{x} + \sqrt{y} + \sqrt{z}$ for $x, y, z \geq 0$, we can bound $\mathrm{E}[|p_{t,u}^* - p_{t,g}|]$ with

$$4\sqrt{2} \cdot \frac{p_{\max}(\bar{f} + \sigma + \bar{b} p_{\max})}{\delta \sqrt{\lambda_{\min}(M)}} \cdot \frac{\sqrt{m+1}}{t^{1/3}} + O(\sqrt{\frac{m+1}{t-1}})$$

Dependence on $m, \underline{b}, \bar{b}$ and other parameters By combining constant factors, the expected regret of the RPS algorithm over T periods can be bounded by

$$O\left((|b| p_{\max} + \bar{f}) \frac{p_{\max}(\bar{f} + \sigma + \bar{b} p_{\max})}{\underline{\delta} \sqrt{\lambda_{\min}(M)}} \sqrt{m+1} T^{2/3} \right) + O\left(\sqrt{(m+1)T} \right),$$

where the pre-factor in the first big O notation only contains an absolute constant. □

C.5. Proof of Proposition 2

Proof. Consider the optimization problem

$$\max_{\alpha, \beta, \gamma} \sum_{t=1}^{T} \mathrm{E}[p_t^* D_t(p_t^*) | \{x_1, \ldots, x_T\}] = \max_{\alpha, \beta, \gamma} \sum_{t=1}^{T} (-\frac{\alpha + \gamma^\mathsf{T} x_t}{2\beta})(b(-\frac{\alpha + \gamma^\mathsf{T} x_t}{2\beta}) + f(x_t)).$$

It is easy to see that for any optimal solution $(\alpha^*, \beta^*, \gamma^*)$, $(\alpha^* \frac{b}{\beta^*}, b, \gamma^* \frac{b}{\beta^*})$ is another optimal solution. Thus, setting $\beta = b$, we have the equivalent optimization problem

$$\max_{\alpha, \gamma} \sum_{t=1}^{T} (\alpha + \gamma^\mathsf{T} x_t)(2f(x_t) - (\alpha + \gamma^\mathsf{T} x_t)).$$

Finally, note that

$$\arg\max_{\alpha, \gamma} \sum_{t=1}^{T} (\alpha + \gamma^\mathsf{T} x_t)(2f(x_t) - (\alpha + \gamma^\mathsf{T} x_t)) = -\arg\min_{\alpha, \gamma} \sum_{t=1}^{T} (f(x_t) - (\alpha + \gamma^\mathsf{T} x_t))^2,$$

which proves Proposition 2. □

C.6. Proof of Theorem 4.

Proof. We decompose the regret as

$$\begin{aligned}\text{Regret}(T) &= \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_t D(p_t)] \\ &= \sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_{g,t}^u D(p_{g,t}^u)] \\ &+ \mathrm{E}[p_{g,t}^u D(p_{g,t}^u)] - \mathrm{E}[p_{g,t} D(p_{g,t})] + \mathrm{E}[p_{g,t} D(p_{g,t})] - \mathrm{E}[p_t D(p_t)].\end{aligned}$$

Following the proof of the regret bound in the IID setting, the quantity in the final line is upper bounded by

$$2\sum_{t=1}^{T} \bar{b}\delta_t^2 = 2\sum_{t=1}^{T} \frac{\bar{b}\delta^2}{4}\frac{1}{t^{1/3}} \leq \frac{3}{2}\bar{b}\delta^2 T^{2/3}.$$

