DYNAMIC INVENTORY AND PRICE CONTROLS INVOLVING UNKNOWN DEMAND ON DISCRETE NONPERISHABLE ITEMS

By

TINGTING ZHOU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

Graduate Program in Management

Written under the direction of Michael Katehakis and Jian Yang

And approved by

_____

_____

_____

_____

_____

Newark, New Jersey

May, 2018

Abstract

Rutgers, The State University of New Jersey

School of Graduate Studies

Newark, New Jersey

Graduate Program in Management

Announcement of Ph.D. Dissertation Defense

TINGTING ZHOU

Dynamic Inventory and Price Controls Involving Unknown Demand on Discrete Nonperishable Items

Committee: Michael Katehakis, Jian Yang, Andrzej Ruszczynski, Hui Xiong, Junmin Shi

Date: April 3, 2018

Time: 3pm

Location: Room 1027 in 1WP, Newark Campus

Abstract: We study adaptive policies that handle dynamic inventory and price controls when the random demand for discrete nonperishable items is unknown. Pure inventory control is achieved by targeting newsvendor ordering quantities that correspond to empirical demand distributions learned over time. On this basis we conduct the more complex joint inventory-price control, where demand-affecting prices await to be evaluated as well. We identify policies that strive to balance between exploration and exploitation, and measure their performances via regrets, i.e., the prices to pay for not knowing demand distributions *a priori* over a given horizon. Multiple bounds are derived on regrets' growth rates; they vary with how thoroughly unknown the demand distributions are and whether nonperishability has indeed been accounted for. A simulation study shows that our policies compare favorably with other potential candidates.

Contents

1 Introduction

For a given firm, inventory control is about dynamically adjusting ordering quantities to minimize the total long-run expected cost stemming from mismatches between inventory levels and random demand realizations. When unit prices influence the random demand pattern that it faces, the firm can further exert joint inventory-price control to attain the maximum total long-run expected profit. Traditional models took the probabilistic distribution associated with the random demand pattern as a known factor. In many real situations, however, even the knowledge on demand distributions can be elusive. When the firm has just introduced a new product or when its external environment has just transitioned to a previously unfamiliar phase, e.g., a severe economic downturn, it will not be sure of the random demand pattern to come. One way out is adopting the Bayesian approach. In it, the firm possesses prior distributions on potential demand patterns. Then, posterior understandings on demand are formed on the basis of new demand realizations. Inventory management taking this approach can be found, for instance, in Scarf [34] and Lariviere and Porteus [30].

Most other times, even prior distributions on demand patterns can be too much to ask for. What meager information one possesses might just be a collection of potential demand distributions. Now, the concerned firm has yet to make decisions using its past observations. But its goal is no longer about catering to specific demand distributions or even sequences of posterior demand distributions. Rather, its history-dependent (henceforward called adaptive) control policy should better yield results that are reasonably good under all potential demand distributions from the given collection. A policy's regret under a given demand pattern and over a fixed time horizon measures the price paid for ambiguity; namely, the difference between the policy's performance and that of the best policy tailor-made for the demand pattern were it known. A policy will be considered good when its worst

regret over all demand patterns in a collection grows over time as slowly as possible.

We let items be discrete. Thus, the underlying firm can, for instance, be a car manufacturer or a wholesaler of bulky items such as bags of grain. Often, the granularity of items would not cause much difference to an inventory and price control setting; besides, continuous- and discrete-item systems are often good approximations of each other. By prohibiting the finest tuning of ordering decisions, our discrete-item setup has the advantage of better reflecting many real firms' inabilities to manage their inventories to exact precisions.

We start with pure inventory control, where the demand distribution f is only known to have an expectation not exceeding some level $\bar{m} > 0$. We adopt a very simple and natural policy that has also been considered by Besbes and Muharremoglu [6]. Recall that the optimal ordering quantity for a newsvendor problem involving an effective holding cost rate $\bar{h}$, an effective backlogging cost rate $\bar{b}$, and a known demand distribution f is the β-quantile of f, where $\beta = \bar{b}/(\bar{h}+\bar{b})$. In every period t, the heuristic policy without knowing the true f advocates ordering up to the β-quantile of the empirical demand distribution $\hat{f}_{t-1}$ that is learned from past demand levels in periods $1, 2, ..., t-1$. A minor modification comes in the form of an artificial upper bound $\bar{d}$ on the order-up-to level. We show that the policy's worst regret over all distributions will not grow faster than the rate $T^{1/2} \cdot (\ln T)^{3/2}$.

A bound linking to the nonperishability of items is one of our most technical accomplishments. It will be repeatedly used in joint inventory-price control.

A good portion of the dissertation is then devoted to the more complex task of joint inventory-price control. Here, the firm can choose from prices $\bar{p}^1, \bar{p}^2, ..., \bar{p}^{\bar{k}}$. But the demand distribution f^k under each choice $k = 1, 2, ..., \bar{k}$ is largely unknown. On top of the common upper bound $\bar{m}$ on the mean demand level, we additionally impose a common upper bound $\bar{s}$ on demands' standard deviations. Now the empirical distribution $\hat{f}^k_{t-1}$ of demand under the price choice k by the beginning of period t

depends on how many times k has been chosen in periods $1, 2, ..., t-1$, and is in turn the product of both demand realizations and the policy adopted. We propose *learning while doing* policies LwD(μ) that are parameterized by some constant μ. A policy from this group ensures that every price is visited often enough; indeed, the number of visits is in the order of t^{μ} by period t. At the same time, it gives prices with more promising historical performances more chances to be visited. Let V_f^k be the best average single-period profit the firm can achieve under the price choice k and demand distribution f, if the latter were known. As the actual f^k under choice k is unknowable to the firm, the policy advocates that, as much as possible, the k achieving the maximum $\tilde{V}_{t-1}^k$, an approximation of the profit $V_{\hat{f}_{t-1}^k}^k$ produced by the empirical distribution, be chosen in period t. The approximation is needed because a cutoff has to be made on the potentially infinite revenue-side computation involving an unbounded support.

As in pure inventory control, the analysis dealing with perishable items draws upon established results in information theory and large deviation, such as Hoeffding's inequality. Basically, we take advantage of the fact that empirical distributions will get ever closer to their generating distributions as increasingly more realizations are observed. The issue of nonperishability necessitates more innovations on our part. We take up the nonperishability-induced bound in pure control and for this to work well, introduce virtual learning periods that accumulate at rates roughly proportional to t^{μ}. This trick allows us to establish that the dominant price, if there ever is one, will be used in long sequences of periods; consequently, the sub-linear bound from pure control will be patched up over various sequences to deliver a reasonable bound for joint control.

The most challenging case is when demand is so utterly unknown that even the existence of a price leading others by a tiny margin cannot be taken for granted. Note distinctions between leaders and followers, whether they be in terms of

distributions or average rewards, are often the enablers of adaptive control's regret analysis; see, e.g., Lai and Robins [29] and Auer, Cesa-Bianchi, and Fischer [2]. To handle the more applicable case without such distinctions, we build on the virtual-learning idea to obtain stickier modifications of the LwD(μ) policies. We name them LwD$'(\mu, \nu, \psi)$, which are parameterized by constants μ, ν, and ψ. These policies are aware of built-in virtual learning periods that accumulate at rates proportional to $t^{1-\psi}$ and in most periods, favor incumbent price choices at degrees expressible in terms of the parameters and the times t. The extra lengths that single prices linger on permit the pure-control bound to take its effect.

In addition to performance guarantees for particular policies, we also attempt to identify lower bounds on regrets that no policy can ever beat. Now, let us detail our main results.

Pure inventory control

$*$ regrets caused by the newsvendor-based policy are of the form $O(T^{1/2} \cdot (\ln T)^{3/2})$ (Theorem 1), while Proposition 2 on nonperishability will remain useful for joint control;

Joint inventory-price control

when items are *perishable*:

when $V^k_{f^k}$ is uniquely maximized and the second best choice is at least $\delta > 0$ behind:

$*$ the LwD(μ) policy with the best known performance guarantee happens at $\mu = 1/2$, whence regrets are of the form $O(T^{1/2} \cdot (\ln T)^{1/2})$ (Proposition 6);

when the demand-distribution vector $\mathbf{f} \equiv (f^k)_{k=1,2,\ldots,\bar{k}}$ can roam more freely:

$*$ the LwD(μ) policy with the best known performance guarantee happens at $\mu = 4/5$, whence regrets are of the form $O(T^{4/5})$ (Proposition 7);

when items are *nonperishable* and $V^k_{f^k}$ is uniquely maximized with a δ margin:

$*$ the LwD(μ) policy with the best known performance guarantee happens at

$\mu = 4/7$, whence regrets are of the form $O(T^{11/14} \cdot (\ln T)^{5/2})$ (Theorem 2);

when items are *nonperishable* and demand patterns are more arbitrary:

$*$ it takes the LwD$'(\mu, \nu, \psi)$ policies to achieve known performance guarantees; the best one, also our ultimate result, is attained at $\mu = 2/3$, an arbitrary ν, and $\psi = 1/3$, whence regrets are of the form $O(T^{5/6} \cdot (\ln T)^{5/2})$ (Theorem 3);

$*$ the tightest lower bound so far achievable is of the form $\Omega(T^{1/2})$ (Theorem 4).

When demand distributions have finite supports, both Proposition 7 and Theorem 2 can be improved. However, Theorem 3 will remain intact.

We have conducted a simulation study. Its pure inventory control part demonstrates the competitiveness of the newsvendor-based policy. The main part concerning joint inventory-price control suggests that more work is probably needed on both the upper-bounding Theorem 3 and lower-bounding Theorem 4. It is likely that $T^{3/5}$- or $T^{2/3}$-sized bounds should prevail for joint inventory-price control just as $T^{1/2}$-sized ones do for pure inventory control. It also indicates that nonperishability does not contribute as much to regret bounds as suggested by our theoretical bounds, on which we have spent major efforts. Therefore, new ideas, especially those that do not rely on the pure-control bound, should be welcomed.

The remainder of the dissertation is organized as follows. We put our contribution in the perspective of existing literature in Section 2. Then, Section 3 introduces pure inventory control along with the newsvendor-based policy; whereas, Section 4 provides various upper bounds. For joint inventory-price control, we use Section 5 to introduce the problem and the LwD(μ) policies. We then spend the next three sections on detailed analyses. Section 6 focuses on the case with perishable items; then, Section 7 moves on to nonperishable items, however, with a slight restriction on demand patterns. In Section 8, we achieve upper bounds for the modified policies LwD$'(\mu, \nu, \psi)$ under the most general conditions involving nonperishable items and unrestricted demand patterns. A lower bound is also derived there. We present our

simulation study in Section 9 and conclude the dissertation in Section 10.

2 Literature Survey

Pioneering works on regret analysis started with adaptive allocation, where the main concern is on dynamically selecting the most promising pool of samples to draw so as to maximize their total sum; see, e.g., Robbins [32], Lai and Robbins [29], Katehakis and Robbins [27], and Auer, Cesa-Bianchi, and Fischer [2]. Also, Auer et al. [3] treated variants where each choice's output is not necessarily an independent sample from a predetermined though unknown distribution; meanwhile, Burnetas, Kanavetas, and Katehakis [11] introduced constraints on the total costs of sampling from various pools. The dynamic pricing portion of our work is akin to picking a winning pool of samples. However, subsequent ordering decisions and inventory carry-overs pose additional challenges.

Regret analysis was also conducted on adaptive Markov decision processes (MDPs) that often involve unknown reward patterns and unknown random state transitions; see, e.g., Burnetas and Katehakis [12], Auer and Ortner [4], Tewari and Bartlett [35], and Jacksh, Ortner, and Auer [26]. Regret bounds derived in this body of literature are often dependent on particular MDPs or to lesser degrees, characteristics of MDPs such as their so-called diameters. So even inventory and price controls are MDP by nature, we need new insights and techniques to achieve regret bounds that countenance all or nearly all possible demand distributions that underlie our particular Markov processes.

