A Additional Technical Discussions

A.1 Relaxing Assumption 1 (f)

In Assumption 1 (f), we make an assumption on the range of the ratio $\frac{\beta^\top x}{\gamma^\top x}$ for all possible x. The assumption seems unnatural at the first glance. The main usage of this assumption is to ensure (i) the dual prices λ^* and λ_T^D to be closed-form and (ii) the primal optimal prices p_t^D and p_t^* to be feasible (Proposition 3). Now we discuss how to relax the assumption. A more conventional assumption for the dynamic pricing literature is to impose a range of allowable prices as follows. The assumption is well-justified from a practical perspective that there will be some constraints for the dynamic prices due to the product cost and market competition. Technically, this assumption makes the linear demand model more relevant in that we only require the linearity to hold for prices within a certain range.

Assumption 2. *All the prices p_t's have to be in the range of $[\underline{p}, \bar{p}]$.*

All the discussions in this subsection are based on Assumption 1 (a)-(e) and Assumption 2. Accordingly, the deterministic benchmark becomes

$$\begin{aligned} R_T^{\mathrm{DQP}}(x_1, x_2, \ldots, x_T) = \max_{p_1,\ldots,p_T \in [\underline{p},\bar{p}]} & \sum_{t=1}^T p_t \left(\beta^\top x_t + (\gamma^\top x_t) p_t\right), \\ \text{s.t.} & \sum_{t=1}^T \beta^\top x_t + (\gamma^\top x_t) p_t \le B, \end{aligned}$$

where we overload the notation of R_T^{DQP} and $x_1, ..., x_T$ are the realized sequence of covariates. From an analysis of the optimality condition, the dual price λ_T^D for the new problem is described by the following theorem.

Proposition 8. *The dual price λ_T^D is the zero of the following function*

$$h\left(\lambda, \{x_1, ..., x_T\}\right) = -bT + \sum_{t=1}^T \mathrm{proj}\left(\frac{\beta^\top x_t}{2} + \frac{\lambda \gamma^\top x_t}{2}, \left[\beta^\top x_t + (\gamma^\top x_t)\underline{p}, \beta^\top x_t + (\gamma^\top x_t)\bar{p}\right]\right),$$

and when the function has no zero on $\mathbb{R}^+$, $\lambda_T^D = 0$. Here the projection function is defined by

$$\mathrm{proj}(z, [\underline{z}, \bar{z}]) = \begin{cases} \underline{z}, & \text{if } z \le \underline{z}, \\ z, & \text{if } z \in (\underline{z}, \bar{z}), \\ \bar{z}, & \text{if } z \ge \bar{z}. \end{cases}$$

Moreover, the function $h\left(\lambda, \{x_1, ..., x_T\}\right)$ is a non-increasing function of λ.

To interpret the proposition, recall that the unconstrained optimal demand at time t is $\frac{\beta^\top x_t}{2}$ and the dual price adjusts the demand downward. The function of $h(\cdot)$ comes from the bindingness of the constraint and it essentially says that under the dual price λ_T^D, the inventory should be exhausted. And when the inventory cannot be exhausted even when $\lambda_T^D = 0$, then the constraint is non-binding and $\lambda_T^D = 0$. The projection operator in the above proposition originates from the extra constraint on the price range of $[\underline{p}, \bar{p}]$. Importantly, this projection step makes the optimal dual price not closed form in terms of the covariates. Similarly, we adjust the definition of the dual price λ^* for the Lagrangian. Let λ^* be the zero of

$$h(\lambda) = -b + \mathbb{E}\left[\text{proj}\left(\frac{\beta^\top \boldsymbol{X}}{2} + \frac{\lambda\gamma^\top \boldsymbol{X}}{2}, [\beta^\top \boldsymbol{X} + (\gamma^\top \boldsymbol{X})\underline{p}, \beta^\top \boldsymbol{X} + (\gamma^\top \boldsymbol{X})\bar{p}]\right)\right]$$

where the expectation is taken with respect to the covariates $\boldsymbol{X}$'s distribution. When the function has no zero on $\mathbb{R}^+$, we set $\lambda^* = 0$.

Based on the new definitions of λ_D^T and λ^*, we have the following upper bound for the regret as an extension of Corollary 1. Compared to Corollary 1, the prices of p_t and p_t^* are restricted to the interval of $[\underline{p}, \bar{p}]$. The proof of the corollary follows a similar argument as that of Proposition 4 and Corollary 1. The corollary provides a similar insight that to achieve a good algorithm performance, one needs to (i) balance the inventory consumption process and (ii) to set prices close to p_t^*.

Corollary 2. *For any B_t-adapted stopping time τ, if $\mathbb{P}(\tau \le \tau_{\bar{M}_\beta}) = 1$, then there exists a constant $\bar{C} > 0$ (not dependent on T) such that*

$$\mathbb{E}R_T^L(x_{1:T}, \lambda^*) - R_T(\pi) \le \bar{C}\mathbb{E}\left[(T - \tau + 1) + \underline{B} + \sum_{t=1}^{\tau-1}(p_t - p_t^*)^2\right] + \bar{C}$$

holds for all $(\beta, \gamma) \in \Theta$, $\mathcal{P} \in \Xi$, and policy $\pi \in \Pi_{na}$. Here p_t is the price specified by the policy π, and $p_t^* = \text{proj}\left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda^*}{2}, [\underline{p}, \bar{p}]\right)$.

The non-closed-form expression of the dual variables brings minor additional computational complexity. Specifically, the function $h(\lambda)$ is a monotonous function so the search of zero can be done via a binary search procedure. In addition, at each time period, we need to project the proposed prices in Algorithm 1 and 2 into the interval of $[\underline{p}, \bar{p}]$. The rest part is identical to the previous algorithm, and the detailed algorithms (Algorithm 3 and Algorithm 4) are provided in Appendix A.3. We remark that the analyses and regret bounds remain similar to the previous case. Specifically, there are mainly two keys for the regret analysis: (i) the gap between λ_D^T and λ^*; (ii) the stability of the resource consumption process. For (i), though both the dual prices are not in closed form, their gap can still be established and bounded by the distance between the empirical distribution and the true distribution of the covariates (as these two distributions distinguish the two equations of $h(\lambda, \{x_1, ..., x_T\})$ and $h(\lambda_D^T)$). For (ii), the analyses of the resource consumption process of both algorithms do not change as the key argument is a martingale analysis where the martingale property is ensured by the optimality condition for the dual variable, and it will not change when we restrict the price to $[\underline{p}, \bar{p}]$.

Unbounded demand shocks and non-binding constraint

In Assumption 1 (d), we require the demand shock ϵ_t to be bounded. This assumption can be relaxed by a sub-Gaussianity on the demand shock. This will not change the results in Section 3 and Section 4 as they are all based on the analyses of the expected revenues. For the algorithm analyses, we need to use the sub-Gaussian version of the concentration inequalities in Proposition 4 and Proposition 6 when bounding $\mathbb{E}[T-\tau]$ and $\mathbb{E}[\underline{B}]$. Essentially, the analyses of these two quantities rely on a concentration argument for the demand shock term, and this concentration of demand shocks will lead to the demand consumption concentrating around its expectation. The concentration property does not rely on the boundedness of the demand shock ϵ_t and can also hold under sub-Gaussian distribution. Specifically, the sub-Gaussianity parameter will replace the variance σ in Assumption 1 (d). One caveat is that the sub-Gaussian structure may result in an unbounded demand shock and consequently a negative demand. This negativeness issue is usually ignored in the analysis for simplicity as noted by (Keskin and Zeevi, 2014).