To bound the difference between the oracle's revenue and the revenue earned by the greedy prices, we let $y_t = D_t - bp_t = f(x_t) + \epsilon_t$. Let $y_t' = D_t - \hat{b}_t p_t$. Let $e_t = (a_t, c_t)$ and let $e_x = (a_x, c_x)$ denote the parameters of the clairvoyant's demand model conditional on the realization $\{x_1, \ldots, x_T\}$. Let E_x denote the expectation conditional on a realization $\{x_1, \ldots, x_T\}$, namely

$$\mathrm{E}_x[\cdot] = \mathrm{E}[\cdot | x_t \text{ for } t = 1 \ldots T].$$

By rewriting the demands and prices in terms of y_t and y_t' we have

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{E}[p_t^* D(p_t^*)] - \mathrm{E}[p_{g,t}^u D_i(p_{g,t}^u)] &= \frac{1}{4|b|}\sum_{t=1}^{T} \mathrm{E}[\mathrm{E}_x[(e_t^\intercal \tilde{x}_t - y_t')^2 - (e_x^\intercal \tilde{x}_t - y_t')^2]] && \text{(EC.16)} \\ &+ \frac{1}{|b|}\mathrm{E}[\mathrm{E}_x[(p_{t,g}^u)^2(b^2 - \hat{b}_t^2)]]. && \text{(EC.17)} \\ &+ \frac{1}{2|b|}\mathrm{E}[\mathrm{E}_x[(y_t' - y_t)(e_t^\intercal \tilde{x}_t - e_x^\intercal \tilde{x}_t)]] && \text{(EC.18)} \\ &+ \frac{1}{|b|}\mathrm{E}[\mathrm{E}_x[y_t p_{t,g}^u(\hat{b}_t - b)]]. && \text{(EC.19)}\end{aligned}$$

First, we will bound (EC.16). Define $M_t = I_{m+1} + \sum_{s=1}^{t} \tilde{x}_s \tilde{x}_s^\intercal$. The closed form expression for the estimator e_t at period t is $(M_t)^{-1}(\sum_{s=1}^{t-1} y_s \tilde{x}_s)$. Expanding the expressions for e_t in (EC.16), we see that most of the terms in the expansion are telescoping, giving

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{E}_x[(e_t^\intercal \tilde{x}_t - y_t')^2 - (e_x^\intercal \tilde{x}_t - y_t')^2] &= \sum_{t=1}^{T} \mathrm{E}_x[(y_t')^2 \tilde{x}_t^\intercal M_t^{-1} \tilde{x}_t] && \text{(EC.20)} \\ &+ \mathrm{E}_x[\|e_x - e_1\|^2 - (e_x - e_{T+1})^\intercal M_T (e_x - e_{T+1})] && \text{(EC.21)} \\ &+ \mathrm{E}_x[(e_1^\intercal \tilde{x}_1)^2 + (e_x^\intercal \tilde{x}_{T+1})^2] && \text{(EC.22)} \\ &- \sum_{t=1}^{T} \mathrm{E}_x[(e_{t+1}^\intercal \tilde{x}_{t+1})^2 \tilde{x}_{t+1}^\intercal M_t^{-1} \tilde{x}_{t+1} - (e_{T+1}^\intercal \tilde{x}_{T+1})^2 - (e_x^\intercal \tilde{x}_1)^2].\end{aligned}$$

Since $e_1 = (I + \tilde{x}_1 \tilde{x}_1^\intercal)^{-1} \cdot 0 = 0$, $\|e_x - e_1\|^2 = \|e_x\|^2$. Then since M_t is positive semi-definite for all t, (EC.21) is upper bounded by $\|e_x\|^2 \leq \bar{a}^2 + \bar{c}^2$. Since $e_1 = 0$ and x_{T+1} can be set to 0, (EC.22) is 0. The final line is upper bounded by 0.