Adaptive policies for inventory control have been considered. Huh and Rusmevichientong [24] focused on a gradient-based policy. It could also be thought of as an extension of stochastic approximation (SA), which was started by Robbins and Monro [33] and Kiefer and Wolfowitz [28]. Huh et al. [25] used the so-called Kaplan-Meier estimator on the distribution function of demand when the latter is censored. Moreover, Besbes and Muharremoglu [6] studied the implications of demand censoring in pure inventory control involving unknown demand. They

focused on the discrete-item case of the repeated newsvendor problem and proposed policies with provably good performance guarantees. In a revenue management setup, Besbes and Zeevi [7] studied the dynamic selection of prices while learning demand on the fly. For the newsvendor problem and its multi-period version involving nonperishable items, Levi, Roundy, and Shmoys [31] relied on randomly generated demand samples to reach solutions with relatively good qualities at high probabilities.

We study both pure inventory and joint inventory-price controls involving the real-time learning of unknown demand patterns regarding discrete nonperishable items, with more emphasis placed on the latter joint control. In pure inventory control, Huh and Rusmevichientong [24] also dealt with items' nonperishability in their main continuous-item part. When items are discrete, their SA-based approach was analyzed for perishable items only. We rely on a newsvendor-based policy involving empirical demand distributions, which was considered by Besbes and Muharremoglu [6] in their study of perishable items. Our contribution is rather on the technical side, with an emphasis on bounding the policy's nonperishability-induced regrets. Proposition 2, especially, enables the analysis of nonperishability-induced regrets in joint inventory-price control.

Pure price control involving unknown demand patterns, as was treated in Besbes and Zeevi [7], is a problem transient in nature. In it, prices are adjusted over time for the firm to reap the highest profit from selling a given initial stock within a fixed time horizon. This area is seeing rapid progress in recent years. For instance, Wang, Deng, and Ye [37] proposed a policy that conducts learning and doing intermittently and achieves very tight regret bounds. Besbes and Zeevi [8] demonstrated that a firm could take its demand function as linear and still manage to avoid severe regrets. Ferreira, Simchi-Levi and Wang [22] applied Thompson Sampling to a network revenue management setting, where different products consume a given set

of resources. Also, Aviv and Pazgal [5], Araman and Caldentey [1], Farias and van Roy [20], Broder and Rusmevichientong [10], and den Boer and Zwart [9] took parametric approaches to such problems. Meanwhile, Cheung, Simchi-Levi, and Wang [16] limited the number of price changes in a setting without inventory constraints.

In contrast to the transient pure price control, joint inventory-price control is recurrent in nature. It deals with the repeated use of pricing and ordering for the attainment of the highest profit in the long run. Our counterpart with known demand patterns and strict discounts over time is Federgruen and Heching [21]. Assuming unknown demand patterns, Burnetas and Smith [13] studied such a problem where demand distributions are continuous and the only information available is whether sales have exceeded ordering quantities. They developed an SA-based method that reached consistency, i.e., asymptotic convergence in time-average profit to the truly optimal; however, rates of convergence were left untouched.

Like us, Chen, Chao, and Ahn [14] also dealt with joint inventory-price control involving unknown demand patterns while providing bounds on regrets' growth rates. Whereas we allow a finite number of price choices and assume an arbitrary price-demand relationship, they let prices come from a compact interval, adopted an either additive or multiplicative demand pattern with average demand decreasing over price, and also made other assumptions like the twice differentiability of the demand-price relationship's deterministic part, concavity of the average revenue function, and strict positivity of average demand levels. They were able to achieve $T^{1/2}$-sized regret bounds. Between the slightly earlier work and ours, we believe there is some trade-off between model flexibility and solution quality. Our avoidance of any assumption on the price-demand relationship reduces the risk of model mis-specification to the minimum. On the flip side, this probably contributes to our

less ideal, $T^{11/14}$- and $T^{5/6}$-sized bounds, as illustrated in Theorems 2 and 3, respectively. Working with a setup similar to that of Chen, Chao, and Ahn [14], recently Chen, Chao, and Shi [15] established a $T^{4/5}$-sized regret bound for the case involving lost sales and censored demand observation.

Our results, especially those allowing demand patterns total freedom in their ranges, such as Theorems 1 and 3, will offer guidance to production, inventory, and sales managers at those critical junctures when for instance, new products have just been rolled out or the economy has just entered a new phase. Besides effective uses of the empirical distribution's properties known in the theories of information and large deviations, we contribute methodologically in the LwD(μ) policies that balance between exploration and exploitation, their stickier variants the LwD$'(\mu, \nu, \psi)$ policies that favor incumbent prices to measured degrees, the virtual-learning trick that regulate frequencies of learning in both directions, and various other regret-analysis techniques. In both pure and joint controls, the nonperishability of discrete items poses as one of the main challenges. Our treatments of Propositions 12 and 16 to this effect might offer ideas to be lent to other applications.

Due to presently unsurmountable difficulties, there is still a sizable gap between the $T^{5/6}$-sized upper bound in Theorem 3 and the $T^{1/2}$-sized lower bound in Theorem 4. Our simulation study adds to the credibility of a $T^{3/5}$- to $T^{2/3}$-sized bound. In addition, it questions whether nonperishable items contribute as much to the final bound as has been depicted in Propositions 12 and 16. We probably need a new proof idea other than the current one focusing on long-lasting single dominant prices. With major efforts so far devoted to the enhancement of the tolerable ambiguity level for demand patterns and also the mitigation of difficulties caused by items' nonperishability, we have left little to show for demand censoring, admittedly another major issue in real applications. Nevertheless, we might have laid some of

the ground work that more realistic and inclusive models could build on. One avenue that the latter could take is substituting the the empirical demand distribution with the Kaplan-Meier estimator as was studied by Huh et al. [25].

3 Pure Inventory Control

Let $\mathbb{N}$ be the set of natural numbers and $\Re$ the set of reals. Also, use $\mathcal{F}_0$ for the collection of all random distributions with $\mathbb{N}$ as their support. A distribution $f \equiv (f(d))_{d\in\mathbb{N}}$ in $\mathcal{F}_0$ satisfies both $f(d) \geq 0$ for every $d \in \mathbb{N}$ and $\sum_{d=0}^{+\infty} f(d) = 1$. Use F_f for the cumulative distribution function (cdf) of any given $f \in \mathcal{F}_0$. It satisfies $F_f(x) = \sum_{d=0}^{\lfloor x \rfloor} f(d)$ for $x \in \Re$. Now consider a multi-period inventory control problem in which the demand D_t in every period $t \in \mathbb{N}$ is a random draw from a collection $\mathcal{F}_1(\bar{m}) \subset \mathcal{F}_0$ of distributions with a uniformly bounded mean $\bar{m} > 0$. A distribution $f \equiv (f(d))_{d\in\mathbb{N}}$ in $\mathcal{F}_1(\bar{m})$ for a random demand D further satisfies

$$\mathbb{E}_f[D] \equiv \sum_{d=0}^{+\infty} d \cdot f(d) = \sum_{d=0}^{+\infty} (1 - F_f(d)) \leq \bar{m}. \tag{1}$$

Suppose the firm starts with nothing at the beginning of period 1 over a T-period horizon.

We primarily consider the case where items are nonperishable and unsatisfied requests are backlogged. Suppose unit holding and backlogging cost rates are some strictly positive $\bar{h}$ and $\bar{b}$, respectively. Also, in any period $t = 1, 2, ..., T$, denote the order-up-to level by y_t and the realized demand level by d_t. Then, with

$$q(y, d) \equiv \bar{h} \cdot (y - d)^+ + \bar{b} \cdot (d - y)^+, \tag{2}$$

the relevant total cost over the T-period horizon will come at $\sum_{t=1}^{T} q(y_t, d_t)$. Indeed, according to the discussion in Section 1 of Appendix 12 which serves as this dissertation's supplement, we can use the above to handle all four combinations where unsatisfied demands are either backlogged or lost and where leftover items are either perishable or nonperishable. When items are perishable, $\bar{h}$ can be understood as the difference $\bar{c} - \bar{s}$, where $\bar{c}$ is the unit production cost and $\bar{s}$ the

unit salvage value. When unsatisfied requests are lost, $\bar{b}$ can be treated as the difference $\bar{l} - \bar{c}$, where $\bar{l}$ is the cost for not satisfying a unit demand. When items are nonperishable, we further require

$$y_t \geq y_{t-1} - d_{t-1}, \qquad \forall t = 2, 3, ..., T. \tag{3}$$

Later it will be clear that this could severely complicate our analysis.

For every $f \in \mathcal{F}_1(\bar{m})$ and $y \in \mathbb{N}$, let $Q_f(y)$ be the single-period average cost under the demand pattern f and order-up-to level y:

$$\begin{aligned} Q_f(y) &\equiv \mathbb{E}_f[q(y, D)] = \textstyle\sum_{d=0}^{+\infty} f(d) \cdot [\bar{h} \cdot (y - d)^+ + \bar{b} \cdot (d - y)^+] \\ &= \bar{h} \cdot \textstyle\sum_{d=0}^{y-1} F_f(d) + \bar{b} \cdot \sum_{d=y}^{+\infty} (1 - F_f(d)). \end{aligned} \tag{4}$$

Also, let $Q_f^* = \min_{y \in \mathbb{N}} Q_f(y)$ be the minimum cost in one period under f. Suppose y_f^* is an order-up-to level that achieves the one-period minimum. Then, when facing a T-period horizon, an optimal policy with a known f will be to repeatedly order up to this level. Thus, the minimum cost over T periods is $Q_f^* \cdot T$.

A salient feature of our current problem, though, is that f is not known beforehand. So instead of any f-dependent policy, we seek a good f-independent policy which takes advantage of demand levels observed in the past. A deterministic policy $\mathbf{y} \equiv (y_t)_{t=1,2,...,T}$ is such that, for $t = 1, 2, ..., T$, each $y_t \in \mathbb{N}$ is a function of the historical demand vector $\mathbf{d}_{[1,t-1]} \equiv (d_s)_{s=1,2,...,t-1}$. Under it, the T-period total average cost is

$$Q_f^T(\mathbf{y}) \equiv \sum_{t=1}^{T} \mathbb{E}_f[\bar{h} \cdot (y_t(\mathbf{D}_{[1,t-1]}) - D_t)^+ + \bar{b} \cdot (D_t - y_t(\mathbf{D}_{[1,t-1]}))^+], \tag{5}$$

which, due to the independence between $\mathbf{D}_{[1,t-1]}$ and D_t, is equal to $\sum_{t=1}^{T} \mathbb{E}_f[Q_f(y_t(\mathbf{D}_{[1,t-1]}))]$. Define the T-period regret $R_f^T(\mathbf{y})$ of using the policy $\mathbf{y}$

against the unknown distribution f:

$$R_f^T(\mathbf{y}) \equiv Q_f^T(\mathbf{y}) - Q_f^* \cdot T = \sum_{t=1}^{T} \mathbb{E}_f[Q_f(y_t(\mathbf{D}_{[1,t-1]}))] - Q_f^* \cdot T. \tag{6}$$

Our goal is to prevent $R_f^T(\mathbf{y})$ from growing too fast in T under all f's within $\mathcal{F}_1(\bar{m})$. We concentrate on one policy inspired by an optimal y_f^* when f is known. From (4), we see that necessary and also sufficient conditions for optimality of any y are

$$Q_f(y+1) - Q_f(y) = (\bar{h} + \bar{b}) \cdot F_f(y) - \bar{b} \geq 0, \tag{7}$$

and

$$Q_f(y) - Q_f(y-1) = (\bar{h} + \bar{b}) \cdot F_f(y-1) - \bar{b} \leq 0, \tag{8}$$

Let $\beta = \bar{b}/(\bar{h} + \bar{b})$ be the famous newsvendor parameter that lies in $(0, 1)$. For $f \in \mathcal{F}_0$, let y_f^* be the associated newsvendor order-up-to level, such that

$$y_f^* \equiv {F_f}^{-1}(\beta) \equiv \min\{d \in \mathbb{N} : \ F_f(d) \geq \beta\}. \tag{9}$$

By definition, $F_f(y_f^*) \geq \beta$ and hence $Q_f(y_f^* + 1) - Q_f(y_f^*) \geq 0$ by (7); also, $F_f(y_f^* - 1) < \beta$ and hence $Q_f(y_f^*) - Q_f(y_f^* - 1) < 0$ by (8). Therefore, $Q_f(y_f^*) = Q_f^*$, meaning that y_f^* is an optimal order-up-to level for the one-period problem when f is known.