Other than the closed-form property, Assumption 1 (f) ensures both λ^* and λ_T^D to be positive. Once we relax the assumption as above, both dual variables can be zero and in that case, the inventory constraint becomes non-binding (we exclude the degenerate case when the dual variable is zero and the constraint is binding). Intuitively, the problem then reduces to the unconstrained dynamic pricing problem. In the analyses of the constraint process for both algorithms, it essentially establishes that the average remaining resource $\frac{B_t}{T-t+1}$ will stay within an interval of $[b-\delta, b+\delta]$ until the very end of the horizon with a high probability. The same argument will establish that if the inventory constraint is non-binding initially, it will remain so throughout the horizon with high probability, and thus it verifies the reduction to the unconstrained dynamic pricing problem.

A.2 Generalized Linear Demand Model

The results in the paper can also be extended to the case of the generalized linear demand model (Ban and Keskin, 2021). Consider the following demand model,

$$D_t = g\left(\beta^\top x_t + (\gamma^\top x_t)p_t\right) + \epsilon_t$$

where the link function $g(\cdot)$ satisfies the following condition. The condition aims to ensure the identifiability of (β, γ) in the unknown setting. In particular, the discussion of the generalized model usually also comes with the restricted price (Assumption 2).

Assumption 3 (Properties of $g(\cdot)$)**.** *Assume $g(\cdot)$ is strictly increasing and differentiable, with a bounded* *derivative over its domain. Specifically, there exist constants* $\underline{g}, \bar{g} \in \mathbb{R}$ *such that* $0 < \underline{g} \le g'(z) \le \bar{g} < \infty$ *for all* $z = x^\top \beta + x^\top \gamma \cdot p$ *where* $x \in \mathcal{X}$, $\theta \in \Theta$, *and* $p \in [\underline{p}, \bar{p}]$.

Under Assumptions 2-3, the deterministic benchmark becomes

$$R_T^{\mathrm{DQP}}(x_1, x_2, \ldots, x_T) = \max_{p_1,\ldots,p_T \in [\underline{p},\bar{p}]} \sum_{t=1}^{T} p_t g\left(\beta^\top x_t + (\gamma^\top x_t) p_t\right),$$

$$\text{s.t.} \sum_{t=1}^{T} g(\beta^\top x_t + (\gamma^\top x_t) p_t) \le B.$$

The associated dual prices λ_T^D is determined by the zero of the following equation

$$\sum_{t=1}^{T} g(\beta^\top x_t + (\gamma^\top x_t) p_t^*) = B$$

where

$$p_t^* = \arg\max_{p \in [\underline{p},\bar{p}]} p \cdot g\left(\beta^\top x_t + (\gamma^\top x_t) p\right) - \lambda \cdot g\left(\beta^\top x_t + (\gamma^\top x_t) p\right).$$

Here when the equation has no zeros, $\lambda_T^D = 0$. The equation is an implicit function of λ, and Assumption 3 ensures that the left-hand-side is monotonous in λ. One additional caveat for this more general model is that at each time period, both algorithms need to solve the price optimization problem given certain dual price λ as above. The function can be non-convex for a general function of $g(\cdot)$. While the allowable price lies in a one-dimensional interval, the optimization problem can be approximately solved through a grid search.

As in the last subsection, the generalized linear demand model does not change the nature of the two algorithms. For both algorithms, we need to change the way that the dual prices are calculated, and for Algorithm 6, the parameter estimation of (β, γ) needs to be done through quasi-maximum likelihood estimation as in (Ban and Keskin, 2021). The analysis of the resource consumption process is the same as before. We remark that Assumption 3 ensures the parameter estimation of (β, γ) has the same order of error rate as the case of the linear demand model. When such continuity structure is violated, den Boer and Keskin (2020) consider the covariate-free setting and provides an elegant analysis of the piecewise-continuous case. In Assumption 3, the positiveness of the gradient essentially imposes some convexity structure for the demand function. When such structure is removed, the dynamic pricing problem (without the inventory constraint) is usually solved by a partitioning of the parameter space (Mao et al., 2018; Chen and Gallego, 2021) and $O(\sqrt{T})$ regret bound is usually no longer achievable. Modeling-wise, another direction to generalize the demand model is to have the parameters (β, γ) dependent on the covariates; Keskin et al. (2022b) explore such dependency through clustering of the covariates and allows different values of parameters (β, γ) for each cluster.

A.3 Algorithms for Section A.1 and Section A.2

Algorithm 3 Inventory-Driven Dynamic Pricing (IDP) Algorithm with Restricted Price

```
Input: $\beta$, $\gamma$, Distribution $\mathcal{P}$, $B$, $T$, price interval
Initialize $B_1 = B$
for $t = 1, ..., T$ do
    if $B_t = 0$ then
        Terminate the procedure
        break
    end if
    Solve the following equation to obtain $\lambda_t$
        $-B_t/(T-t+1) + \mathbb{E}\left[\text{proj}\left(\frac{\beta^\top \boldsymbol{X}}{2} + \frac{\lambda\gamma^\top \boldsymbol{X}}{2}, \left[\beta^\top \boldsymbol{X} + (\gamma^\top \boldsymbol{X})\underline{p}, \beta^\top \boldsymbol{X} + (\gamma^\top \boldsymbol{X})\bar{p}\right]\right)\right] = 0$
    when the equation has no zeros, set $\lambda_t = 0$
    Set the price
        $p_t = \text{proj}\left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda_t}{2}, [\underline{p}, \bar{p}]\right)$
    Execute with price $p_t$ and observe demand
        $D_t = \beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t$
        $d_t = D_t \wedge B_t$
    Update the remaining inventory
        $B_{t+1} = B_t - d_t$
end for
```

Algorithm 3 describes a variant of Algorithm 1 when the prices are restricted to the interval of $[\underline{p}, \bar{p}]$. It solves the optimality condition to obtain λ_t and utilizes λ_t to set the price at time t. This search procedure can be approximated by using a sampled empirical distribution to replace the true distribution of $\boldsymbol{X}$ and doing a binary search for the zero λ_t.

In a similar manner, Algorithm 4 changes Algorithm 2 in two ways: (i) first, the dual price is estimated based on the realized samples for the first $\lfloor\sqrt{T}\rfloor + 1$ time periods instead of being computed in a closed form; (ii) there is an extra projection step for the proposed price at each time t.

B Proofs of Section 3 and Section 4

B.1 Additional lemma and proof

The following lemma states that the function $f_{\beta,\gamma}(x)$ (frequently used in the later derivations) has μ-Lipshitz gradient.

Lemma 1. *Define function* $f_{\beta,\gamma} : \mathcal{X} \to \mathbb{R}$:

$$f_{\beta,\gamma}(x) \triangleq \frac{(2b - \beta^\top x)^2}{\gamma^\top x}.$$

Algorithm 4 First-Explore-Then-Exploit (FETE) Algorithm with Restricted Price

```
Input $B, T$
Initialize $B_1 = B$
for $t = 1, ..., \lfloor\sqrt{T}\rfloor + 1$ do
    Set the price
        $p_t = \begin{cases} z_1, & t(\mathrm{mod}2) = 1 \\ z_2, & t(\mathrm{mod}2) = 0 \end{cases}$
    Observe the demand $d_t$ and update $B_{t+1} = B_t - d_t$
end for
Obtain least square estimates $\hat{\beta}_0$ and $\hat{\gamma}_0$; then project on the parameter set $\Theta$
    $\left(\hat{\beta}_0, \hat{\gamma}_0\right) = \arg\min_{\beta,\gamma} \sum_{t=1}^{\lfloor\sqrt{T}\rfloor+1} \left(d_t - \beta^\top x_t - (\gamma^\top x_t)p_t\right)^2$
    $\left(\hat{\beta}, \hat{\gamma}\right) = \arg\min_{(x,y)\in\Theta} \left\|(x,y) - \left(\hat{\beta}_0, \hat{\gamma}_0\right)\right\|_2^2$
Solve the following equation
    $-b(\lfloor\sqrt{T}\rfloor + 1) + \sum_{t=1}^{\lfloor\sqrt{T}\rfloor+1} \mathrm{proj}\left(\frac{\beta^\top x_t}{2} + \frac{\lambda\gamma^\top x_t}{2}, \left[\beta^\top x_t + (\gamma^\top x_t)\underline{p}, \beta^\top x_t + (\gamma^\top x_t)\bar{p}\right]\right) = 0$
and obtain the dual price $\hat{\lambda}$. When the above equation has no solution, set $\hat{\lambda} = 0$.
for $t = \lfloor\sqrt{T}\rfloor + 2, ..., T$ do
    if $B_t = 0$ then
        Set $d_t = ... = d_T = 0$
        Set $p_k = -\frac{\beta^\top x_k}{2\gamma^\top x_k}$ for $k = t, ..., T$
        break
    end if
    Assign
        $p_t = \mathrm{proj}\left(-\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{\hat{\lambda}}{2}, [\underline{p}, \bar{p}]\right)$
    Execute with price $p_t$ and observe demand
        $D_t = \beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t$
        $d_t = D_t \wedge B_t$
    Update the remaining inventory
        $B_{t+1} = B_t - d_t$
end for
```

with $(\beta, \gamma) \in \Theta$. *Then, there exists a constant* $\mu_1 > 0$ *not dependent on* (β, γ) *such that*

$$-\mu\|x - x_0\|_2^2 \leq f_{\beta,\gamma}(x) - f_{\beta,\gamma}(x_0) - \nabla f_{\beta,\gamma}(x_0)^\top (x - x_0) \leq \mu\|x - x_0\|_2^2$$

for any $x, x_0 \in \mathcal{X}$.

Proof. Based on Assumption 1 we know $\gamma^\top x \leq -\underline{M}_\gamma$ is negative for all $(\beta, \gamma) \in \Theta$ and $x \in \mathcal{X}$. Also, the function $f_{\beta,\gamma}$ is second-order differentiable and it has well-defined gradient and Hessian. Consider that the sets Θ and $\mathcal{X}$ are bounded (Assumption 1), we can choose μ_1 to be the maximum of absolute eigenvalues of the Hessian matrices for all $(\gamma, \theta) \in \Theta$ and $x \in \mathcal{X}$, i.e.

$$\mu = \max_{(\beta,\gamma)\in\Theta} \max_{x\in\mathcal{X}} \left|\text{eigs}(\nabla^2 f_{\beta,\gamma}(x))\right|.$$

Given the compactness of Θ and $\mathcal{X}$, we know $\mu \in \mathbb{R}^+$ exists. Then,

$$-\mu\|x - x_0\|_2^2 \le f_{\beta,\gamma}(x) - f_{\beta,\gamma}(x_0) - \nabla f_{\beta,\gamma}(x_0)^\top (x - x_0) \le \mu\|x - x_0\|_2^2$$

□

B.2 Proof of Theorem 2

Proof. By spelling out the optimization problem $R_T^{\text{DQP}}(x_1, ..., x_T)$,

$$\begin{aligned} \max_{p_1,\dots,p_T} & \sum_{t=1}^{T} p_t \left(\beta^\top x_t + (\gamma^\top x_t) p_t\right), \\ \text{s.t.} & \sum_{t=1}^{T} \beta^\top x_t + (\gamma^\top x_t) p_t \le B, \end{aligned}$$

which is a linear-constrained quadratic program. Under Assumption 1 (b), we know the objective is concave. Then, the optimal solution $(p_{1:T}, \lambda)$ should satisfy the KKT conditions:

$$\beta^\top x_t + 2(\gamma^\top x_t) p_t - \lambda(\gamma^\top x_t) = 0,$$

$$\lambda \left(\sum_{t=1}^{T} \beta^\top x_t + (\gamma^\top x_t) p_t - B\right) = 0$$

where $\lambda \ge 0$ is the dual variable. From the KKT condition, we can solve the optimal solution. □

B.3 Proof of Proposition 1

Proof. By plugging in the optimal value of $p_k, ..., p_T$, we have

$$\begin{aligned} v_t(b_t, x_{t:T}) &= \sum_{k=t}^{T} \frac{-\beta^\top x_k + \lambda_{t:T}^D \gamma^\top x_k}{2\gamma^\top x_k} \cdot \left(\frac{1}{2}\beta^\top x_k + \frac{\lambda_{t:T}^D}{2}\gamma^\top x_k\right) \\ &= \sum_{k=t}^{T} \frac{(\lambda_{t:T}^D)^2}{4}\gamma^\top x_k - \frac{(\beta^\top x_k)^2}{4\gamma^\top x_k} \\ &= \sum_{k=t}^{T} \frac{\gamma^\top x_k}{4} \cdot \left(\frac{2b_t - \sum_{k=t}^{T} \beta^\top x_k}{\sum_{k=t}^{T} \gamma^\top x_k}\right)^2 - \sum_{k=t}^{T} \frac{(\beta^\top x_k)^2}{4\gamma^\top x_k}, \end{aligned}$$

where

$$\lambda_{t:T}^{D} = \frac{2b_t - \sum_{k=t}^{T} \beta^\top x_k}{\sum_{k=t}^{T} \gamma^\top x_k}$$

Given the negativeness of $\gamma^\top x_k$'s and $v_t(b_t, x_{t:T})$ quadratic in b_t, the function $v_t(b_t, x_{t:T})$ is concave in b_t. Moreover, the result holds for all $t = 1, ..., T$. $\square$

B.4 Proof of Theorem 3

Proof. The proof is based on backward induction. First, we verify that

$$V_T(b_T, x_T) \le v_T(b_T, x_T)$$

for all b_T and x_T. Indeed,

$$\begin{aligned} V_T(b_T, x_T) &= \max_{p \ge 0} \mathbb{E}[d_T p | x_T, b_T] \\ &= \max_{p \ge 0} \mathbb{E}\left[\left((\beta^\top x_T + (\gamma^\top x_T)p + \epsilon_T) \wedge b_T\right) p | x_T\right] \\ &\le \max_{p \ge 0} \left((\beta^\top x_T + (\gamma^\top x_T)p) \wedge b_T\right) p \\ &= v_T(b_T, x_T). \end{aligned}$$

In the above, the second line and fourth line are simply from the definition of V_T^* and v_T, respectively. Note that the expression in the second line is concave in ϵ_T, the third line comes from an application of Jensen's Inequality in regard to ϵ_T.

Assume the relation holds for $t + 1$,

$$\begin{aligned} V_t(b_t, x_{t:T}) &= \max_{p \ge 0} \mathbb{E}[d_t p | x_t, b_t] + \mathbb{E}[V_{t+1}(b_t - d_t, x_{t+1:T})] \\ &\le \max_{p \ge 0} \mathbb{E}[d_t p | x_t, b_t] + \mathbb{E}[v_{t+1}(b_t - d_t, x_{t+1:T})] \\ &\le \max_{p \ge 0} \mathbb{E}[d_t p | x_t, b_t] + v_{t+1}(b_t - \mathbb{E}[d_t | x_t, b_t], x_{t+1:T}) \\ &= v_t(b_t, x_{t:T}), \end{aligned}$$

where $d_t = b_t \wedge \left(\beta^\top x_t + (\gamma^\top x_t)p + \epsilon_t\right)$. The first line comes from Theorem 1. The second line comes from the induction assumption. The third line comes from the concavity of the function v_{k+1} (Proposition 1). The last line comes from the definition of the optimization problem in the deterministic regime.