Finally, to bound Eq (EC.20), we can write

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}_x[(y_t')^2\tilde{x}_t^\intercal M_t^{-1}\tilde{x}_t] &= \sum_{t=1}^{T}\mathrm{E}_x[(f(x_t)+\epsilon_t+(b-\hat{b}_t)p_t)^2\tilde{x}_t^\intercal M_t^{-1}\tilde{x}_t] \\ &\leq \sum_{t=1}^{T}(2\sigma^2+2\bar{f}^2+4\bar{b}^2p_{\max}^2)\tilde{x}_t^\intercal M_t^{-1}\tilde{x}_t.\end{aligned}$$

The second line follows from the fact that ϵ_t is independent of the other terms and that it is mean 0 and variance σ^2. We also use the boundedness of f, $\hat{b}_t$ and p_t. Finally, using the identity

$$x^\intercal(\Sigma+xx^\intercal)^{-1}x=\frac{\det(\Sigma)}{\det(\Sigma+xx^\intercal)}$$

for any matrix Σ, we have

$$\begin{aligned}(2\sigma^2+2\bar{f}^2+4\bar{b}^2p_{\max}^2)\sum_{t=1}^{T}\tilde{x}_t^\intercal M_t^{-1}\tilde{x}_t &\leq (2\sigma^2+2\bar{f}^2+4\bar{b}^2p_{\max}^2)\sum_{t=1}^{T}1-\frac{\det(M_{t-1})}{\det(M_t)} \\ &\leq (2\sigma^2+2\bar{f}^2+4\bar{b}^2p_{\max}^2)\sum_{k=1}^{m+1}\log(1+\lambda_k),\end{aligned}$$

where the λ_js are the eigenvalues of $\sum_{t=1}^{T}\tilde{x}_t\tilde{x}_t^\intercal$. The sum of the λ_js is at most $T\cdot\max_t\|\tilde{x}_t\|^2$, which in turn is at most $\sqrt{m+1}T$. Thus the last line is $O((m+1)\cdot\log(T(m+1)))$. Then (EC.16) does not dominate the regret bound.

Now we will bound Eq (EC.17). Using the definition $p_{g,t}^u=-\frac{e_t^T\tilde{x}_t}{2b_t}$ and the fact that $|b_t|\geq \underline{b}$ gives

$$\begin{aligned}\mathrm{E}_x[(p_{t,g}^u)^2(b^2-\hat{b}_t^2)] &\leq \frac{1}{2\underline{b}}\mathrm{E}_x[(e_t^T\tilde{x}_t)^2(b^2-\hat{b}_t^2)] \\ &\leq \frac{1}{\underline{b}}\mathrm{E}_x[((e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2+(e_x^T\tilde{x}_t)^2)(b^2-\hat{b}_t^2)] \\ &\leq \frac{1}{\underline{b}}((2\bar{b}^2)\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]+(\bar{a}+\bar{c})^2\mathrm{E}_x[b^2-\hat{b}_t^2]). \qquad \text{(EC.23)}\end{aligned}$$

The second line follows from the identity $(x+y)^2\leq 2x^2+2y^2$. The third line follows from the fact that $b^2+\hat{b}_t^2\leq 2\bar{b}^2$ due to the assumptions on b and the projection step in the algorithm, as well as from the assumptions on e_x. Now, to bound $\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]$ in Eq (EC.23), note that we have

$$\sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]=\sum_{t=1}^{T}\mathrm{E}_x[(e_x^T\tilde{x}_t-y_t)^2-(e_t^T\tilde{x}_t-y_t)^2]. \qquad \text{(EC.24)}$$

This is because

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-y_t)^2-(e_x^T\tilde{x}_t-y_t)^2] &= \sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]+\mathrm{E}_x[(e_x^T\tilde{x}_t-y_t)\tilde{x}_t^T(e_t-e_x)] \\ &= \sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]+\mathrm{E}_x[(e_x^T\tilde{x}_t-y_t)\tilde{x}_t^T(e_t-e_x)] \\ &= \sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2],\end{aligned}$$

where the second line follows from the fact that $y_t=f(x_t)+\epsilon_t$ and $\mathrm{E}_x[\epsilon_t]=0$, ϵ_t independent of x_t, e_x, e_t, and the final line follows from the first order conditions of the minimization problem Eq (10), as given by Eq (11). Eq (EC.24) thus implies that $\sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t-e_x^T\tilde{x}_t)^2]$ is $O((m+1)\log(T(m+1))$.