Now with f unknown, we might adopt the newsvendor level $y_{f_{t-1}}^*$ for some good estimate f_{t-1} of f. The primary candidate for f_{t-1} is the empirical distribution $\hat{f}_{t-1}$. For $t = 2, 3, ...$, denote the empirical distribution $\hat{f}_{t-1} \in \mathcal{F}_0$ by $(\hat{f}_{t-1}(d))_{d \in \mathbb{N}}$, where

$$\hat{f}_{t-1}(d) = \frac{\sum_{s=1}^{t-1} \mathbf{1}(d_s = d)}{t-1}, \qquad \forall d \in \mathbb{N}. \tag{10}$$

Each $\hat{f}_{t-1}$ has its corresponding cdf $\hat{F}_{t-1} \equiv F_{\hat{f}_{t-1}}$. When f is truly arbitrary from $\mathcal{F}_0$, the closeness between f and $\hat{f}_{t-1}$ would be hard to guage. This is why we confine the f's to $\mathcal{F}_1(\bar{m})$, where (1) is satisfied. Still, there is no guarantee for $\hat{f}_{t-1} \in \mathcal{F}_0$ to be in $\mathcal{F}_1(\bar{m})$; yet, we need the resulting order-up-to level to be close to y_f^* when $f \in \mathcal{F}_1(\bar{m})$. For this purpose, we introduce an artificial bound. From (1), we have for any $f \in \mathcal{F}_1(\bar{m})$ and $y \in \mathbb{N}\backslash\{0\}$ that

$$y \cdot (1 - F_f(y-1)) \leq \sum_{d=0}^{y-1}(1 - F_f(d)) \leq \sum_{d=0}^{+\infty}(1 - F_f(d)) \leq \bar{m}. \tag{11}$$

Meanwhile, (9) dictates that $y_f^* = 0$ or $F_f(y_f^* - 1) < \beta$. So in combination with (11),

$$y_f^* \leq \frac{\bar{m}}{1 - F_f(y_f^* - 1)} < \frac{\bar{m}}{1-\beta}. \tag{12}$$

Our heuristic lets the firm order nothing in period 1; that is, $y_1 = \hat{y}_1 = 0$. For any $t = 2, 3, ...$ and $\bar{d} \equiv \lceil 2\bar{m}/(1-\beta) \rceil$ which is safely above the bound $\bar{m}/(1-\beta)$ in (12), it lets

$$\hat{y}_t = y^*_{\hat{f}_{t-1}} \wedge \bar{d} = \hat{F}^{-1}_{t-1}(\beta) \wedge \bar{d} = \min\left\{ d = 0, 1, ..., \bar{d} - 1 : \sum_{s=1}^{t-1} \mathbf{1}(d_s \leq d) \geq \beta \cdot (t-1) \right\}, \tag{13}$$

where the last formula is understood as $\bar{d}$ when none of the $d = 0, 1, ..., \bar{d} - 1$ satisfies the inequality. Furthermore, the heuristic advises the firm to order up to $y_t = \hat{y}_t$ in case items are perishable; otherwise, in order that (3) is satisfied, it advises the firm to order up to

$$y_t = \hat{y}_t \vee (y_{t-1} - d_{t-1}). \tag{14}$$

Due to the hard bound $\bar{d}$ used in (13), information gleaned from any d_s exceeding the $\bar{d}$ level can be disregarded. Within any $\hat{f}_{t-1} \equiv (\hat{f}_{t-1}(d))_{d \in \mathbb{N}}$ defined at (10), the

portion that is useful to the ordering decision comes from its first $\bar{d}$ components:

$$\hat{f}_{t-1}(0), \hat{f}_{t-1}(1), ..., \hat{f}_{t-1}(\bar{d}-1).$$

4 Bounds for Pure Control

We show that the newsvendor-based policy $\mathbf{y}$ described through (13) and (14) will incur a regret $R_f^T(\mathbf{y})$ that is slow-growing in the horizon length T. By (6) and (14),

$$R_f^T(\mathbf{y}) = R_f^{T1}(\mathbf{y}) + R_f^{T2}(\mathbf{y}), \tag{15}$$

where

$$R_f^{T1}(\mathbf{y}) = \sum_{t=1}^{T} \mathbb{E}_f[Q_f(\hat{y}_t)] - Q_f^* \cdot T, \tag{16}$$

and since $y_1 = \hat{y}_1 = 0$ by design and hence $y_2 = \hat{y}_2$,

$$R_f^{T2}(\mathbf{y}) = \sum_{t=3}^{T} \mathbb{E}_f[Q_f(y_t) - Q_f(\hat{y}_t)]. \tag{17}$$

In view of (4) and (17), over-payment in holding might be more than offset by under-payment in backlogging. So $R_{\mathbf{f}}^{T2}(\mathbf{y})$ for an arbitrary policy $\mathbf{y}$ might even be strictly negative. Still, it can be said that $R_f^{T1}(\mathbf{y})$ represents the price paid for the regrettable fact that the policy $\mathbf{y}$ was not designed with the particular distribution f in mind; meanwhile, $R_f^{T2}(\mathbf{y})$ captures the additional "cost" due to the nonperishability of items.

Our derivation will rely on the convergence of the empirical distribution $\hat{f}_{t-1} \in \mathcal{F}_0$ to the true distribution $f \in \mathcal{F}_1(\bar{m})$. Let

$$\delta_V(f, g, d) \equiv \max_{d'=0}^{d-1} |F_f(d') - F_g(d')|, \tag{18}$$

i.e., the maximum difference between the cdf's of f and g up to the level $d-1$. We

have

$$\mathbb{P}_f\left[\delta_V(f,\hat{f}_{t-1},d)\geq\varepsilon\right]\leq 2d\cdot\exp\left(-2\varepsilon^2\cdot(t-1)\right). \tag{19}$$

Indeed, let $X_1,...,X_n$ be independent random variables where each X_i is bounded by the interval $[a_i,b_i]$. Then for any $\varepsilon\geq 0$, Theorem 2 of Hoeffding [23] stated that

$$\mathbb{P}\left[|\frac{\sum_{i=1}^n X_i}{n}-\mathbb{E}[\frac{\sum_{i=1}^n X_i}{n}]|\geq\varepsilon\right]\leq 2\cdot\exp\left(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n(b_i-a_i)^2}\right). \tag{20}$$

If we make the dependence on $\mathbf{D}_{[1,t-1]}$ of the entity $\hat{f}_{t-1}$ defined through (10) and hence that of $\hat{F}_{t-1}$ explicit, we have $[\hat{F}_{t-1}(\mathbf{D}_{[1,t-1]})](d')=\sum_{s=1}^{t-1}\mathbf{1}(D_s\leq d')/(t-1)$ for every $d'\in\mathbb{N}$. Note that every $\mathbf{1}(D_s\leq d')$ is in the interval $[0,1]$ and

$$\mathbb{E}_f\left[[\hat{F}_{t-1}(\mathbf{D}_{[1,t-1]})](d')\right]=\mathbb{E}_f\left[\frac{\sum_{s=1}^{t-1}\mathbf{1}(D_s\leq d')}{t-1}\right]=\frac{\sum_{s=1}^{t-1}\mathbb{P}_f[D_s\leq d']}{t-1}=F_f(d'). \tag{21}$$

Therefore, (20) will result in

$$\mathbb{P}_f[|\hat{F}_{t-1}(d')-F_f(d')|\geq\varepsilon]\leq 2\cdot\exp(-2\varepsilon^2\cdot(t-1)),\qquad\forall d'\in\mathbb{N}. \tag{22}$$

The above will then lead to

$$\mathbb{P}_f\left[\max_{d'=0}^{d-1}|\hat{F}_{t-1}(d')-F_f(d')|\geq\varepsilon\right]\leq\sum_{d'=0}^{d-1}\mathbb{P}_f[|\hat{F}_{t-1}(d')-\hat{F}_f(d')|\geq\varepsilon], \tag{23}$$

which will yield (19). We also need $y^*_{\hat{f}_{t-1}}$ to not exceed $\bar{d}$ too often. By (11),

$$F_f(\bar{d})\geq 1-\frac{\bar{m}}{\bar{d}+1}>\frac{1+\beta}{2}. \tag{24}$$

In view of (9), this will lead to

$$\mathbb{P}_f\left[y^*_{\hat{f}_{t-1}} \geq \bar{d}+1\right] = \mathbb{P}_f\left[\hat{F}_{t-1}(\bar{d}) < \beta\right] \leq \mathbb{P}_f\left[F_f(\bar{d}) - \hat{F}_{t-1}(\bar{d}) > \frac{1-\beta}{2}\right], \qquad (25)$$

which, by (22), amounts to

$$\mathbb{P}_f\left[y^*_{\hat{f}_{t-1}} \geq \bar{d}+1\right] \leq 2\cdot\exp\left(-(1-\beta)^2\cdot(t-1)/2\right). \qquad (26)$$

Without any prior knowledge on the $f \in \mathcal{F}_1(\bar{m})$, we can manage to obtain a $T^{1/2}\cdot(\ln T)^{1/2}$-sized bound on the $R_f^{T1}(\mathbf{y})$ defined in (16). Due to (13), the key is to show that $Q_f(y^*_{\hat{f}_{t-1}} \wedge \bar{d}) - Q_f(y^*_f)$ will converge to 0 quickly. This will be achievable if we can show that $Q_f(y^*_g \wedge \bar{d}) - Q_f(y^*_f)$ will be small when g and f are close by and that $y^*_{\hat{f}_{t-1}}$ will not exceed $\bar{d}$ too often. These are when (19), (26), and other properties related to the inventory management problem such as the optimality of y^*_f to $Q_f(\cdot)$, will be useful. The final form of the bound comes from the estimation of certain summations through integrations.

Proposition 1 *For any $f \in \mathcal{F}_1(\bar{m})$ and $g \in \mathcal{F}_0$, as well as $y^0 \equiv y^*_g \wedge \bar{d}$,*

$$Q_f(y^0) - Q_f(y^*_f) \leq (\bar{h}\bar{d} + \bar{b}\bar{d})\cdot[\delta_V(f,g,\bar{d}) + \mathbf{1}(y^*_g \geq \bar{d}+1)].$$

Consequently, there are positive constants A^{Prop1} and B^{Prop1}, such that

$$R_f^{T1}(\mathbf{y}) \leq A^{Prop1} + B^{Prop1}\cdot T^{1/2}\cdot(\ln T)^{1/2}.$$

All remaining proofs of this section have been relegated to Appendix 11. Next, we obtain a bound in the same order of magnitude for $R_f^{T2}(\mathbf{y})$ as defined by (17). The

process is much more involved than that for $R_f^{T1}(\mathbf{y})$. From (14), we see that

$$y_t = \hat{y}_t \vee (\hat{y}_{t-1} - d_{t-1}) \vee (\hat{y}_{t-2} - d_{t-2} - d_{t-1}) \vee \cdots \vee (\hat{y}_1 - d_1 - d_2 - \cdots - d_{t-1}). \quad (27)$$

There is a latest s so that

$$y_t = \hat{y}_s - d_s - d_{s+1} - \cdots - d_{t-1}, \quad (28)$$

which occurs exactly when either $s = t$ or $s \leq t-1$ and

$$\hat{y}_s - d_s - d_{s+1} - \cdots - d_{t-1} - 1 \geq \hat{y}_t \vee (\hat{y}_{t-1} - d_{t-1}) \vee \cdots \vee (\hat{y}_{s+1} - d_{s+1} - \cdots - d_{t-1}), \quad (29)$$

and regardless,

$$\hat{y}_s \geq (\hat{y}_{s-1} - d_{s-1}) \vee (\hat{y}_{s-2} - d_{s-2} - d_{s-1}) \vee \cdots \vee (\hat{y}_1 - d_1 - d_2 - \cdots - d_{s-1}). \quad (30)$$

Inspired by the above, we define random variables $I \geq 1$ and $S_1, S_2, ..., S_I, S_{I+1}$ in an iterative fashion as follows. First, let $S_1 = 1$. Now for some $i = 1, 2, ...$, suppose S_i has been settled. Then, let S_{i+1} be the first t after S_i such that

$$\hat{y}_t \geq \hat{y}_{S_i} - D_{S_i} - D_{S_i+1} - \cdots - D_{t-1}, \quad (31)$$

if such a $t \leq T$ can be identified. If not, mark the latest i as I and let $S_{I+1} = T + 1$. For any t, let $L(t)$ be the largest $S_i \leq t$. This $L(t)$ can serve as the earlier s satisfying (29) and (30) that corresponds to t. Note that $L(t)$ along with $D_{L(t)}, D_{L(t)+1}, ..., D_{t-1}$ are independent of D_t. So by (17), as well as (28) to (31),

$$R_f^{T2}(\mathbf{y}) = \sum_{t=3}^{T} \mathbb{E}_f \left[Q_f(\hat{y}_{L(t)} - D_{L(t)} - \cdots - D_{t-1}) - Q_f(\hat{y}_t) \right]. \quad (32)$$

Now we are in a position to derive the nonperishability-induced bound.