$\square$

B.5 Proof of Proposition 2

Proof. For the problem $R_T^L(x_{1:T},\lambda^*)$, it is easy to verify that the maximum is achieved when

$$p_t = -\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda^*}{2}$$

for $t = 1, ..., T$. Then,

$$\begin{aligned} R_T^L(x_{1:T},\lambda^*) &= \sum_{t=1}^{T}\left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda^*}{2}\right)\left(\beta^\top x_t + \gamma^\top x_t \cdot \left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda^*}{2}\right)\right) \\ &\quad - \lambda^* \sum_{t=1}^{T}\left(\beta^\top x_t + \gamma^\top x_t \cdot \left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{\lambda^*}{2}\right) - b\right) \\ &= -\sum_{t=1}^{T}\frac{(\beta^\top x_t)^2}{4\gamma^\top x_t} - \lambda^*\sum_{t=1}^{T}\frac{\beta^\top x_t}{2} - (\lambda^*)^2\sum_{t=1}^{T}\frac{\gamma^\top x_t}{4} + bT\lambda^*. \end{aligned}$$

Taking expectation, also plugging in the definition of λ^*,

$$\mathbb{E}[R_T^L(x_{1:T},\lambda^*)] = \frac{(\lambda^*)^2 T\gamma^\top \mathbb{E}\boldsymbol{X}}{4} - \mathbb{E}\left[\sum_{t=1}^{T}\frac{(\beta^\top x_t)^2}{4\gamma^\top x_t}\right].$$

In a similar way, we can derive,

$$\mathbb{E}[R_T^{DQP}(x_{1:T})] = \mathbb{E}\left[\frac{(\lambda_T^D)^2}{4}\sum_{t=1}^{T}\gamma^\top x_t\right] - \mathbb{E}\left[\sum_{t=1}^{T}\frac{(\beta^\top x_t)^2}{4\gamma^\top x_t}\right].$$

Thus, the problem is reduced to show

$$\left|\frac{(\lambda^*)^2 T\gamma^\top \mathbb{E}\boldsymbol{X}}{4} - \mathbb{E}\left[\frac{(\lambda_T^D)^2}{4}\sum_{t=1}^{T}\gamma^\top x_t\right]\right| \le C_1.$$

By plugging in the definition of λ^* and λ_T^D, it is further reduced to

$$\left|\frac{\left(2B - T\beta^\top \mathbb{E}\boldsymbol{X}\right)^2}{4T\gamma^\top \mathbb{E}\boldsymbol{X}} - \mathbb{E}\left[\frac{\left(2B - \sum_{t=1}^{T}\beta^\top x_t\right)^2}{4\sum_{t=1}^{T}\gamma^\top x_t}\right]\right| \le C_1,$$

or equivalently,

$$\left|\frac{T\left(2b - \beta^\top \mathbb{E}\boldsymbol{X}\right)^2}{4\gamma^\top \mathbb{E}\boldsymbol{X}} - \mathbb{E}\left[\frac{T\left(2b - \sum_{t=1}^{T}\beta^\top x_t/T\right)^2}{4\sum_{t=1}^{T}\gamma^\top x_t/T}\right]\right| \le C_1. \tag{9}$$

Then, applying Lemma 1 with $x_0 = \mathbb{E}\boldsymbol{X}$, we obtain the following two inequalities

$$\frac{\left(2b - \sum_{t=1}^{T}\beta^\top x_t/T\right)^2}{4\sum_{t=1}^{T}\gamma^\top x_t/T} - \frac{\left(2b - \beta^\top \mathbb{E}\boldsymbol{X}\right)^2}{4\gamma^\top \mathbb{E}\boldsymbol{X}} - \frac{1}{4}\nabla f_{\beta,\gamma}(\mathbb{E}\boldsymbol{X})\left(\frac{\sum_{t=1}^{T}x_t}{T} - \mathbb{E}\boldsymbol{X}\right) \le \frac{\mu}{4}\left\|\frac{\sum_{t=1}^{T}x_t}{T} - \mathbb{E}\boldsymbol{X}\right\|_2^2,$$

$$\frac{\left(2b-\sum_{t=1}^{T}\beta^{\top}x_t/T\right)^2}{4\sum_{t=1}^{T}\gamma^{\top}x_t/T}-\frac{\left(2b-\beta^{\top}\mathbb{E}\boldsymbol{X}\right)^2}{4\gamma^{\top}\mathbb{E}\boldsymbol{X}}-\frac{1}{4}\nabla f_{\beta,\gamma}(\mathbb{E}\boldsymbol{X})\left(\frac{\sum_{t=1}^{T}x_t}{T}-\mathbb{E}\boldsymbol{X}\right)\geq-\frac{\mu}{4}\left\|\frac{\sum_{t=1}^{T}x_t}{T}-\mathbb{E}\boldsymbol{X}\right\|_2^2.$$

Taking expectation will cancel out the first order term, namely the last term on the left hand side of the above two inequalities. Then, we scale the inequalities with a factor of T and it matches the left hand side of (9)

$$\begin{aligned}\left|\frac{T\left(2b-\beta^{\top}\mathbb{E}\boldsymbol{X}\right)^2}{4\gamma^{\top}\mathbb{E}\boldsymbol{X}}-\mathbb{E}\left[\frac{T\left(2b-\sum_{t=1}^{T}\beta^{\top}x_t/T\right)^2}{4\sum_{t=1}^{T}\gamma^{\top}x_t/T}\right]\right| &\leq\frac{T\mu}{4}\mathbb{E}\left\|\frac{\sum_{t=1}^{T}x_t}{T}-\mathbb{E}\boldsymbol{X}\right\|_2^2\\ &=\frac{\mu}{4}\mathbb{E}\left\|\boldsymbol{X}-\mathbb{E}\boldsymbol{X}\right\|_2^2.\end{aligned}$$

Here we define C_1 according to the maximum of the last part above over all distributions $\mathcal{P}$,

$$C_1=\frac{\mu}{4}\cdot\max_{\boldsymbol{X}\sim\mathcal{P}\in\Xi}\mathbb{E}\left\|\boldsymbol{X}-\mathbb{E}\boldsymbol{X}\right\|_2^2$$

where μ is defined in Lemma 1. We know this maximum exists because of the compactness of $\mathcal{X}$. Then, (9) is true and Proposition 2 is proved. □

B.6 Proof of Theorem 4

Proof. To compare $R_T(\pi)$ and $R_T^L(x_{1:T},\lambda)$, we first introduce a surrogate for $R_T(\pi)$ that also takes a Lagrangian form. Define

$$\tilde{R}_T(\pi)\triangleq\sum_{t=1}^{T}p_td_t-\lambda^*\sum_{t=1}^{T}(d_t-b)\,.$$

On each realized trial,

$$\sum_{t=1}^{T}(d_t-b)=\sum_{t=1}^{T}d_t-B=-\underline{B}\leq 0.$$

We know

$$\tilde{R}_T(\pi)-R_T(\pi)=-\lambda^*\sum_{t=1}^{T}(d_t-b)=\lambda^*\underline{B},$$

and the fact that $\lambda^*\geq 0$, we have

$$\tilde{R}_T(\pi)-R_T(\pi)=\lambda^*\underline{B}\geq 0. \tag{10}$$

Next, we proceed to compare $\tilde{R}_T(\pi)$ and $R_T^L(x_{1:T},\lambda)$. Given the choice of $\overline{M}_\beta$ as the maximum potential market size and the stopping time $\tau_{\overline{M}_\beta}$. When $t<\tau_{\overline{M}_\beta}$, i.e., $B_t\geq\overline{M}_\beta$, it means the available inventory exceeds the maximal possible demand. Hence, the true demand will not be truncated by the inventory, that is,

$$\mathbb{E}[d_t|p_t,B_t]=\beta^{\top}x_t+(\gamma^{\top}x_t)\cdot p_t$$

when $B_t \geq \overline{M}_\beta$. Based on this fact, we analyze $\tilde{R}_T(\pi)$ by conditioning on the covariates $x_{1:T}$,