To bound $\mathrm{E}_x[b^2 - \hat{b}_t^2])$ in Eq (EC.23), we can write

$$\begin{aligned}
\mathrm{E}_x[b^2 - \hat{b}_t^2] &= \mathrm{E}_x[(b - \hat{b}_t)(b + \hat{b}_t)]) \\
&\leq 2\bar{b}\mathrm{E}_x[|b - \hat{b}_t|] \\
&\leq 2\bar{b}\sqrt{\mathrm{E}[(\hat{b}_t - b)^2]} \\
&\leq 8\bar{b}\frac{\sqrt{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}}{\delta}\frac{1}{t^{1/3}},
\end{aligned}$$

where the second line follows from our assumed bounds on b and the projection step in the algorithm, the third line follows from Jensen's inequality since the function $x \mapsto x^2$ is convex, and the final line follows from Lemma EC.1. Then, $\sum_{t=1}^{T} \mathrm{E}_x([b^2 - \hat{b}_t^2]) \leq \frac{32}{3}\bar{b}\frac{\sqrt{\bar{f}^2+\sigma^2+\bar{b}^2 p_{\max}^2}}{\delta}T^{2/3}$, which dominates the $O((m+1\log(T(m+1))$ term $\sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)^2]$, and implies that Eq (EC.17) is $O(T^{2/3})$.

Similar ideas can be used to bound Eq (EC.18) and (EC.19). For Eq (EC.18), using the identity $y_t' - y_t = (b - \hat{b}_t)p_t$, we have

$$\begin{aligned}
\sum_{t=1}^{T}\mathrm{E}_x[(y_t' - y_t)(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)] &\leq 2\bar{b}p_{\max}\sum_{t=1}^{T}\mathrm{E}_x[|e_t^T\tilde{x}_t - e_x^T\tilde{x}_t|] \\
&\leq 2\bar{b}p_{\max}\sum_{t=1}^{T}\sqrt{\mathrm{E}_x[(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)^2]} \\
&\leq 2\bar{b}p_{\max}\sqrt{T}\sqrt{\sum_{t=1}^{T}\mathrm{E}_x[(e_t^T\tilde{x}_t - e_x^T\tilde{x}_t)^2]}.
\end{aligned}$$

The first line follows from our assumption on b, and that $\hat{b}$ and p_t are projections onto bounded sets. The second line follows from using Jensen's inequality again, and the final step follows from the Cauchy-Schwarz theorem. Then, applying Eq (EC.24) again, we see that Eq (EC.18) is $O(\sqrt{(m+1)T}\log(T(m+1)))$.

Finally, each term of Eq (EC.19) can be written as

$$\begin{aligned}
\mathrm{E}_x[y_t p_{t,g}^u(\hat{b}_t - b)] &= \mathrm{E}_x[f(x_t)p_{t,g}^u(\hat{b}_t - b)] \\
&\leq \frac{\bar{f}}{2\underline{b}}\mathrm{E}_x[(e_t^T\tilde{x}_t)(\hat{b}_t - b)] \\
&= \frac{\bar{f}}{2\underline{b}}\mathrm{E}_x[(e_t^T\tilde{x}_t - e^T\tilde{x}_t)(\hat{b}_t - b) + (e^T\tilde{x}_t)(\hat{b}_t - b)] \\
&\leq \frac{\bar{f}}{2\underline{b}}(2\bar{b}\mathrm{E}[|e_t^T\tilde{x}_t - e^T\tilde{x}_t|] + (\bar{a}+\bar{c})\mathrm{E}_x[|\hat{b}_t - b|]).
\end{aligned}$$

The second line follows from the definition of y_t and the fact that $\mathrm{E}[\epsilon_t] = 0$ and ϵ_t is independent from $p_{t,g}^u$ and $\hat{b}_t$. The final line follows from our assumption on b, and that $\hat{b}$ and p_t are projections onto bounded sets. We have already shown that $\sum_{t=1}^{T}\mathrm{E}[|e_t^T\tilde{x}_t - e^T\tilde{x}_t|]$ is $O((m+1\log(T(m+1))$, and that $\sum_{t=1}^{T}\mathrm{E}_x[|\hat{b}_t - b|]$ is $O(T^{2/3})$. Then Eq (EC.19) is $O(T^{2/3})$, which implies that the RPS algorithm is $O(T^{2/3})$ as well, thus concluding the proof.