Proposition 2 *There exist positive constants A^{Prop2} and B^{Prop2} such that*

$$R_f^{T2}(\mathbf{y}) \leq A^{Prop2} + B^{Prop2} \cdot T^{1/2} \cdot (\ln T)^{3/2}.$$

This is one of our most demanding results. It is also a building block for the analysis of the joint-control case involving nonperishable items. For its proof, we exploit the observations made from (27) to (32), all the while understanding that the actual order-up-to level y_t will be $\hat{y}_s - D_s - \cdots - D_{t-1}$ for some $s \leq t$. We are tasked to show that the term $Q_f(\hat{y}_s - D_s - \cdots - D_{t-1}) - Q_f(\hat{y}_t)$ can be bounded. For $\gamma = 1 - f(0)$, we divide the proof into two cases, the one with $\gamma \geq (1-\beta)/2$ and the other with $\gamma < (1-\beta)/2$.

In the former large-γ case, demand will accumulate over time with a guaranteed speed and $\hat{y}_t \geq \hat{y}_s - D_s - \cdots - D_{t-1}$ will occur ever more surely as $t - s$ increases. This is an effect similar to that achieved by Chen, Chao, and Ahn [14]'s assumption on the strict positivity of average demand levels. Then, for the minority case where $t - s$ is small, by exploiting natures of the empirical distribution and the newsvendor formula, we can come up with bounds related to $|Q_f(\hat{y}_s - D_s - \cdots - D_{t-1}) - Q_f(\hat{y}_t)| \cdot \mathbf{1}(\hat{y}_t \leq \hat{y}_s - D_s - \cdots - D_{t-1} - 1)$. Especially important is the observation that $\hat{y}_t \leq \hat{y}_s - D_s - \cdots - D_{t-1} - 1$ only if

$$\beta \leq \hat{F}_{t-1}(\hat{y}_t) \leq \hat{F}_{t-1}(\hat{y}_s - D_s - \cdots - D_{t-1} - 1) < \beta + \frac{t-s}{s}. \tag{33}$$

We will end up with a trade-off already encountered in the proof of Proposition 1. This is the source of the $T^{1/2} \cdot (\ln T)^{3/2}$-sized growth rate. However, the constants will grow as γ shrinks, because it takes ever longer for demand to accumulate. Therefore, we seek a different approach for the latter small-γ case, when $\gamma < (1-\beta)/2 < 1 - \beta$. This is the time when $y_f^* = \hat{F}_f^{-1}(\beta) = 0$ because

$F_f(0) = f(0) = 1 - \gamma > \beta$. We utilize the fact that $\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1$ is the bare minimum for $\hat{y}_s - D_s - \cdots - D_{t-1} \geq \hat{y}_t + 1$. But the latter will be true only if both $\hat{y}_s \geq 1$ and for some $d = 1, 2, ..., \bar{d}$, both $\hat{y}_s \geq d$ and $D_s + \cdots + D_{t-1} \leq d - 1$. For all γ's in the interval $[0, (1 - \beta)/2)$, we achieve a uniform bound in the order of $\ln T$, which is dominated by the one obtained in the first case.

Besides Hoeffding's inequality, the proof also exploits Markov's inequality in bounding $\mathbb{P}_f[\hat{y}_s \geq d]$, which, through (9) to (13), is the chance for the portion of earlier demand levels at or exceeding d to be greater than $1 - \beta$. Our proof has not been helped by the fact that the $\hat{y}_t$'s as defined through (13) can be time-varying. Had it not been so, a simpler proof like that for Proposition 5 of Chen, Chao, and Shi [15] might have been achievable.

Combining Propositions 1 and 2, we get a bound for the T-period regret $R_f^T(\mathbf{y})$.

Theorem 1 *Let* $\mathbf{y}$ *be the newsvendor-based adaptive policy. Then, there are positive constants* A^{Them1} *and* B^{Them1} *such that*

$$\sup_{f \in \mathcal{F}_1(\bar{m})} R_f^T(\mathbf{y}) \leq A^{Them1} + B^{Them1} \cdot T^{1/2} \cdot (\ln T)^{3/2}.$$

The constants involved can depend on the problem's parameters $\bar{h}$, $\bar{b}$, and $\bar{m}$. However, they are uniform across all f's in $\mathcal{F}_1(\bar{m})$. For the repeated newsvendor problem, Besbes and Muharremoglu [6] have already shown a $T^{1/2}$-sized lower bound (Lemma 4, with ε replaced by $1/T^{1/2}$ in its (C-8)). According to (14), the current case merely adds the restriction $y_t(\mathbf{d}_{[1,t-1]}) \geq y_{t-1}(\mathbf{d}_{[1,t-2]}) - d_{t-1}$ to the adaptive policy considered. So the lower bound can be no better. In view of this, the above is almost the best one can hope for. Huh and Rusmevichientong's [24] SA-based policy was shown to have a $T^{1/2}$-sized bound when items are continuous. Its modification for the discrete-item case would achieve a similar bound in the repeated newsvendor setting. Since we believe a bound involving nonperishable items would not be far off,

our own Theorem 1 will not seem too surprising. This being said, our pure-control study has set the stage for the more involved joint-control study to come. Especially useful will be the nonperishability-related effect already captured by Proposition 2.

5 Joint Inventory-price Control

We now combine pricing with inventory control. For some finite integer $\bar{k} = 2, 3, ...,$ suppose prices can be chosen from among the different levels $\bar{p}^1, \bar{p}^2, ..., \bar{p}^{\bar{k}}$ that are above the unit production cost $\bar{c}$. Under any price $\bar{p}^k$, supose demand will be randomly drawn from some $f^k \equiv (f^k(d))_{d\in\mathbb{N}}$ in the collection $\mathcal{F}_2(\bar{m}, \bar{s})$ which, for some $\bar{s} \geq 0$, is a strict subset of $\mathcal{F}_1(\bar{m})$. On top of the bound (1) on the mean, any $f \in \mathcal{F}_2(\bar{m}, \bar{s})$ has the following additional bound:

$$\mathbb{E}_f[D^2] \equiv \sum_{d=0}^{+\infty} d^2 \cdot f(d) = \sum_{d=0}^{+\infty} (2d+1) \cdot (1 - F_f(d)) \leq \bar{m}^2 + \bar{s}^2. \tag{34}$$

Basically, the standard deviation of any $f \in \mathcal{F}_2(\bar{m}, \bar{s})$ is bounded by $\bar{s}$. The new restriction is prompted by the joint control's need on revenue-side considerations. Let the firm immediately earn revenue $p \cdot d$ when it charges price p in any period in which demand realization happens to be d. In the backlogging case, the total cost of dealing with the demanded units is certainly captured by $\bar{c} \cdot \sum_{t=1}^{T} d_t + \sum_{t=1}^{T} q(y_t, d_t)$, with $q(y, d)$ defined at (2) posing as the total period-wise holding-backlogging cost. This can also be true for a lost sales case slightly different from the setting studied in Chen, Chao, and Shi [15]. Here, we assume that any demand unit not satisfied within a period will cost the firm an extra $\bar{l} - p = \bar{b} + \bar{c} - p$ on top of the price not earned. Basically, we are assuming a fixed total lost sales cost of $\bar{l}$. In the other setting, however, the extra cost on top of the lost revenue is assumed to be independent of the price charged.

For either the backlogging or fixed-total lost sales case, the firm's T-period profit is

$$\sum_{t=1}^{T} p_t \cdot d_t - \bar{c} \cdot \sum_{t=1}^{T} d_t - \bar{h} \cdot \sum_{t=1}^{T} (y_t - d_t)^+ - \bar{b} \cdot \sum_{t=1}^{T} (d_t - y_t)^+. \tag{35}$$

That is, it equals $\sum_{t=1}^{T} v(p_t, y_t, d_t)$, where

$$v(p, y, d) \equiv (p - \bar{c}) \cdot d - q(y, d) = (p - \bar{c}) \cdot d - \bar{h} \cdot (y - d)^+ - \bar{b} \cdot (d - y)^+. \tag{36}$$

Here, $v(p, y, d)$ can be understood as the profit the firm can make in one single period when it charges price p, orders y items, and faces a realized demand level of d. As before, whether or not (35) reflects the profit for nonperishable items depends on (i) whether the constant $\bar{h}$ is treated merely as the holding cost rate or the difference between the unit production cost $\bar{c}$ and the unit salvage value $\bar{s}$ and (ii) whether or not (3) is enforced.

Now, let $V_f(p, y)$ be the average of the profit $v(p, y, d)$ defined at (36) that the firm can make when it faces any demand distribution f. Note that

$$V_f(p, y) \equiv \mathbb{E}_f[v(p, y, D)] = (p - \bar{c}) \cdot \mathbb{E}_f[D] - Q_f(y), \tag{37}$$

where the average cost $Q_f(y)$ is defined at (4). Furthermore, for any price choice k, let

$$V_f^k \equiv \max_{y \in \mathbb{N}} V_f(\bar{p}^k, y) = V_f(\bar{p}^k, y_f^*), \tag{38}$$

i.e., the most that the firm can earn while charging price $\bar{p}^k$ and facing demand distribution f, where the optimal ordering level y_f^* is given in (9). When each price $\bar{p}^k$ corresponds to some demand distribution f^k, we can use the vector $\mathbf{f} \equiv (f^k)_{k=1,2,\ldots,\bar{k}}$ to reflect the current price-demand relationship. When given such a vector $\mathbf{f} \in (\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$, let

$$V_{\mathbf{f}}^* \equiv \max_{k=1}^{\bar{k}} V_{f^k}^k, \tag{39}$$

i.e., the maximum average profit that the firm can earn in one period. Suppose $k^*_{\mathbf{f}}$ solves (39); namely, $V^*_{\mathbf{f}} = V^{k^*_{\mathbf{f}}}_{f^{k^*_{\mathbf{f}}}}$. Then, when facing a T-period horizon with a known demand-distribution vector $\mathbf{f}$, an optimal policy will be to repeatedly charge the price $\bar{p}^{k^*_{\mathbf{f}}}$ and order up to $y^*_{f^{k^*_{\mathbf{f}}}}$. Thus, the maximum profit over T periods will be

$$V^*_{\mathbf{f}} \cdot T = V^{k^*_{\mathbf{f}}}_{f^{k^*_{\mathbf{f}}}} \cdot T = V_{f^{k^*_{\mathbf{f}}}}\left(\bar{p}^{k^*_{\mathbf{f}}}, y^*_{f^{k^*_{\mathbf{f}}}}\right) \cdot T.$$

When $\mathbf{f} \in (\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$ is unknown, we again seek a good adaptive policy. Such a policy $(\mathbf{k}, \mathbf{y}) \equiv (k_t, y_t)_{t=1,2,...,T}$ satisfies that, for $t = 1, 2, ..., T$, each price choice $k_t = 1, 2, ..., \bar{k}$ is a function of the historical demand vector $\mathbf{d}_{[1,t-1]}$, and so is each order-up-to level y_t. Under it, the T-period total average profit will be

$$V^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) \equiv \sum_{t=1}^{T} \mathbb{E}_{\mathbf{f}}\left[V_{f^{k_t}}\left(\bar{p}^{k_t}, y_t\right)\right]. \tag{40}$$

Note the average profit $V_{f^{k_t}}(\bar{p}^{k_t}, y_t)$ as defined in (37) can be used for each period t because, given a price choice k_t, the demand in that period is independent of earlier demands which determine the pricing and ordering decisions k_t and y_t. Now define the T-period regret $R^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y})$ of using the adaptive policy $(\mathbf{k}, \mathbf{y})$ under a demand-distribution vector $\mathbf{f}$:

$$R^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) \equiv V^*_{\mathbf{f}} \cdot T - V^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}). \tag{41}$$

We aim to identify adaptive policies $(\mathbf{k}, \mathbf{y})$ that prevent $R^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y})$ from growing too fast in T for "most" or even all $\mathbf{f}$'s within $(\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$.