$$\begin{aligned}
\mathbb{E}\left[\tilde{R}_T(\pi)\middle|x_{1:T}\right] &= \mathbb{E}\left[\sum_{t=1}^{T}(p_t-\lambda^*)d_t\middle|x_{1:T}\right]+\lambda^*B \\
&= \sum_{t=1}^{T}\mathbb{E}\left[(p_t-\lambda^*)d_t\middle|x_{1:T}\right]+\lambda^*B \\
&= \sum_{t=1}^{T}\mathbb{E}\left[(p_t-\lambda^*)d_tI(\tau_{\overline{M}_\beta}>t)\middle|x_{1:T}\right]+\sum_{t=1}^{T}\mathbb{E}\left[(p_t-\lambda^*)d_tI(\tau_{\overline{M}_\beta}\leq t)\middle|x_{1:T}\right]+\lambda^*B \\
&= \sum_{t=1}^{T}\mathbb{E}\left[(p_t-\lambda^*)(\beta^\top x_t+(\gamma^\top x_t)\cdot p_t)I(\tau_{\overline{M}_\beta}>t)\middle|x_{1:T}\right]+\sum_{t=1}^{T}\mathbb{E}\left[(p_t-\lambda^*)d_tI(\tau_{\overline{M}_\beta}\leq t)\middle|x_{1:T}\right]+\lambda^*B \\
&= \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t-\lambda^*)(\beta^\top x_t+(\gamma^\top x_t)\cdot p_t)\middle|x_{1:T}\right]+\mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}(p_t-\lambda^*)d_t\middle|x_{1:T}\right]+\lambda^*B
\end{aligned} \tag{11}$$

The second line and last line come from the exchange of summation and expectation. In the middle part, we only need to notice that, as argued in above, the demand will not be truncated before time $\tau_{\overline{M}_\beta}$.

Now, we analyze the second term. On one hand, we have

$$\mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}(p_t-\lambda^*)d_t\middle|x_{1:T}\right]\geq -\lambda^*\mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}d_t|x_{1:T}\right]\geq -\lambda^*\overline{M}_\beta \tag{12}$$

which is due to the definition of $\tau_{\overline{M}_\beta}$ and the non-negativeness of the price and the demand. On the other hand,

$$\mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}(p_t-\lambda^*)d_t\middle|x_{1:T}\right]\leq \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}p_td_t\middle|x_{1:T}\right]\leq \bar{p}\overline{M}_\beta \tag{13}$$

where

$$\bar{p}=\max_{(\beta,\gamma)\in\Theta}\max_{x\in\mathcal{X}}-\frac{\beta^\top x}{\gamma^\top x}$$

is the maximum possible price to guarantee a non-negative expected demand. Specifically, the intuition for (13) is that the revenue generated from $\overline{M}_\beta$ units of inventory grows at most linearly with $\overline{M}_\beta$. The two inequalities (12) and (13) together tell that the revenue generated after $\overline{M}_\beta$ is $O(1)$.

Gluing together (10), (11), (12), (13), we obtain

$$\mathbb{E}[R_T(\pi)|x_{1:T}]\geq \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t-\lambda^*)(\beta^\top x_t+(\gamma^\top x_t)\cdot p_t)\middle|x_{1:T}\right]-\lambda^*\overline{M}_\beta+\lambda^*B-\lambda^*\underline{B},$$

$$\mathbb{E}[R_T(\pi)|x_{1:T}]\leq \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t-\lambda^*)(\beta^\top x_t+(\gamma^\top x_t)\cdot p_t)\middle|x_{1:T}\right]+\bar{p}\overline{M}_\beta+\lambda^*B.$$

Finally, for the lower bound on the difference,

$$
\begin{aligned}
\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) \geq\ & \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1} (p_t^* - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*) - \sum_{t=1}^{\tau_{\overline{M}_\beta}-1} (p_t - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t)\right] \\
&+ \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T} (p_t^* - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] - \bar{p}\overline{M}_\beta \\
=\ & \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1} (p_t^* - p_t)\left(\beta^\top x_t + \gamma^\top x_t(p_t^* + p_t) - \lambda^*\gamma^\top x_t\right)\right] \\
&+ \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T} (p_t^* - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] - \bar{p}\overline{M}_\beta \\
=\ & -\mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1} \frac{\gamma^\top x_t}{2}(p_t^* - p_t)^2\right] + \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T} (p_t^* - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] - \bar{p}\overline{M}_\beta
\end{aligned}
$$

The second term

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T} (p_t^* - \lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] &= \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T} -\frac{\gamma^\top x_t}{4}\left(\frac{\beta^\top x_t}{\gamma^\top x_t} + \lambda^*\right)^2\right] \\
&= \mathbb{E}\left[-\frac{\gamma^\top \boldsymbol{X}}{4}\left(\frac{\beta^\top \boldsymbol{X}}{\gamma^\top \boldsymbol{X}} + \lambda^*\right)^2\right]\cdot \mathbb{E}[T - \tau_{\overline{M}_\beta} + 1]. \qquad (14)
\end{aligned}
$$

where $\boldsymbol{X} \sim \mathcal{P}$ denotes the random variable for x_t's. Let

$$
C_{2,0} = \frac{\mathbb{E}\left[-\frac{\gamma^\top \boldsymbol{X}}{4}\left(\frac{\beta^\top \boldsymbol{X}}{\gamma^\top \boldsymbol{X}} + \lambda^*\right)^2\right]}{\mathbb{E}\left[\max\left\{(p_t^*)^2, \left(p_t^* + \frac{\beta^\top \boldsymbol{X}}{\gamma^\top \boldsymbol{X}}\right)^2\right\}\right]}
$$

and choose

$$
C_2' = \max\left\{\overline{M}_\beta, C_{2,0}, \bar{p}\overline{M}_\beta\right\}.
$$

We have

$$
\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) \geq C_2'\mathbb{E}\left[\sum_{t=1}^{T}(p_t^* - p_t)^2\right] - C_2'. \qquad (15)
$$

and it completes the lower bound part. For the upper bound part on the difference,

$$\begin{aligned}
\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) &\geq \mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t^*-\lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*) - \sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t-\lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t)\right] \\
&\quad + \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}(p_t^*-\lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] + \lambda^*\overline{M}_\beta + \lambda^*\mathbb{E}[\underline{B}] \\
&= -\mathbb{E}\left[\sum_{t=1}^{\tau_{\overline{M}_\beta}-1}\frac{\gamma^\top x_t}{2}(p_t^*-p_t)^2\right] + \mathbb{E}\left[\sum_{t=\tau_{\overline{M}_\beta}}^{T}(p_t^*-\lambda^*)(\beta^\top x_t + (\gamma^\top x_t)\cdot p_t^*)\right] + \lambda^*\overline{M}_\beta + \lambda^*\mathbb{E}[\underline{B}]
\end{aligned}$$

With the same analysis of the second term as in (14), we choose

$$C_2 = \max \left\{\overline{M}_\beta, \mathbb{E}\left[-\frac{\gamma^\top \boldsymbol{X}}{4}\left(\frac{\beta^\top \boldsymbol{X}}{\gamma^\top \boldsymbol{X}} + \lambda^*\right)^2\right], \lambda^*\overline{M}_\beta, \lambda^*\right\}.$$

Then, we have

$$\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) \leq C_2\mathbb{E}\left[(T-\tau_{\overline{M}_\beta}+1) + \underline{B} + \sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t^*-p_t)^2\right] + C_2. \tag{16}$$

Thus the upper bound part is obtained and the proof is completed with (15) and (16). □

B.7 Proof of Corollary 1

Proof. From Theorem 4, we know that

$$\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) \leq C_2\mathbb{E}\left[(T-\tau_{\overline{M}_\beta}+1) + \underline{B} + \sum_{t=1}^{\tau_{\overline{M}_\beta}-1}(p_t-p_t^*)^2\right] + C_2.$$