Dependence on $m, \bar{a}, \bar{b}, \bar{c}$ and other parameters By combining constant factors, the expected regret of the RPS algorithm over T periods can be bounded by

$$\begin{aligned}
&O\left(\bar{b}\delta^2 + \frac{(\bar{a}+\bar{c})^2}{\underline{b}}\frac{\sqrt{\bar{f}^2+\sigma^2+\bar{b}^2p_{\max}^2}}{\delta}\left((\bar{a}+\bar{c})\bar{b} + \frac{1}{\underline{b}}\right)T^{2/3}\right) \\
&+ O\left(\sqrt{(m+1)T}\log(T(m+1)) + (m+1)\log(T(m+1))\right)
\end{aligned}$$

where the pre-factor in the first big O notation only contains an absolute constant. □

C.7. Lemmas

LEMMA EC.1 **(Bound on $\hat{b}_t$).** $\mathrm{E}[(\hat{b}_t - b)^2]$ *can be bounded as follows:*

- *When Algorithm 1 is applied to the IID setting, for $t \geq 4$, we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{\sqrt{t}}.$$

- *When Algorithm 2 is applied to the price ladder setting, for $t \geq 2$, we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 4 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{t^{2/3}}.$$

- *When Algorithm 3 is applied to the non IID setting, for $t \geq 4$ we have*

$$\mathrm{E}[(\hat{b}_t - b)^2] \leq 12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{t^{2/3}}.$$

Proof. Define the constant α_1 such that

$$\alpha_1 = \begin{cases} \frac{1}{4} & \text{in the IID setting,} \\ \frac{1}{6} & \text{in the price ladder setting,} \\ \frac{1}{6} & \text{in the non IID setting.} \end{cases}$$

We will first consider the **IID and non IID settings**, where prices are drawn from continuous price intervals at each time period. Using the definitions of b_t in Algorithms 1 and 3, $\hat{b}_t = \mathrm{Proj}(\hat{b}_t^u, B)$, where $\hat{b}_t^u = \frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1} \Delta p_s^2}$. Since the true parameter $b \in B$, we have

$$\begin{aligned}
\mathrm{E}[(\hat{b}_t - b)^2] &\leq \mathrm{E}[(\hat{b}_t^u - b)^2] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1} \Delta p_s^2} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_s) + \epsilon_s + b p_{g,s} + b \Delta p_s)}{\sum_{s=1}^{t-1} \Delta p_s^2} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_s) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \Delta p_s^2}\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_s) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1}}\right)^2\right].
\end{aligned}$$

In the last equality, we used the fact that $\Delta p_s^2 = \frac{\delta^2}{4} s^{-2\alpha_1}$.

Note that Δp_s's for all s are mutually independent, independent of x_s, and have mean 0, so

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_t) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1}}\right)^2\right] &= \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2 (f(x_s) + \epsilon_s + b p_{g,s})^2}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
&\leq \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3 \Delta p_s^2 (f(x_s)^2 + \epsilon_t^2 + b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1} \frac{\delta^2}{4} s^{-2\alpha_1})^2}\right] \\
&\leq 12 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\delta^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}} \qquad \text{(EC.25)}
\end{aligned}$$

We used the fact that $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$. In the last step, we used the definition that $\Delta p_s^2 = \frac{\delta^2}{4} s^{-2\alpha_1}$ and the assumption that $f(x_s), b, p_{g,s}$ are bounded.