A good policy should test each price $\bar{p}^k$ often enough to learn the corresponding distribution f^k well; yet, it should not dwell on the price for too long if $V^k_{f^k}$ is strictly below $V^*_{\mathbf{f}}$. As the f^k's are unknown, substitutes that are acquirable from past experiences can be used in their stead. Since inventory-related errors already amount to the order of $t^{1/2}$, we can use a term roughly proportional to t^{μ} at some $\mu \in [1/2, 1)$ as the guaranteed number that any price $\bar{p}^k$ will have been visited by

time t. At the same time, we should limit the visits to $\bar{p}^k$ when $V^k_{f^k}$, as approximated from its surrogate, does not seem promising. These inspire our policy. It keeps track of the number of times price $\bar{p}^k$ is charged in periods 1 through t:

$$\mathcal{N}^k_t \equiv \sum_{s=1}^{t} \mathbf{1}(p_s = \bar{p}^k). \tag{42}$$

The policy also designates the mode of each period t as either *learning*, with $m_t = 0$ or *doing*, with $m_t = 1$. It is certainly true that

$$\mathcal{N}^k_t = \mathcal{N}^k_{t,0} + \mathcal{N}^k_{t,1}, \tag{43}$$

with

$$\mathcal{N}^k_{t,m} \equiv \sum_{s=1}^{t} \mathbf{1}(m_s = m \text{ and } p_s = \bar{p}^k), \qquad \forall m = 0, 1. \tag{44}$$

As in pure inventory control, the policy uses empirical demand distributions $\hat{f}^k_{t-1} \equiv (\hat{f}^k_{t-1}(d))_{d\in\mathbb{N}}$ as surrogates of the acutal distributions. The newsvendor formula (13) effectively introduces a fixed cutoff point $\bar{d}$ for the pure control—any data collected about demand levels beyond $\bar{d}$ are not used. Here for joint inventory-price control, we will also use a cutoff point. However, due to the current need to estimate the revenue side as well, the new cutoff point will be higher and will also grow with the number of observations. In particular, when $\mathcal{N}^k_{t-1} \geq 1$, let $\tilde{f}^k_{t-1} \in \mathcal{F}_0$ stand for the empirical distribution of the demand under the price $\bar{p}^k$ observed over the past $t-1$ periods; however, with a cutoff

$$\tilde{d}^k_{t-1} \equiv (\mathcal{N}^k_{t-1})^{1/4} \vee \bar{d}. \tag{45}$$

With $\tilde{d}^k_{t-1}$ given by (45), we let

$$\tilde{f}^k_{t-1}(d) = \hat{f}^k_{t-1}(d) = \frac{\sum_{s=1}^{t-1} \mathbf{1}(p_s = \bar{p}^k \text{ and } d_s = d)}{\mathcal{N}^k_{t-1}}, \qquad \forall d = 0, 1, ..., \tilde{d}^k_{t-1} - 1, \tag{46}$$

$\tilde{f}^k_{t-1}(\tilde{d}^k_{t-1}) = 1 - \sum_{d=0}^{\tilde{d}^k_{t-1}-1} \tilde{f}^k_{t-1}(d)$, and $\tilde{f}^k_{t-1}(d) = 0$ for $d = \tilde{d}^k_{t-1} + 1, \tilde{d}^k_{t-1} + 2, ...$ This $\tilde{f}^k_{t-1} \equiv (\tilde{f}^k_{t-1}(d))_{d\in\mathbb{N}}$ is the surrogate demand distribution under the price $\bar{p}^k$. For any price choice $k = 1, 2, ...\bar{k}$, let $\tilde{V}^k_{t-1} \equiv V_{\tilde{f}^k_{t-1}}\left(\bar{p}^k, y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}\right)$, which by (37) further equals

$$\mathbb{E}_{\tilde{f}^k_{t-1}}\left[v\left(\bar{p}^k, y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}, D\right)\right] \equiv (\bar{p}^k - \bar{c}) \cdot \mathbb{E}_{\tilde{f}^k_{t-1}}[D] - Q_{\tilde{f}^k_{t-1}}\left(y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}\right). \tag{47}$$

These will be the approximate price-k profits on which we base pricing decisions. Their approximation powers can be seen by comparing (37) with (47). In the latter, while the cutoff point $\tilde{d}^k_{t-1}$ for demand learning grows with $\mathcal{N}^k_{t-1}$ in the fashion of (45), the bound $\bar{d}$ for ordering is still fixed at $\lceil 2\bar{m}/(1-\beta)\rceil$.

Here comes a detailed description of our LwD(μ) (*learning* while *doing*) policy at a parameter $\mu \in [1/2, 1)$. Initially, the policy lets $\mathcal{N}^k_{0,0} = \mathcal{N}^k_{0,1} = \mathcal{N}^k_0 = 0$ for $k = 1, 2, ..., \bar{k}$. Then, in every period $t = 1, 2, ...$, suppose $\kappa_{t-1}(1)$ is a k that minimizes the number $\mathcal{N}^k_{t-1}$. Subsequently, if $\mathcal{N}^{\kappa_{t-1}(1)}_{t-1} < (t/\bar{k})^\mu$, the policy will recommend the following:

0.1. set the mode of period t as *learning*, with $m_t = 0$;

0.2. also, let the price choice $k_t = \kappa_{t-1}(1)$;

0.3. next, conduct bookkeeping in the fashion of

$$\mathcal{N}^{k_t}_{t,0} = \mathcal{N}^{k_t}_{t-1,0} + 1, \qquad \mathcal{N}^{k_t}_{t,1} = \mathcal{N}^{k_t}_{t-1,1}, \qquad \mathcal{N}^{k_t}_{t} = \mathcal{N}^{k_t}_{t-1} + 1. \tag{48}$$

Otherwise, with $\mathcal{N}^{\kappa_{t-1}(1)}_{t-1} \geq (t/\bar{k})^\mu$ which necessitates that $\mathcal{N}^k_{t-1} \geq 1$ at every $k = 1, 2, ..., \bar{k}$, the policy will recommend the following:

1.1. set the mode of period t as *doing*, with $m_t = 1$;

1.2. also, let the price choice k_t be a maximizer of the profit estimate $\tilde{V}_{t-1}^k$ as defined at (47) from among $k = 1, 2, ..., \bar{k}$;

1.3. next, conduct bookkeeping in the fashion of

$$\mathcal{N}_{t,0}^{k_t} = \mathcal{N}_{t-1,0}^{k_t}, \qquad \mathcal{N}_{t,1}^{k_t} = \mathcal{N}_{t-1,1}^{k_t} + 1, \qquad \mathcal{N}_t^{k_t} = \mathcal{N}_{t-1}^{k_t} + 1. \tag{49}$$

Finally, for those k's unequal to k_t, we keep

$$\mathcal{N}_{t,0}^{k} = \mathcal{N}_{t-1,0}^{k}, \qquad \mathcal{N}_{t,1}^{k} = \mathcal{N}_{t-1,1}^{k}, \qquad \mathcal{N}_t^{k} = \mathcal{N}_{t-1}^{k}. \tag{50}$$

After the price choice k_t has been settled, the policy lets the firm charge the price $\bar{p}^{k_t}$. When $t = 1$, the policy has ordering facilitated through $y_1 = \hat{y}_1 = 0$. For $t = 2, 3, ...$, it advises

$$\hat{y}_t = y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d} = \min\left\{ d = 0, 1, ..., \bar{d} - 1 : \sum_{s=1}^{t-1} \mathbf{1}(\bar{p}_s = \bar{p}^k \text{ and } d_s \le d) \ge \beta \cdot \mathcal{N}_{t-1}^k \right\}, \tag{51}$$

when $\mathcal{N}_{t-1}^k \ge 1$ and $\hat{y}_t = 0$ when $\mathcal{N}_{t-1}^k = 0$, which is again followed by (14). Just as for (13), the last formula in (51) is understood as $\bar{d}$ when none of the $d = 0, 1, ..., \bar{d} - 1$ satisfies the inequality. Because $\tilde{d}_{t-1}^k \ge \bar{d}$, we have from (13) that

$$y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d} = y^*_{\hat{f}^k_{t-1}} \wedge \bar{d}. \tag{52}$$

In the policy, it is easy to see that the updatings (48) to (50) will ensure the satisfaction of (42) to (44) by the $\mathcal{N}_{t,0}^k$'s, $\mathcal{N}_{t,1}^k$'s, $\mathcal{N}_t^k$'s, and m_t in every period $t = 1, 2, ...$ In that exploration is done at controlled paces and exploitation is intended for the maximization of profit per period, the current LwD(μ) bears some resemblance to the pricing policy proposed in Section 4 of Burnetas and Smith [13].

However, our pacing using the $(t/\bar{k})^{\mu}$-function is different; it leads to Propositions 3 and 4 which are essential for our regret analysis. In addition, while we compare the potential profits under observed empirical distributions $\tilde{V}^k_{t-1}$ in the exploitation step 1.2, the earlier work compared the average profits truly experienced:

$$\hat{V}^k_{t-1} \equiv \frac{\sum_{s=1}^{t-1} \mathbf{1}(p_s = \bar{p}^k) \cdot v(\bar{p}^k, y_s, d_s)}{\mathcal{N}^k_{t-1}}, \tag{53}$$

where $v(p, y, d)$ is defined at (36). Our choice is realizable in the current discrete-item setting and being less affected by earlier errors, could encourage faster convergence.

Because $\mu < 1$, there exists the smallest $j = 2, 3, ...$ such that $j \geq (j + 1/\bar{k})^{\mu}$. Also, it will happen that $\mathcal{N}^k_{n\bar{k}} = n$ for $n = 1, ..., j$ and $k = 1, ..., \bar{k}$. Basically, there are initially a j number of $\bar{k}$-long cycles in each of which all prices are tried once in the *learning* mode.

6 Joint-control Bounds when Items are Perishable

For implementation and analysis purposes, it is actually beneficial to keep track of not only a minimizer of $\mathcal{N}_t^k$ for $t = 0, 1, ...$, but also an entire sequence in the ascending order. Thus, let $\kappa_t \equiv (\kappa_t(k))_{k=1,2,...,\bar{k}}$ be a permutation of the numbers $1, 2, ..., \bar{k}$ such that

$$\mathcal{N}_t^{\kappa_t(1)} \leq \mathcal{N}_t^{\kappa_t(2)} \leq \cdots \leq \mathcal{N}_t^{\kappa_t(\bar{k})}. \tag{54}$$

Each $\kappa_t(k)$ is the index of the price that has been visited the k-th fewest times by the end of period t. The policy can maintain such a κ_t for $t = 0, 1, ...$ Details are left to Section 2 of Appendix 12. We now make two important observations about LwD(μ).