When $\tau \leq \tau_{\overline{M}_\beta}$, so,

$$\mathbb{E}R_T^L(x_{1:T},\lambda^*) - \mathbb{E}R_T(\pi) \leq C_2\mathbb{E}\left[(T-\tau_{\overline{M}_\beta}+1) + \underline{B} + \sum_{t=1}^{\tau-1}(p_t-p_t^*)^2 + \sum_{t=\tau}^{\tau_{\overline{M}_\beta}-1}(p_t-p_t^*)^2\right] + C_2.$$

In particular, the term

$$\mathbb{E}\left[\sum_{t=\tau}^{\tau_{\overline{M}_\beta}-1}(p_t-p_t^*)^2\right] \leq \mathbb{E}[\tau_{\overline{M}_\beta}-\tau]\cdot\mathbb{E}\left[\max_{p\in\left[0,-\frac{\beta^\top x_t}{\gamma^\top x_t}\right]}(p-p_t^*)^2\right].$$

With

$$C_2'' = \max\left\{C_2, \mathbb{E}\left[\max_{p\in\left[0,-\frac{\beta^\top x_t}{\gamma^\top x_t}\right]} (p-p_t^*)^2\right]\right\} = \max\left\{C_2, \mathbb{E}\left[\max\left\{(p_t^*)^2, \left(p_t^* + \frac{\beta^\top \boldsymbol{X}}{\gamma^\top \boldsymbol{X}}\right)^2\right\}\right]\right\},$$

the result follows. □

C Proofs of Section 5 and Section 6

C.1 Proof of Proposition 4

Proof. With this definition, τ' is a B_t-adapted stopping time. For any $t < \tau'$, the price p_t set by Algorithm 1 is well defined, i.e., between 0 and $-\frac{\beta^\top x_t}{\gamma^\top x_t}$ and the demand d_t will not be truncated, i.e. $d_t = D_t$. We introduce a new inventory process B_t'. Define

$$B_t' = \begin{cases} B_t, & t < \tau' \\ \frac{(T-t+1)B_{t-1}'}{T-t+2}, & \tau' \le t. \end{cases}.$$

This new process takes the same value as B_t before τ' and it freezes the quantity $\frac{B_t'}{T-t+1}$ from the time τ' on. We are going to use this quantity to study $\mathbb{E}\tau'$ and $\mathbb{E}\underline{B}$. Given the definition of B_t, we know

$$\begin{aligned}\frac{B_{t+1}'}{T-t} &= \frac{B_t'}{T-t+1} I(t\ge\tau') + \frac{1}{T-t} I(t<\tau')\left(B_t' - D_t\right)\\ &= \frac{B_t'}{T-t+1} I(t\ge\tau') + \frac{1}{T-t} I(t<\tau')\left(B_t' - \left(\beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t\right)\right),\\ &= \frac{B_t'}{T-t+1} I(t\ge\tau') + \frac{B_t'}{T-t+1} I(t<\tau') + \frac{1}{T-t}\left(\frac{B_t'}{T-t+1} - \left(\beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t\right)\right) I(t<\tau')\end{aligned}$$

where p_t is specified by Algorithm 1. The second line is because when $t < \tau'$, the process B_t' evolves the same as the process B_t and the demand will not be truncated. Consider

$$\begin{aligned}&\mathbb{E}\left[\frac{B_t'}{T-t+1} - \left(\beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t\right) \Big| B_t, t<\tau'\right]\\ &= \mathbb{E}\left[\frac{B_t'}{T-t+1} - \left(\beta^\top x_t + (\gamma^\top x_t)\left(-\frac{\beta^\top x_t}{2\gamma^\top x_t} + \frac{2B_t'/(T-t+1) - \beta^\top \mathbb{E}\boldsymbol{X}}{2\gamma^\top \boldsymbol{X}}\right)\right) \Big| B_t, t<\tau'\right]\\ &= 0.\end{aligned}$$

We know the process $\{M_t\}_{t=1}^T$,

$$Y_t = \frac{B_t'}{T-t+1}$$

is a martingale adapted to the filtration $\mathcal{F}_t = \{(x_t, \epsilon_t)\}_{s=1}^{t-1}$. Also, we know

$$
\begin{aligned}
|Y_{t+1} - Y_t| &\leq \frac{1}{T-t} \left| \frac{B'_t}{T-t+1} - \left(\beta^\top x_t + (\gamma^\top x_t) p_t + \epsilon_t\right) \right| \\
&\leq \frac{1}{T-t} \left(\bar{b} + \overline{M}_\beta\right)
\end{aligned}
$$

for $t = 1, ..., T-1$. Then, by applying Azuma–Hoeffding inequality for the martingale $\{Y_t\}_{t=1}^T$, we know

$$
\begin{aligned}
\mathbb{P}\left(|Y_t - Y_1| \geq \kappa\right) &\leq 2\exp\left(\frac{-\kappa^2}{2\sum_{s=1}^{t-1} \left(\frac{1}{T-t} \left(\bar{b} + \overline{M}_\beta\right)\right)^2} \right) \\
&\leq 2\exp\left(\frac{-\kappa^2 (T-t)}{(\bar{b} + \overline{M}_\beta)^2} \right)
\end{aligned}
$$

for all $t = 1, ..., T$ and $\kappa > 0$.

Also, we know that when $t \leq \tau'$, $B_t = B'_t$. Therefore,

$$
\{\tau' \leq t\} \subset \left\{ \left| \frac{B'_t}{T-t+1} - b \right| \geq \delta_t \right\} = \{|Y_t - Y_1| \geq \delta_t\}
$$

where

$$
\delta_t = \min\left\{ b - \underline{b}, \bar{b} - b, b - \frac{\bar{M}_\beta}{T-t+1} \right\}.
$$

The third term in the definition of δ_t is negligible until the very end of the horizon where t is close to $T-1$. Let $\delta = \min\{b - \underline{b}, \bar{b} - b\}$. Applying the Chebyshev's Inequality,

$$
\mathbb{P}(\tau' \leq t) \leq \mathbb{P}\left(|Y_t - Y_1| \geq \delta_t\right) \leq 2\exp\left(\frac{-\delta^2 (T-t)}{(\bar{b} + \overline{M}_\beta)^2} \right).
$$

Consequently,

$$
\begin{aligned}
\mathbb{E}[T - \tau'] &= \sum_{t=1}^{T} \mathbb{P}(\tau' \leq t) \\
&\leq \sum_{t=1}^{T} 2\exp\left(\frac{-\delta^2 (T-t)}{(\bar{b} + \overline{M}_\beta)^2} \right) \leq \frac{2}{1 - \exp\left(\frac{-\delta^2}{(\bar{b} + \overline{M}_\beta)^2} \right)}
\end{aligned}
$$

Additionally, we have

$$
\underline{B} \leq B_{\tau'-1} = B'_{\tau'-1}.
$$

So,

$$
\mathbb{E}\underline{B} \leq \mathbb{E}[B'_{\tau'-1}] \leq \mathbb{E}[\bar{b}(T - \tau' + 2)] \leq \frac{2\bar{b}}{1 - \exp\left(\frac{-\delta^2}{(\bar{b} + \overline{M}_\beta)^2} \right)} + 2\bar{b}.
$$