Now consider the **price ladder setting**. Using the definitions of b_t in Algorithm 2, $\hat{b}_t = \mathrm{Proj}(\hat{b}_t^u, B)$, where $\hat{b}_t^u = \frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}$. Since the true parameter $b \in B$, we have

$$
\begin{aligned}
\mathrm{E}[(\hat{b}_t - b)^2] &\le \mathrm{E}[(\hat{b}_t^u - b)^2] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s D_s}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_s) + \epsilon_s + b p_{g,s} + b\Delta p_s)}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}} - b\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_s) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}\right)^2\right]
\end{aligned}
$$

The last line follows from the fact that $\mathrm{E}[\Delta p_s^2 | p_{g,t} = q_{i_s}] = (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}$.

As before, Δp_s's for all s are mutually independent, independent of x_s, and have mean 0, so

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{\sum_{s=1}^{t-1} \Delta p_s (f(x_t) + \epsilon_s + b p_{g,s})}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}\right)^2\right] &= \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} \Delta p_s^2 (f(x_s) + \epsilon_s + b p_{g,s})^2}{(\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1})^2}\right] \\
&\le \mathrm{E}\left[\frac{\sum_{s=1}^{t-1} 3\Delta p_s^2 (f(x_s)^2 + \epsilon_t^2 + b^2 p_{g,j}^2)}{(\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1})^2}\right] \\
&\le 3 \cdot \mathrm{E}\left[\frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\sum_{s=1}^{t-1}(q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}}\right] \\
&\le 3 \cdot \frac{\bar{f}^2 + \sigma^2 + \bar{b}^2 p_{\max}^2}{\underline{\delta}^2} \cdot \frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}}. \qquad \text{(EC.26)}
\end{aligned}
$$

We used the fact that $(x+y+z)^2 \le 3(x^2+y^2+z^2)$. The second to last step uses the definition that $\mathrm{E}[\Delta p_s^2 | p_{g,t} = q_{i_s}] = (q_{i_s} - q_{i_{s-1}})(q_{i_{s+1}} - q_{i_s}) s^{-2\alpha_1}$, and the assumption that $f(x_s), b, p_{g,s}$ are bounded. The last step uses the assumption that $q_i - q_{i-1} \ge \underline{\delta}$ for $i = 1, \ldots, N+1$.

Now for the **IID, non IID and price ladder settings**,

$$
\sum_{s=1}^{t-1} s^{-2\alpha_1} \ge \int_{y=1}^{t} y^{-2\alpha_1} dy = \frac{1}{1-2\alpha_1}(t^{1-2\alpha_1} - 1),
$$

and we have for $t \ge 4$ that

$$
\frac{1}{\sum_{s=1}^{t-1} s^{-2\alpha_1}} \le 2(1-2\alpha_1) t^{2\alpha_1 - 1}. \qquad \text{(EC.27)}
$$

Substituting (EC.27) into (EC.25) and (EC.26) respectively, we prove the lemma in the IID, price ladder and non IID settings. □

Lemma EC.2 **(Bound on $\|QM_t^{-1}Q\|_2$).** *Let $M = \mathrm{E}[\tilde{x}\tilde{x}^\intercal]$, $V = \mathrm{E}[(Q^{-1}\tilde{x}\tilde{x}^\intercal Q^{-1} - I)^2]$ and $M_t = \frac{1}{t-1}\sum_{s=1}^{t-1} \tilde{x}_s \tilde{x}_s^\intercal$. For any $t \ge 2$, M_t is invertible and $\|QM_t^{-1}Q\|_2 \le 2$ with probability at least*

$$
1 - 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).
$$

Proof. For any $s = 1, \ldots, t-1$, we have $\mathrm{E}[I - Q^{-1}\tilde{x}_s\tilde{x}_s^\top Q^{-1}] = 0$, where I is the identity matrix. In addition, for an arbitrary matrix A, it holds that $\|A\|_2 \le \|A\|_F$, so by $\|\tilde{x}_s\|_\infty \le 1$, we have