Proposition 3 *For any period $t = 1, 2, ...$, it is true that that LwD(μ) will ensure*

$$\mathcal{N}_{t,0}^k < \left(\frac{t}{\bar{k}}\right)^{\mu} + 1, \qquad \forall k = 1, 2, ..., \bar{k}.$$

Proposition 4 *For any period $t = 1, 2, ...$, it is true that LwD(μ) will ensure*

$$\mathcal{N}_{t-1}^k \geq \left(\frac{t}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., \bar{k}.$$

Since $(t/\bar{k})^{\mu}$ is not necessarily an integer, being strictly less than $(t/\bar{k})^{\mu} + 1$ is different, albeit inconsequentially, from being less than $(t/\bar{k})^{\mu}$. All proofs of this section have been relegated to Appendix 11. While Proposition 3 gives an upper bound on the time spent on pure *learning*, Proposition 4 gives a guarantee on the amount of time each price choice k will be visisted. We already have (19) as an expression on how close f and the empirical distribution $\hat{f}_{t-1}$ based on it can be as t

grows to $+\infty$. For the current joint control, it can be convenient to have another expression. For

$$\delta_W(f,g,d) \equiv |\sum_{d'=0}^{d-1}[F_f(d') - F_g(d')]|, \tag{55}$$

the time-$(t-1)$ empirical distribution $\hat{f}_{t-1}$ defined through (10) would satisfy

$$\mathbb{P}_f\left[\delta_W(f,\hat{f}_{t-1},d) \geq \varepsilon\right] \leq 2\cdot\exp\left(-2\varepsilon^2\cdot(t-1)/d^2\right). \tag{56}$$

Indeed, note that $\sum_{d'=0}^{d-1}(1-F_f(d')) = \mathbb{E}_f[D\wedge d]$; meanwhile,

$$\begin{aligned}\sum_{d'=0}^{d-1}(1-\hat{F}_{t-1}(d')) &= \sum_{d'=0}^{d-1}\sum_{s=1}^{t-1}\mathbf{1}(D_s \geq d'+1)/(t-1)\\ &= \sum_{s=1}^{t-1}\sum_{d'=1}^{d}\mathbf{1}(D_s\geq d')/(t-1) = \sum_{s=1}^{t-1}(D_s\wedge d)/(t-1),\end{aligned} \tag{57}$$

where the first equality is due to (10) and $\hat{F}_{t-1}(d')$'s definition as the cdf for $\hat{f}_{t-1}$. Moreover, note the $D_s\wedge d$'s are independent random variables bounded in the range of $[0,d]$; also, $\mathbb{E}_f[D\wedge d] = \mathbb{E}_f[\sum_{s=1}^{t-1}(D_s\wedge d)/(t-1)]$. Thus, by Hoeffding's (20),

$$\begin{aligned}\mathbb{P}_f\left[\delta_W(f,\hat{f}_{t-1},d)\geq\varepsilon\right] &= \mathbb{P}_f\left[|\sum_{d'=0}^{d-1}(1-F_f(d')) - \sum_{d'=0}^{d-1}(1-\hat{F}_{t-1}(d'))|\geq\varepsilon\right]\\ &= \mathbb{P}_f\left[|\mathbb{E}_f[\sum_{s=1}^{t-1}(D_s\wedge d)/(t-1)] - \sum_{s=1}^{t-1}(D_s\wedge d)/(t-1)|\geq\varepsilon\right],\end{aligned} \tag{58}$$

and hence (56). Another useful bound is that, from (34), any $y\in\mathbb{N}$ would satisfy

$$(2y+1)\cdot\sum_{d=y}^{+\infty}(1-F_f(d)) \leq \sum_{d=y}^{+\infty}(2d+1)\cdot(1-F_f(d)) \leq \sum_{d=0}^{+\infty}(2d+1)\cdot(1-F_f(d)) \leq \bar{m}^2+\bar{s}^2. \tag{59}$$

Recall that $\tilde{V}_{t-1}^k$ defined at (47) provides a time-$(t-1)$ estimate on the per-period average profit of the price $\bar{p}^k$. Using (19), (26), (56), and (59), as well as Propositions 1 and 4, we can show the probabilistic convergence of the estimate $\tilde{V}_{t-1}^k$ to the true per-period profit $V_{f^k}^k \equiv V_{f^k}(\bar{p}^k, y_{f^k}^*)$ as defined through (37) and (38).

Proposition 5 *There exist positive constants* A^{Prop5}, B^{Prop5}, C^{Prop5}, D^{Prop5}, *and* E^{Prop5}, *such that for any* $k = 1, 2, ..., \bar{k}$ *and* $\varepsilon > 0$, *the probability* $\mathbb{P}_f\left[|\tilde{V}^k_{t-1} - V^k_{f^k}| \geq \varepsilon\right]$ *will be below*

$$A^{Prop5} \cdot \exp\left(-B^{Prop5} \cdot \varepsilon^2 \cdot t^{\mu/2}\right) + 2 \cdot \exp\left(-C^{Prop5} \cdot t^{\mu}\right),$$

when t *is greater than* $D^{Prop5} + E^{Prop5}/\varepsilon^{4/\mu}$. *The upper bound can be further written as* $A^{Prop5} \cdot \exp\left(-B^{Prop5} \cdot (\varepsilon^2 \wedge 1) \cdot t^{\mu/2}\right)$.

In the current joint control, all constants can be functions of the parameters $\bar{k}$, $\bar{p}^1, \bar{p}^2, ..., \bar{p}^{\bar{k}}$, $\bar{c}$, $\bar{h}$, $\bar{b}$, $\bar{m}$, and $\bar{s}$. For the regret of any joint-control policy $(\mathbf{k}, \mathbf{y})$ defined at (41), note that

$$\begin{aligned} R^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) &= V^*_{\mathbf{f}} \cdot T - V^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) = \textstyle\sum_{t=1}^T \left(V^{k^*_{\mathbf{f}}}_{f^{k^*_{\mathbf{f}}}} - \mathbb{E}_{\mathbf{f}}[V_{f^{k_t}}(\bar{p}^{k_t}, y_t)]\right) \\ &= \textstyle\sum_{t=1}^T \mathbb{E}_{\mathbf{f}}\left[V^{k^*_{\mathbf{f}}}_{f^{k^*_{\mathbf{f}}}} - V^{k_t}_{f^{k_t}}\right] + \sum_{t=1}^T \mathbb{E}_{\mathbf{f}}\left[V^{k_t}_{f^{k_t}} - V_{f^{k_t}}(\bar{p}^{k_t}, y_t)\right], \end{aligned} \tag{60}$$

where the first equality is from (41), the second equality is from (38) to (40), as well as the $V^k_{f^k}$-maximizing nature of $k^*_{\mathbf{f}}$, and the third equality is just an identity. In the second line, the first term can be attributed to sub-optimal prices, while the second term can be attributed to sub-optimal order-up-to levels. Suppose furthermore that, the policy $(\mathbf{k}, \mathbf{y})$ represents LwD(μ). Then, due to (37), we can rewrite (60) as something similar to (15):

$$R^T_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) = R^{T1}_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) + R^{T2}_{\mathbf{f}}(\mathbf{k}, \mathbf{y}), \tag{61}$$

where

$$R^{T1}_{\mathbf{f}}(\mathbf{k}, \mathbf{y}) = \sum_{t=1}^T \mathbb{E}_{\mathbf{f}}\left[V^{k^*_{\mathbf{f}}}_{f^{k^*_{\mathbf{f}}}} - V^{k_t}_{f^{k_t}}\right] + \sum_{t=1}^T \mathbb{E}_{\mathbf{f}}\left[V^{k_t}_{f^{k_t}} - V_{f^{k_t}}\left(\bar{p}^{k_t}, \hat{y}_t\right)\right], \tag{62}$$

$$R_{\mathbf{f}}^{T2}(\mathbf{k},\mathbf{y}) = \sum_{k=1}^{\bar{k}} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{f}} \left[\mathbf{1}(k_t = k) \cdot (Q_{f^k}(y_t) - Q_{f^k}(\hat{y}_t)) \right], \tag{63}$$

with all the y_t's and $\hat{y}_t$'s iteratively provided by (14) and (51). In (61), the first term $R_{\mathbf{f}}^{T1}(\mathbf{k},\mathbf{y})$ stands for the regret that an LwD(μ) policy will accrue if items are allowed to perish at the end of each period; the second term $R_{\mathbf{f}}^{T2}(\mathbf{k},\mathbf{y})$ captures the additional regret due to inventory carry-overs. In the remainder of this section, we focus on bounding the first term $R_{\mathbf{f}}^{T1}(\mathbf{k},\mathbf{y})$ given in (62). To this end, define

$$\delta V_{\mathbf{f}}^{k} \equiv V_{\mathbf{f}}^{*} - V_{f^k}^{k} = V_{f^{k_{\mathbf{f}}^*}}^{k_{\mathbf{f}}^*} - V_{f^k}^{k} \geq 0. \tag{64}$$

It measures the difference in average single-period profit between using the price choice k and making the best choice $k_{\mathbf{f}}^*$. Due to the nature of the LwD(μ) policy, (62) will become

$$R_{\mathbf{f}}^{T1}(\mathbf{k},\mathbf{y}) = T_1 + T_2 + T_3, \tag{65}$$

with

$$T_1 = \sum_{k \neq k_{\mathbf{f}}^*} \delta V_{\mathbf{f}}^{k} \cdot \mathbb{E}_{\mathbf{f}}[\mathcal{N}_{T,0}^{k}], \tag{66}$$

$$T_2 = \sum_{k \neq k_{\mathbf{f}}^*} \delta V_{\mathbf{f}}^{k} \cdot \sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}} \left[m_t = 1 \text{ and } \max_{k' \neq k} \tilde{V}_{t-1}^{k'} \leq \tilde{V}_{t-1}^{k} \right], \tag{67}$$

where $m_t = 0$ or 1 stands for *learning* or *doing* while $\delta V_{\mathbf{f}}^{k}$ has been defined at (64), and

$$T_3 = \sum_{k=1}^{\bar{k}} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{f}} \left[\mathbf{1}(k_t = k) \cdot \left(V_{f^k}^{k} - V_{f^k}\left(\bar{p}^k, \hat{y}_t\right) \right) \right]. \tag{68}$$

The terms in (65) blame the perishable-item regret on three sources: time that is

spent on *learning*, errors caused by inaccuracies of the $\tilde{V}^k_{t-1}$'s in representing the actual $V^k_{f^k}$'s—the k that maximizes the former is not necessarily the $V^k_{f^k}$-maximizing $k^*_{\mathbf{f}}$, and the non-optimality of the $\hat{y}_t$'s under the actual demand distributions f^k. It is quite clear that Proposition 3 is tailor-made for bounding the first term (66). For the second term (67), note $m_t = 1$ is not always true in period t; also, for $\tilde{V}^k_{t-1}$ defined in (47) to achieve the maximum among $k = 1, 2, ..., \bar{k}$, it must happen that $\tilde{V}^{k^*_{\mathbf{f}}}_{t-1} \leq \tilde{V}^k_{t-1}$. So in view of (67), we can obtain

$$T_2 \leq \sum_{k \neq k^*_{\mathbf{f}}} \delta V^k_{\mathbf{f}} \cdot \sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}^{k^*_{\mathbf{f}}}_{t-1} \leq \tilde{V}^k_{t-1} \right]. \tag{69}$$

Going forward, the inequalities (56) and (59), as well as Proposition 5 will be useful. Meanwhile, the bounding of the third term (68) can resort to Proposition 1. The situation where one price clearly dominates all other prices by a guaranteed margin is easier to tackle. Let us deal with this first. Now let

$$\delta V^*_{\mathbf{f}} \equiv \min_{k \neq k^*_{\mathbf{f}}} \delta V^k_{\mathbf{f}}, \tag{70}$$

i.e., the minimum gap in single-period profits between optimal and non-optimal price choices. For any $\delta > 0$, we use $\mathcal{F}^{\bar{k}}_2(\bar{m}, \bar{s}, \delta)$ to denote the subset of $\mathbf{f}$'s in $(\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$ that have one price leading other choices by at least a δ-margin:

$$\mathcal{F}^{\bar{k}}_2(\bar{m}, \bar{s}, \delta) \equiv \left\{ \mathbf{f} \in (\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}} : \ k^*_{\mathbf{f}} \text{ is unique and } \delta V^*_{\mathbf{f}} \geq \delta \right\} \subset (\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}. \tag{71}$$

The performance of LwD(μ) can be well bounded when $\mathbf{f}$ is known to come from $\mathcal{F}^{\bar{k}}_2(\bar{m}, \bar{s}, \delta)$.