□

C.2 Proof of Theorem 5

Proof. According to Proposition 4 and Corollary 1, we only need to show there exists a constant C_3',

$$\mathbb{E}\left[\sum_{t=1}^{\tau'-1}(p_t - p_t^*)^2\right] \leq C_3' \log T$$

where p_t is specified by π_{IDP}. When $t < \tau'$, in the proof of Proposition 4, we show that

$$\begin{aligned}
\mathbb{E}\left(\frac{B_t}{T-t+1} - b\right)^2 &\leq \int_0^\infty \kappa^2 d\mathbb{P}\left(|Y_t - Y_1| \geq \kappa\right) \\
&\leq \int_0^\infty \frac{4\kappa^3(T-t)}{(\bar{b}+\overline{M}_\beta)^2} \exp\left(\frac{-\kappa^2(T-t)}{(\bar{b}+\overline{M}_\beta)^2}\right) d\kappa \\
&\leq \frac{C_3'}{T-t}
\end{aligned}$$

for constant C_3' not dependent on T or t. In Algorithm 1,

$$\begin{aligned}
\mathbb{E}(p_t - p_t^*)^2 &= \frac{1}{4}\mathbb{E}(\lambda_t - \lambda^*)^2 \\
&= \frac{1}{\left(\gamma^\top \mathbb{E}\boldsymbol{X}\right)^2}\mathbb{E}\left(\frac{B_t}{T-t+1} - b\right)^2 \\
&\leq \frac{2C_3'}{(T-t)\left(\gamma^\top \mathbb{E}\boldsymbol{X}\right)^2} \leq \frac{2C_3'}{(T-t)K_4^2}.
\end{aligned}$$

Taking the summation, we obtain

$$\mathbb{E}\left[\sum_{t=1}^{\tau'-1}(p_t - p_t^*)^2\right] \leq \frac{2C_3' \log T}{K_4^2}. \tag{17}$$

Therefore, by choosing $C_3 = \frac{2C_2'' C_3' \log T}{K_4^2}$, the proof is completed by combining Corollary 1 and Proposition 4 with (17). $\square$

C.3 Proof of Proposition 5

Proof. Given that the exploration phase adopts the same one as in the papers (Keskin and Zeevi, 2014; Ban and Keskin, 2021). Specifically, the two-price switching part in FETE algorithm is the same as the strategy in ILSX algorithm in (Keskin and Zeevi, 2014) and ISL-d algorithm in Besbes and Zeevi (2009). A proof for the convergence results on β and γ can be directly implied from the related results in (Keskin and Zeevi, 2014; Ban and Keskin, 2021). Lastly, for the sample-averaging estimator $\hat{X}$, notice

$$\mathbb{E}\left\|\hat{X} - \mathbb{E}\boldsymbol{X}\right\|_2^2 = \frac{1}{T}\mathrm{tr}(\mathrm{cov}(\boldsymbol{X})).$$

and hence,

$$\mathbb{E}\left\|\hat{X} - \mathbb{E}\boldsymbol{X}\right\|_2 = \frac{1}{\sqrt{T}}\sqrt{\mathrm{tr}(\mathrm{cov}(\boldsymbol{X}))}.$$

□

C.4 Proof of Proposition 6

Proof. First, for $\lfloor\sqrt{T}\rfloor + 2 \le t \le \tau'$,

$$\begin{aligned}\mathbb{E}[p_t|x_{1:t}] &= \mathbb{E}\left[-\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{\hat{\lambda}}{2}\Big|x_{1:t}\right]\\ &= \mathbb{E}\left[-\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{2b - \hat{\beta}^\top\hat{X}}{2\hat{\gamma}^\top\hat{X}}\Big|x_{1:t}\right].\end{aligned}$$

Define function $g : \Theta$

$$g_{x'}(\beta,\gamma,x) = -\frac{\beta x'}{\gamma x'} + \frac{2b - \beta x}{2\gamma x}.$$

With the same approach as in Lemma 1, we can show that there exists μ_2, such that

$$g_{x'}(\beta_1,\gamma_1,x_1) - g_{x'}(\beta_2,\gamma_2,x_2) - \nabla g_{x'}(\beta_2,\gamma_2,x_2)^\top(\beta_1-\beta_2,\gamma_1-\gamma_2,x_1-x_2) \le \mu_3\left(\|\beta_1-\beta_2\|_2^2 + \|\gamma_1-\gamma_2\|_2^2 + \|x_1-x_2\|_2^2\right)$$

$$g_{x'}(\beta_1,\gamma_1,x_1) - g_{x'}(\beta_2,\gamma_2,x_2) - \nabla g_{x'}(\beta_2,\gamma_2,x_2)^\top(\beta_1-\beta_2,\gamma_1-\gamma_2,x_1-x_2) \ge -\mu_3\left(\|\beta_1-\beta_2\|_2^2 + \|\gamma_1-\gamma_2\|_2^2 + \|x_1-x_2\|_2^2\right).$$

hold for all $x' \in \mathcal{X}$ and $(\beta_1,\gamma_1,x_1),(\beta_2,\gamma_2,x_2) \in \Theta\times\mathcal{X}$. In fact, we can simply take μ_3 to be the largest absolute eigenvalue over all $x' \in \mathcal{X}$ and $(\beta_2,\gamma_2,x_2) \in \Theta\times\mathcal{X}$.

Then, by setting $x_t = x'$, $(\beta_1,\gamma_1,x_1) = (\hat{\beta},\hat{\gamma},\hat{X})$ and $(\beta_2,\gamma_2,x_2) = (\beta,\gamma,\mathbb{E}\boldsymbol{X})$, we obtain

$$\begin{aligned}\mathbb{E}[p_t - p_t^*|x_t] &= \mathbb{E}[g_{x_t}(\hat{\beta},\hat{\gamma},\hat{X}) - g_{x_t}(\beta,\gamma,\mathbb{E}\boldsymbol{X})|x_t]\\ &\le \mu_3\mathbb{E}\left[\|\hat{\beta}-\beta\|_2^2 + \|\hat{\gamma}-\ \gamma\|_2^2 + \|\hat{X}-\mathbb{E}\boldsymbol{X}\|_2^2\Big|x_t\right]\\ &\le \frac{3\mu_3 C_4\log T}{\sqrt{T}}.\end{aligned}$$

where the second line comes from the unbiasedness of linear regression and sample mean estimators to remove the first-order terms, and the third line comes Proposition 5. In the same way, we can obtain,

$$\mathbb{E}[p_t - p_t^*|x_t] \ge -\frac{3\mu_3 C_4\log T}{\sqrt{T}}.$$

Now, we compute $\mathbb{E}[T - \tau_{\bar{M}_\beta}]$ by analyzing the dynamics of B_t. By the definition of $\tau_{\bar{M}_\beta}$, when $t \le \tau_{\bar{M}_\beta}$, the demand $d_t = D_t$ will not be truncated. Consider the sequence of untruncated demand $\{D_t\}_{t=1}^T$ such that

$$p_t = -\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{\hat{\lambda}}{2}$$

$$D_t = \beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t.$$

We know that

$$\left\{B_t > \bar{M}_\beta\right\} = \{\tau \geq t+1\}$$

Then

$$\left\{bT - \sum_{k=1}^{t-1} D_k > \bar{M}_\beta\right\} = \{\tau \geq t+1\}.$$

This is true because when $\tau \geq t+1$, $B_t = bT - \sum_{k=1}^{t-1} D_k$. Then,

$$\left\{bT - K_3(\lfloor\sqrt{T}\rfloor + 1) - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k > \bar{M}_\beta\right\} \subset \left\{bT - \sum_{k=1}^{t-1} D_k > \bar{M}_\beta\right\} = \{\tau \geq t+1\}.$$

Next, we analyze the mean and variance of the demand sum,

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right] &= \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}\left[D_k - b\right] + b(t - \lfloor\sqrt{T}\rfloor - 2) \\
&= \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}\left[\beta^\top x_t + (\gamma^\top x_t)p_t + \epsilon_t - \beta^\top x_t - (\gamma^\top x_t)p_t^* - \epsilon_t\right] + b(t - \lfloor\sqrt{T}\rfloor - 2) \\
&= \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[(\gamma^\top x_t)(p_t - p_t^*)] + b(t - \lfloor\sqrt{T}\rfloor - 2) \\
&= \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[(\gamma^\top x_t)\mathbb{E}[p_t - p_t^*|x_t]] + b(t - \lfloor\sqrt{T}\rfloor - 2) \\
&\leq \frac{3\mu_3 C_4 K_3(t-1)\log T}{\sqrt{T}} + b(t - \lfloor\sqrt{T}\rfloor - 2).
\end{aligned}$$