$$\begin{aligned}\lambda_{\max}(I - Q^{-1}\tilde{x}_s\tilde{x}_s^\top Q^{-1}) &\le \|I - Q^{-1}\tilde{x}_s\tilde{x}_s^\top Q^{-1}\|_2 \\ &\le \|Q^{-1}\|_2\|M - \tilde{x}_s\tilde{x}_s^\top\|_2\|Q^{-1}\|_2 \\ &\le \|Q^{-1}\|_2\|M - \tilde{x}_s\tilde{x}_s^\top\|_F\|Q^{-1}\|_2 \\ &\le \frac{1}{\sqrt{\lambda_{\min}(M)}} \cdot 2(m+1) \cdot \frac{1}{\sqrt{\lambda_{\min}(M)}} = \frac{2(m+1)}{\lambda_{\min}(M)}.\end{aligned}$$

Note that we used the submultiplicative property of the spectral norm. Since $\{\tilde{x}_s\}$ are independent and identically distributed, we apply the matrix Bernstein bound (Lemma EC.4) with $\alpha = (t-1)/2$ to yield

$$\begin{aligned}\mathbb{P}\left[\lambda_{\max}\left(\sum_{s=1}^{t-1}\frac{I - Q^{-1}\tilde{x}_s\tilde{x}_s^\top Q^{-1}}{t-1}\right) > \frac{1}{2}\right] &\le (m+1)\exp\left(-\frac{t^2/2}{\|(t-1)V\|_2 + 2(m+1)t/(3\lambda_{\min}(M))}\right) \\ &= (m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).\end{aligned}$$

By an identical argument, we also have

$$\mathbb{P}\left[\lambda_{\max}\left(-\sum_{s=1}^{t-1}\frac{I - Q^{-1}\tilde{x}_s\tilde{x}_s^\top Q^{-1}}{t-1}\right) > \frac{1}{2}\right] \le (m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).$$

Thus we have

$$\begin{aligned}\mathbb{P}[\|I - Q^{-1}M_tQ^{-1}\|_2 > \frac{1}{2}] &= \mathbb{P}[\max\{\lambda_{\max}(I - Q^{-1}M_tQ^{-1}), \lambda_{\max}(Q^{-1}M_tQ^{-1} - I)\} > \frac{1}{2}] \\ &\le 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right).\end{aligned} \tag{EC.28}$$

We can write $Q^{-1}M_tQ^{-1} = I + (Q^{-1}M_tQ^{-1} - I)$, then by Weyl's inequality,

$$\begin{aligned}\lambda_{\min}(Q^{-1}M_tQ^{-1}) &\ge \lambda_{\min}(I) + \lambda_{\min}(Q^{-1}M_tQ^{-1} - I) \\ &\ge 1 - \|Q^{-1}M_tQ^{-1} - I\|_2\end{aligned}$$

By Eq (EC.28), with probability at least

$$1 - 2(m+1)\exp\left(-\frac{3\lambda_{\min}(M)(t-1)}{24\lambda_{\min}(M)\|V\|_2 + 8(m+1)}\right),$$

we have $\lambda_{\min}(Q^{-1}M_tQ^{-1}) \ge 1/2$. Since $Q^{-1}M_tQ^{-1} = Q^{-1}\frac{\sum_{s=1}^{t-1}\tilde{x}_s\tilde{x}_s^\top}{t-1}Q^{-1}$ is positive semidefinite, $\lambda_{\min}(Q^{-1}M_tQ^{-1}) > 0$ implies that it is invertible. Then

$$\|QM_t^{-1}Q\|_2 = \frac{1}{\lambda_{\min}(Q^{-1}M_tQ^{-1})} \le 2.$$

This proves the lemma. □

LEMMA EC.3 **(Optimal Policy Structure for Linear Demand)**. *Suppose the true demand function is linear, given by*

$$D_t(p) = a + bp + c^\top x_t + \epsilon.$$

Then, it is optimal for the seller to use a linear pricing policy of the form $p_t = S_t + (U_t)^\top x_t$, *where* S_t *and* U_t *are measurable with respect to* $\mathcal{H}_{t-1}$.