Proposition 6 *Let* $(\mathbf{k}, \mathbf{y})$ *be the adaptive policy generated from following LwD(μ) for some* $\mu \in [1/2, 1)$. *Then,*

$$T_1 \leq A'' + B'' \cdot T^{\mu}, \tag{72}$$

for some positive constants A'' and B''. Also,

$$\sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^{k_{\mathbf{f}}^*} \leq \tilde{V}_{t-1}^{k}\right] \leq C'' + \frac{D''}{(\delta V_{\mathbf{f}}^{k})^{4/\mu}}, \tag{73}$$

for some positive constants C'' and D''; this then leads to

$$T_2 \leq E'' + (\bar{k}-1)\cdot \frac{F''}{\delta^{\,4/\mu-1}}, \tag{74}$$

for some positive constants E'' and F''. In addition,

$$T_3 \leq G'' + H''\cdot T^{1/2}\cdot (\ln T)^{1/2}, \tag{75}$$

for some positive constants G'' and H''. Consequently, for any $\delta > 0$, there are constants $A^{Prop6}{}_{\delta}$, B^{Prop6}, and C^{Prop6}, such that

$$R_{\mathbf{f}}^{T1}(\mathbf{y},\mathbf{k}) \leq A^{Prop6}{}_{\delta} + B^{Prop6}\cdot T^{\mu} + C^{Prop6}\cdot T^{1/2}\cdot (\ln T)^{1/2},$$

for any $\mathbf{f} \in \mathcal{F}_2^{\bar{k}}(\bar{m},\bar{s},\delta)$; however, $\lim_{\delta\to 0^+} A^{Prop6}{}_{\delta} = +\infty$. Also, $\mu = 1/2$ is the choice with the tightest guarantee among the LwD(μ) policies. It will achieve an $O(T^{1/2}\cdot(\ln T)^{1/2})$-bound.

The task of bounding will be more demanding when $\mathbf{f}$ is truly free to roam in $(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}$. Analysis relying on the inequalities stated in Proposition 6 will result in the following.

Proposition 7 *Let $(\mathbf{k},\mathbf{y})$ be the adaptive policy generated from following LwD(μ) for some $\mu \in [1/2, 1)$. Then, there are positive constants A^{Prop7} and B^{Prop7} such that*

$$\sup_{\mathbf{f}\in(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}} R_{\mathbf{f}}^{T1}(\mathbf{k},\mathbf{y}) \leq A^{Prop7} + B^{Prop7}\cdot T^{\mu\vee(1-\mu/4)}.$$

The choice $\mu = 4/5$ will achieve an $O(T^{4/5})$-bound.

Comparing Propositions 6 and 7, one might say that those incidences $\mathbf{f}$ residing in $(\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}} \setminus \mathcal{F}_2^{\bar{k}}(\bar{m}, \bar{s}, \delta)$ for ever smaller δ's are "trouble makers" that render the optimal price illusory to catch. On the other hand, the nonperishability of items seems to present an even more challenging problem for the current joint inventory-price control. We will spend the next two sections on it. In Section 7, we first suppose that the demand-distribution vector $\mathbf{f}$ is restricted to $\mathcal{F}_2^{\bar{k}}(\bar{m}, \bar{s}, \delta)$ for some $\delta > 0$. Then in Section 8, we move on to the general case where $\mathbf{f}$ can come from anywhere in $(\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$.

7 Nonperishability with Restricted Demand Patterns

This section is devoted to the bounding of $R_{\mathbf{f}}^{T2}(\mathbf{k},\mathbf{y})$ as given in (63) for $\mathbf{f}\in\mathcal{F}_2^{\bar{k}}(\bar{m},\bar{s},\delta)$ where $\delta>0$. Our plan is to show that the dominant price say $\bar{p}^1$ will be used uninterruptedly for long sequences of periods; meanwhile, the sub-linear growth of nonperishability-induced regrets in sequences' lengths, as evidenced in Proposition 2, will help to bound the T-period regret overall. Another idea involves the use of *virtual* learning periods.

To this end, let $r_f^t(x_1)$ be almost the same nonperishability-induced pure-control regret under demand distribution f as the $R_f^{T2}(\mathbf{y})$-term defined at (17). But now, we let it be from period 1 to a variable period t, and let the starting inventory level be some arbitrary integer $x_1\leq\bar{d}$. Due to (51), the order-up-to levels used by a LwD(μ) policy in the current joint-control case are also bounded by $\bar{d}$, and hence the resultant starting inventory levels under it are below $\bar{d}$ as well. Proposition 2 can be understood as a bound for $r_f^T(0)$. But a close scrutiny would reveal that a bound of the same form works for $r_f^t(x_1)$ regardless of the valuation of x_1. Actually, the only changes needed in the proof would be to replace "$t=3$" with "$t=1$" and "$T-2$" with "T". Hence, for the positive constants A^{Prop2} and B^{Prop2},

$$r_f^t(x_1)\leq A^{Prop2}+B^{Prop2}\cdot t^{1/2}\cdot(\ln t)^{3/2}, \tag{76}$$

for any $t\in\mathbb{N}$, $f\in\mathcal{F}_1(\bar{m})$, and integer $x_1\leq\bar{d}$.

We also find it convenient to condition on when *learning* has happened. For any $t\in\mathbb{N}$, let $\mathcal{M}(t)\subseteq\{0,1\}^t$ be the set of all potential *learning/doing*-mode sequences $m\equiv(m_s)_{s=1,2,...,t}$ over the first t periods. Given any $m\equiv(m_s)_{s=1,2,...,t}\in\mathcal{M}(t)$, we can use $\mathcal{N}_{t,0}(m)=\sum_{s=1}^t(1-m_s)$ to denote the total number of learning periods

under the mode sequence m of the first t periods. Due to Proposition 3,

$$\mathcal{N}_{t,0}(m) \leq \bar{k} \cdot \left[\left(\frac{t}{\bar{k}} \right)^{\mu} + 1 \right] = \bar{k}^{1-\mu} \cdot t^{\mu} + \bar{k}, \qquad \forall m \in \mathcal{M}(t). \tag{77}$$

Let $s_1(m)$, $s_2(m)$, ..., $s_{\mathcal{N}_{t,0}(m)}(m)$ be the periods in which learning takes place. Certainly,

$$s_i(m) = \min \left\{ s = 1, ..., t : \sum_{\tau=1}^{s} (1 - m_\tau) \geq i \right\}, \qquad \forall i = 1, .., \mathcal{N}_{t,0}(m), \tag{78}$$

and hence

$$\mathcal{N}_{t,0}(m) = \max\{i = 1, 2, ... : \ s_i(m) \leq t\}. \tag{79}$$

For convenience, use $M(t)$ for the random mode sequence that has actually happened. Note that $\sum_{m \in \mathcal{M}(t)} \mathbb{P}_{\mathbf{f}}[M(t) = m] = 1$. Without loss of generality, designate the choice 1 as $k^*_{\mathbf{f}}$. Conditioned on $M(t) = m$, the chance for $\tilde{V}^1_{t-1}$ to be greater than any other $\tilde{V}^k_{t-1}$ plus a margin say $\delta V^*_{\mathbf{f}}/2$ can be shown to be ever closer to one as t increases. Recall that the $\tilde{V}^l_{t-1}$'s defined at (47) are used in the learning mode for the selection of the winning price, and $\delta V^*_{\mathbf{f}}$ defined at (70) registers the profit gap between the best and second best prices.

Proposition 8 *There exist positive constants A^{Prop8}, B^{Prop8}, C^{Prop8}, and D^{Prop8} such that for any mode sequence $m \in \mathcal{M}(t)$ realized for $M(t)$, whenever $t \geq A^{Prop8} + B^{Prop8}/(\delta V^*_{\mathbf{f}})^{4/\mu}$,*

$$\mathbb{P}_{\mathbf{f}} \left[\tilde{V}^1_{t-1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{t-1} + \frac{\delta V^*_{\mathbf{f}}}{2} | M(t) = m \right] \geq 1 - C^{Prop8} \cdot \exp\left(-D^{Prop8} \cdot (\delta V^*_{\mathbf{f}} \wedge 1)^2 \cdot t^{\mu/2} \right).$$

Proofs of this section can be found in Appendix 11. Furthermore, given that $\tilde{V}^1_{t-1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{t-1} + \delta V^*_{\mathbf{f}}/2$ has already occurred, we can show that the status of

$\tilde{V}^1_{t+\tau-1} \geq \tilde{V}^k_{t+\tau-1}$ will be maintainable for quite large τ's. We first present an intermediate result.

Proposition 9 *For any price choice $k = 1, 2, ..., \bar{k}$ and periods t and t',*

$$|\tilde{V}^k_{t'-1} - \tilde{V}^k_{t-1}| \leq (\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \delta_V(\tilde{f}^k_{t-1}, \tilde{f}^k_{t'-1}, \bar{d}) \\ +(\bar{p}^k - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}^k_{t-1} + 1) + (\bar{p}^k - \bar{c} + \bar{b}) \cdot \delta_W(\tilde{f}^k_{t-1}, \hat{f}^k_{t'-1}, \tilde{d}^k_{t-1}).$$

In addition, for any integer $\tau \geq 0$ and $d = 0, 1, ..., \tilde{d}^k_{t-1} - 1$, whenever $\mathcal{N}^k_{t+\tau-1} = \mathcal{N}^k_{t-1} + \tau$,

$$|\tilde{F}^k_{t-1}(d) - \tilde{F}^k_{t+\tau-1}(d)| \leq \frac{\tau}{\mathcal{N}^1_{t-1}}.$$

The cutoff level $\tilde{d}^k_{t-1}$ is defined at (45), the distribution $\tilde{f}^k_{t-1}$ is defined at (46), and the distance δ_W is defined at (55). We can now reach the longevity of one dominant price.

Proposition 10 *There are positive constants A^{Prop10}, B^{Prop10}, and C^{Prop10} such that whenever $t \geq A^{Prop10} + B^{Prop10}/(\delta V^*_{\mathbf{f}})^{4/\mu}$ and $\tilde{V}^1_{t-1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{t-1} + \delta V^*_{\mathbf{f}}/2$, the LwD($\mu$) policy will ensure that $k_t = k_{t+1} = \cdots = k_{t+t'-1} = 1$ for t' as large as $C^{Prop10} \cdot \delta V^*_{\mathbf{f}} \cdot \left((t/\bar{k})^\mu - 1\right)^{3/4}$, as long as no learning is to emerge in the periods $t, t+1, ..., t+t'-1$.*

Without loss of generality, we can suppose that $A^{Prop8} \leq A^{Prop10}$ and $B^{Prop8} \leq B^{Prop10}$. Let period $t \geq A^{Prop10} + B^{Prop10}/(\delta V^*_{\mathbf{f}})^{4/\mu}$. Proposition 8 predicts that going forward, the future profit indicator of price choice $k^*_{\mathbf{f}} = 1$ will lead those of the other prices by a comfortable margin with a probability that converges to 1 quickly as t grows; also, Proposition 10 predicts that once leading by a comfortable margin, the price choice 1 will be maintained for a long time unless it is interrupted by *learning*. So far we have upper bounds on learning frequencies in the forms of Proposition 3 and (77). To utilize Propositions 8 and 10, it will help to have lower bounds as well.

To this end, we introduce *virtual* learning periods that will allow the frequencies of learning, be it actual or virtual, to be regulated on both sides. Let $G_{\mu,\delta}$ be a constant strictly above $(4 \cdot \bar{k}^{3\mu/4})/(C^{Prop10} \cdot \delta)$, where C^{Prop10} is the constant in Proposition 10 and δ is the margin that appeared in (71); for instance,

$$G_{\mu,\delta} = \frac{4 \cdot \bar{k}^{3\mu/4}}{C^{Prop10} \cdot \delta} + 1. \tag{80}$$

Also, let $I_{\mu,\delta} \geq \bar{k}$ be large enough so that both $\lceil (1 + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \rceil \geq 1$ and $\lceil (2 + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \rceil - \lceil (1 + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \rceil \geq 2$. Note that both $G_{\mu,\delta}$ and $I_{\mu,\delta}$ will grow to $+\infty$ when δ approaches 0^+. Now define $s'_i = \lceil (i + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \rceil$ for $i = 1, 2, ...$ as virtual learning periods. We have omitted expressing the dependence of the s'_i's on (μ, δ) for simplicity. The size of $I_{\mu,\delta}$ will guarantee that $1 \leq s'_1 < s'_2 < \cdots$. Similarly to (79), let

$$\mathcal{N}'_{t,0} = \max\{i = 1, 2, ... : \ s'_i \leq t\}. \tag{81}$$

It stands for the number of virtual learning periods by time t. Let $L'(t) = \{s'_1, s'_2, ..., s'_{\mathcal{N}'_{t,0}}\}$ be the set of virtual learning periods up to t. For $m \in \mathcal{M}(t)$, let $L(m,t) = \{s_1(m), s_2(m), ..., s_{\mathcal{N}_{t,0}(m)}(m)\}$ be the set of actual learning periods up to t. A *virtual learning* period can be either an actual *learning* or actual *doing* period. Now consider the combined set $L''(m,t) = L(m,t) \cup L'(t)$. We can write $L''(m,t)$ as $\{s''_1(m), s''_2(m), ..., s''_{\mathcal{N}''_{t,0}(m)}(m)\}$ with $1 \leq s''_1(m) < s''_2(m) < \cdots < s''_{\mathcal{N}''_{t,0}(m)}(m) \leq t$ and each $s''_i(m)$ being either some $s_j(m)$ or some s'_l or both. Certainly,

$$\mathcal{N}''_{t,0}(m) \leq \mathcal{N}_{t,0}(m) + \mathcal{N}'_{t,0}, \tag{82}$$

The frequencies at which the combined learning periods $s''_i(m)$ arise can be

constrained in both directions.