As for the variance,

$$\begin{aligned}
\mathrm{Var}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right] &= \mathbb{E}\left(\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k - \mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right]\right)^2 \\
&= \mathbb{E}\left[\mathrm{Var}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \middle| \hat{\beta}, \hat{\gamma}, \hat{X}\right]\right] + \mathbb{E}\left(\mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \middle| \hat{\beta}, \hat{\gamma}, \hat{X}\right] - \mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right]\right)^2
\end{aligned}$$

Since the demand D_k for $k \geq \lfloor\sqrt{T}\rfloor + 2$ is conditional independent given $\hat{\beta}, \hat{\gamma}$ and $\hat{X}$, the first term can be bounded by for $\nu_0(t - \lfloor\sqrt{T}\rfloor - 2)$ some constant ν_0. While we can show that the function,

$$E\left[\beta^\top x_t + (\gamma^\top x_t) \cdot \left(-\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{2b - \hat{\beta}^\top \hat{X}}{2\hat{\gamma}^\top \hat{X}}\right)\right]$$

is L-Lipschitz in $(\hat{\beta}, \hat{\gamma}, \hat{X})$ for some constant L. Here the expectation is taken with respect to x_t. The second term

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \middle| \hat{\beta}, \hat{\gamma}, \hat{X}\right] - \mathbb{E}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right]\right)^2 \\
\leq & L^2(t - \lfloor\sqrt{T}\rfloor - 2)\mathbb{E}\left(\|\hat{\beta} - \beta\|_2 + \|\hat{\gamma} - \ \gamma\|_2 + \|\hat{X} - \mathbb{E}\boldsymbol{X}\|_2\right)^2 \\
\leq & \frac{9(t - \lfloor\sqrt{T}\rfloor - 2)L^2\mu_3 C_4 \log T}{\sqrt{T}}
\end{aligned}$$

Therefore, we can always choose a constant $\nu \geq \nu_0$ such that

$$\operatorname{Var}\left[\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k\right] \leq \nu(t - \lfloor\sqrt{T}\rfloor - 2).$$

Then,

$$\begin{aligned}
&\mathbb{P}\left(bT - K_3(\lfloor\sqrt{T}\rfloor + 1) - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \leq \bar{M}_\beta\right) \\
\leq\ & \mathbb{P}\left(\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \geq bT - K_3(\lfloor\sqrt{T}\rfloor + 1) - \bar{M}_\beta\right) \\
\leq\ & \mathbb{P}\left(\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[D_k] \geq bT - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[D_k] - K_3(\lfloor\sqrt{T}\rfloor + 1) - \bar{M}_\beta\right) \\
\leq\ & \mathbb{P}\left(\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[D_k] \geq bT - \frac{3\mu_3 C_4 K_3 (t-1)\log T}{\sqrt{T}} - b(t - \lfloor\sqrt{T}\rfloor - 2) - K_3(\lfloor\sqrt{T}\rfloor + 1) - \bar{M}_\beta\right) \\
\leq\ & \mathbb{P}\left(\sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} \mathbb{E}[D_k] \geq b(T - t) - C_0\sqrt{T}\log T\right)
\end{aligned}$$

where $C_0 = 3\mu_3 C_4 K_3 + b + K_3 + \bar{M}_\beta$. Applying Chebyshev's Inequality,

$$\mathbb{P}\left(bT - K_3(\lfloor\sqrt{T}\rfloor + 1) - \sum_{k=\lfloor\sqrt{T}\rfloor+2}^{t-1} D_k \leq \bar{M}_\beta\right) \leq \frac{\nu(t - \lfloor\sqrt{T}\rfloor - 2)}{(b(T-t) - C_0\sqrt{T}\log T)^2}.$$

Then,

$$\mathbb{E}[T-\tau] = \sum_{t=\lfloor\sqrt{T}\rfloor+2}^{T} \mathbb{P}(\tau \le t)$$
$$\le \sum_{t=\lfloor\sqrt{T}\rfloor+2}^{T} \max\left\{\frac{\nu(t-\lfloor\sqrt{T}\rfloor-2)}{(b(T-t)-C_0\sqrt{T}\log T)^2}, 1\right\} \le C_4'\sqrt{T}\log T.$$

With a similar method, we can show the other direction,

$$\mathbb{E}[\underline{B}] \le C_4'\sqrt{T}\log T$$

□

C.5 Proof of Theorem 6

Proof. Based on Theorem 4 and the result in Proposition 6, it is left to show that there exists a constant C_p such that

$$\mathbb{E}\left[\sum_{t=1}^{\tau_{\bar{M}_\beta}-1} (p_t - p_t^*)^2\right] \le C_p\sqrt{T}\log T,$$

for all $(\beta,\gamma)\in\Theta$ and $\mathcal{P}\in\Xi$. Also, the exploration phase lasts for $\sqrt{T}$ steps and that incurs at most $O(\sqrt{T})$ regret. We only need to show,

$$\mathbb{E}\left[\sum_{t=1}^{\tau_{\bar{M}_\beta}-1} (p_t - p_t^*)^2\right] \le C_p\sqrt{T}\log T. \tag{18}$$

Given that

$$p_t = -\frac{\hat{\beta}^\top x_t}{2\hat{\gamma}^\top x_t} + \frac{2b-\hat{\beta}\hat{X}}{2\hat{\gamma}\hat{X}}$$

$$p_t^* = -\frac{\beta^\top \mathbb{E}x_t}{2\gamma x_t} + \frac{2b-\beta^\top\mathbb{E}\boldsymbol{X}}{2\gamma^\top\mathbb{E}\boldsymbol{X}}.$$

Consider function $g:\Theta$

$$g_{x'}(\beta,\gamma,x) = -\frac{\beta x'}{\gamma x'} + \frac{2b-\beta x}{2\gamma x}.$$

Given the boundedness of β, γ, x, and x', there exists a constant L such that

$$|g_{x'}(\beta_1,\gamma_1,x_1) - g_{x'}(\beta_2,\gamma_2,x_2)| \le L\left(\|\beta_1-\beta_2\|_2 + \|\gamma_1-\gamma_2\|_2 + \|x_1-x_2\|_2\right).$$

So,

$$\begin{aligned}\mathbb{E}(p_t - p_t^*)^2 &\leq L^2 \mathbb{E}\left(\|\beta_1 - \beta_2\|_2 + \|\gamma_1 - \gamma_2\|_2 + \|x_1 - x_2\|_2\right)^2 \\ &\leq \frac{3C_4 L^2 \log T}{\sqrt{T}}.\end{aligned}$$

for all $t = \lfloor \sqrt{T} \rfloor + 2, ..., T$ Therefore, (18) follows and the theorem is proved. ☐

C.6 Proof of Proposition 7

Proof. We refer to the proof of Theorem 1 in the paper (Ban and Keskin, 2021). Differently, there are extra d parameters to estimate in our setting, which is $\mathbb{E}\boldsymbol{X}$. So, the Fisher information matrix $\tilde{J}_t$ is a $3d$-by-$3d$ matrix. By the upper-left $2d$-by-$2d$ matrix is the same as the J_t in the paper (Ban and Keskin, 2021). We can simply choose

$$\tilde{C}_t(\theta, x_t) = [C_t(\theta, x_t), \underbrace{0, ..., 0}_{d}]$$

and then all the proof therein still go through for our setting. ☐