Proof. Suppose the seller uses a pricing policy $\pi(\mathcal{H}_{t-1}, x_t) = \pi_t(x_t)$ at period t, where function $\pi_t(\cdot)$ is measurable with respect to t and could be nonlinear. We denote by $\tilde{\mathbb{E}}[\cdot]$ the conditional expectation operator $\mathrm{E}[\cdot \mid \mathcal{H}_{t-1}]$. Let S and U be the optimal solution of the following least squares problem:

$$\max_{s\in\mathbb{R}, u\in\mathbb{R}^m} \tilde{\mathbb{E}}\left[\left(\pi_t(x_t) - s - u^\top x_t\right)^2\right].$$

Clearly, S and U are measurable with respect to $\mathcal{H}_{t-1}$. By the first order condition, the optimal solution (S, U) satisfies

$$\tilde{\mathbb{E}}[\pi_t(x_t) - S - U^\top x_t] = 0, \quad \tilde{\mathbb{E}}\left[x_t\left(\pi_t(x_t) - S - U^\top x_t\right)\right] = 0. \tag{EC.29}$$

Now, let us compare the conditional expected revenue of price $\pi_t(x_t)$ and price $S + U^\top x_t$. We have

$$\begin{aligned}
&\tilde{\mathbb{E}}\left[\pi_t(x_t) D_t(\pi_t(x_t)) - (S + U^\top x_t) D_t(S + U^\top x_t)\right] \\
=&\tilde{\mathbb{E}}\left[\pi_t(x_t)\cdot\left(a + b\pi_t(x_t) + c^\top x_t\right) - (S + U^\top x_t)(a + b\cdot(S + U^\top x_t) + c^\top x_t)\right] \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t))^2 - (S + U^\top x_t)^2\right] + \tilde{\mathbb{E}}\left[(a + c^\top x_t)(\pi_t(x_t) - S - U^\top x_t)\right] && \text{(EC.30)} \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t))^2 - (S + U^\top x_t)^2\right] \\
=&b\left\{\tilde{\mathbb{E}}\left[(\pi_t(x_t) - S - U^\top x_t)^2\right] + 2\tilde{\mathbb{E}}\left[(S + U^\top x_t)(\pi_t(x_t) - S - U^\top x_t)\right]\right\} && \text{(EC.31)} \\
=&b\tilde{\mathbb{E}}\left[(\pi_t(x_t) - S - U^\top x_t)^2\right] \le 0.
\end{aligned}$$

The second term of Eq (EC.30) and the second term of Eq (EC.31) are both zero because of the first order condition Eq (EC.29). In the last step, recall that the price sensitivity parameter $b < 0$.

By taking the expectation over history $\mathcal{H}_{t-1}$, we have

$$\mathrm{E}\left[\pi_t(x_t) D_t(\pi_t(x_t)) - (S + U^\top x_t) D_t(S + U^\top x_t)\right] \le 0,$$

so if $p_t = \pi_t(x_t)$ is a nonlinear pricing policy, it is dominated by a linear pricing policy $p_t = S + U^\top x_t$. $\square$

Lemma EC.4 **(Matrix Bernstein bound, Tropp (2012))**. *Consider a finite sequence X_k of independent, random, self-adjoint matrices with dimension d. Assume that each random matrix satisfies*

$$\mathrm{E}[X_k] = 0 \text{ and } \lambda_{\max}(X_k) \le R \text{ almost surely,}$$

then for all $t \ge 0$,

$$\mathbb{P}\left[\lambda_{\max}(\sum_k X_k) \ge t\right] \le d \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right) \text{ where } \sigma^2 = \|\sum_k \mathrm{E}[X_k^2]\|_2.$$

References

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434.