Proposition 11 *We have the following useful inequalities:*

$$\mathcal{N}''_{t,0}(m) \leq H_{\mu,\delta} \cdot t^{\mu}, \tag{83}$$

for some constant $H_{\mu,\delta}$ *which is above* $G_{\mu,\delta}$ *and hence in satisfaction of* $\lim_{\delta \to 0^+} H_{\mu,\delta} = +\infty$*;*

$$s''_i(m) \geq \left(\frac{1}{H_{\mu,\delta}^{\ 1/\mu}} \right) \cdot i^{1/\mu}; \tag{84}$$

$$s''_i(m) \leq \left(\frac{1}{G_{\mu,\delta}^{\ 1/\mu}} \right) \cdot (i + I_{\mu,\delta})^{1/\mu} + 1; \tag{85}$$

$$s''_{i+1}(m) - s''_i(m) \leq \left[\left(\frac{4}{G_{\mu,\delta}} \right) \cdot \left(s''_i(m) \right)^{1-\mu} + 1 \right] \vee \left\lceil \frac{(1 + I_{\mu,\delta})^{1/\mu}}{G_{\mu,\delta}^{\ 1/\mu}} \right\rceil. \tag{86}$$

Note that (83) and (84) bound the combined learning frequencies from above, while (85) and (86) bound them from below.

We now utilize (76) and Propositions 8 to 11 to bound $R^{T2}_{\mathbf{f}}(\mathbf{k}, \mathbf{y})$ when $(\mathbf{k}, \mathbf{y})$ comes from the LwD(μ) policy and $\mathbf{f} \in \mathcal{F}_2^{\bar{k}}(\bar{m}, \bar{s}, \delta)$. For convenience, let $s''_{\mathcal{N}''_{T,0}(m)+1}(m) = T + 1$. Now for any $i = 1, 2, ..., \mathcal{N}''_{T,0}(m)$, let $N_i(m)$ be the random number of consecutive same-price *doing* sequences from period $s''_i(m) + 1$ to period $s''_{i+1}(m) - 1$. We do allow $N_i(m) = 0$ when $s''_{i+1}(m) = s''_i(m) + 1$; otherwise, $N_i(m)$ is integer-valued between 1 and $s''_{i+1}(m) - s''_i(m) - 1$. Suppose, for instance, that $s''_i(m) = 10$, $s''_{i+1}(m) = 16$, the price choice 1 is used in periods 11 and 12, the price choice 2 is used in periods 13 and 14, and the price choice 1 is used again in period 15. Then, $N_i(m)$ would be 3 because there have been three same-price consecutive sequences in periods 11 to 15. Define $U_{i,1}(m), ..., U_{i,N_i(m)+1}(m)$ so that

$$s''_i(m) + 1 = U_{i,1}(m) < U_{i,2}(m) < \cdots < U_{i,N_i(m)}(m) < U_{i,N_i(m)+1}(m) = s''_{i+1}(m), \tag{87}$$

and for each $j = 1, ..., N_i(m)$, periods $U_{i,j}(m), U_{i,j}(m)+1, ..., U_{i,j+1}(m)-1$ form the (i,j)-segment, i.e., a consecutive same-price *doing* sequence. For the segment, denote the price choice by $K_{i,j}(m) \equiv k_{U_{i,j}(m)} = k_{U_{i,j}(m)+1} = \cdots = k_{U_{i,j+1}(m)-1}$ and the starting inventory level by $X_{i,j}(m)$. For the previous example, $U_{i.1}(m) = 11$, $U_{i,2}(m) = 13$, $U_{i,3}(m) = 15$, and $U_{i,4}(m) = 16$; also, $K_{i,1}(m) = 1$, $K_{i,2}(m) = 2$, and $K_{i,3}(m) = 1$.

Now (63) can be rewritten as

$$R_{\mathbf{f}}^{T2}(\mathbf{k}, \mathbf{y}) = T_1 + T_2, \tag{88}$$

where

$$T_1 = \sum_{m \in \mathcal{M}(T)} \mathbb{P}_{\mathbf{f}}[M(T) = m] \cdot \mathbb{E}_{\mathbf{f}} \left[\sum_{i=1}^{\mathcal{N}''_{T,0}(m)} r^1_{f^{K_{s''_i(m)}}} \left(X_{s''_i(m)} \right) | M(T) = m \right], \tag{89}$$

$$T_2 = \sum_{m \in \mathcal{M}(T)} \mathbb{P}_{\mathbf{f}}[M(T) = m] \cdot \theta_2(m), \tag{90}$$

and

$$\theta_2(m) = \mathbb{E}_{\mathbf{f}} \left[\sum_{i=1}^{\mathcal{N}''_{T,0}(m)} \sum_{j=1}^{N_i(m)} r^{U_{i,j+1}(m) - U_{i,j}(m)}_{f^{K_{i,j}(m)}} \left(X_{i.j}(m) \right) | M(T) = m \right]. \tag{91}$$

In (88), the nonperishability-induced regret is partitioned into two parts, the part T_1 that is accrued over *combined-learning* periods and the part T_2 that is accrued over *doing* periods that are not even *virtual learning* ones. In both (89) and (90), we sum over all potential $m \in \mathcal{M}(t)$ that could be realized for the random $M(t)$; recall that $r^t_f(x_1)$ is defined around (76). The way the term $\theta_2(m)$ is expressed in (91) keeps track of price-switching epochs.

A T^{μ}-sized bound can be easily identified for T_1 due to (83). From (91), we also

have $\theta_2(m) = \eta_2(m) + \zeta_2(m)$, where

$$\eta_2(m) = \sum_{i=1}^{\mathcal{N}''_{T,0}(m)} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}^1_{s''_i(m)+1} \leq \max_{k=2}^{\bar{k}} \tilde{V}^k_{s''_i(m)+1} + \delta V^*_{\mathbf{f}}/2 | M(T) = m \right] \times \\ \times \mathbb{E}_{\mathbf{f}} [\sum_{j=1}^{N_i(m)} r^{U_{i,j+1}(m) - U_{i,j}(m)}_{f^{K_{i,j}(m)}} (X_{i.j}(m))| \\ | M(T) = m \text{ and } \tilde{V}^1_{s''_i(m)+1} \leq \max_{k=2}^{\bar{k}} \tilde{V}^k_{s''_i(m)+1} + \delta V^*_{\mathbf{f}}/2], \tag{92}$$

and

$$\zeta_2(m) = \sum_{i=1}^{\mathcal{N}''_{T,0}(m)} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}^1_{s''_i(m)+1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{s''_i(m)+1} + \delta V^*_{\mathbf{f}}/2 | M(T) = m \right] \times \\ \times \mathbb{E}_{\mathbf{f}} [\sum_{j=1}^{N_i(m)} r^{U_{i,j+1}(m) - U_{i,j}(m)}_{f^{K_{i,j}(m)}} (X_{i.j}(m))| \\ | M(T) = m \text{ and } \tilde{V}^1_{s''_i(m)+1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{s''_i(m)+1} + \delta V^*_{\mathbf{f}}/2]. \tag{93}$$

For every mode sequence m, the terms $\eta_2(m)$ and $\xi_2(m)$ capture two different types of cases for the total regret over the i-th inter-learning interval lasting from $s''_i(m)+1$ to $s''_{i+1}(m)-1$ for all the i's equaling $1, 2, ..., \mathcal{N}''_{T,0}(m)$. They take advantage of the interchangeability between expectation and summation. The first term is dedicated to the cases where the price-1 profit estimates $\tilde{V}^1_{s''_i(m)+1}$ fail to lead other estimates by the margin $\delta V^*_{\mathbf{f}}/2$ in the starting periods $s''_i(m)+1$, and the second term $\xi_2(m)$ is devoted to the opposite cases. By Proposition 8, the probabilities within (92) are small. This point will eventually lead to a constant bound for $\eta_2(m)$. Due to Proposition 10 and some bounds in Proposition 11, the number $N_i(m)$ can be shown to be 1 under an m specified in (93). We can then utilize (76) and other bounds in Proposition 11 to bound $\zeta_2(m)$ by a $T^{(1+\mu)/2}$-sized term.

Proposition 12 *When $\mu \in [4/7, 1)$, it is true that T_1 of (89) has a T^{μ}-sized bound and T_2 of (90) has a $T^{(1+\mu)/2}$-sized bound. Overall, there are positive constants $A^{Prop12}_{\mu,\delta}$, $B^{Prop12}_{\mu,\delta}$, and $C^{Prop12}_{\mu,\delta}$, such that for the LwD(μ) policy $(\mathbf{k}, \mathbf{y})$ and any*

$\mathbf{f} \in \mathcal{F}_2^{\bar{k}}(\bar{m}, \bar{s}, \delta)$,

$$R_{\mathbf{f}}^{T2}(\mathbf{k}, \mathbf{y}) \leq A^{Prop12}_{\mu,\delta} + B^{Prop12}_{\mu,\delta} \cdot T^{\mu} + C^{Prop12}_{\mu,\delta} \cdot T^{(1+\mu)/2} \cdot (\ln T)^{5/2};$$

however, $\lim_{\delta \to 0^+} B^{Prop12}_{\mu,\delta} = +\infty$ *while it can happen that* $\lim_{\delta \to 0^+} C^{Prop12}_{\mu,\delta} = 0$.

Even though the third term on the right-hand side of Proposition 12's signature inequality dominates the second term, the limiting behaviors of the coefficients have prompted us to keep the dominated term. They also give us some hope that a bound somewhere between the orders of T^{μ} and $T^{(1+\mu)/2}$ might be achievable for the limiting case where $\delta = 0$. In addition, the requirement that $\mu \geq 4/7$ is newly added to ensure the analysis.

For the time being, by putting Propositions 6 and 12 together, we can obtain a bound for the LwD(μ) policy that accounts for the nonperishability of items.

Theorem 2 *Let* $(\mathbf{k}, \mathbf{y})$ *be the policy generated from following LwD(*μ*) for* $\mu \in [4/7, 1)$. *Then, for any* $\delta > 0$, *there are constants* $A^{Them2}_{\mu,\delta}$, $B^{Them2}_{\mu,\delta}$, *and* $C^{Them2}_{\mu,\delta}$, *such that*

$$\sup_{\mathbf{f} \in \mathcal{F}_2^{\bar{k}}(\bar{m}, \bar{s}, \delta)} R_{\mathbf{f}}^{T}(\mathbf{y}, \mathbf{k}) \leq A^{Them2}_{\mu,\delta} + B^{Them2}_{\mu,\delta} \cdot T^{\mu} + C^{Them2}_{\mu,\delta} \cdot T^{(1+\mu)/2} \cdot (\ln T)^{5/2}.$$

Also, the choice $\mu = 4/7$ *will achieve an* $O(T^{11/14} \cdot (\ln T)^{5/2})$*-bound.*

The restriction on $\delta > 0$ is the last barrier for us to overcome.

8 Nonperishability with Arbitrary Demand Patterns

We now come around to bound the general case involving both nonperishable items and more arbitrary demand patterns. For the LwD(μ) policies, there is now no guarantee that any single price choice would be kept long enough. While still utilizing (76), our remedy is to consider *stickier* variants that allow longer reigns of incumbent price choices.

Fix some $\nu > 2\bar{k}^{\mu/4} \cdot (\bar{m}^2 + \bar{s}^2) \cdot \max_{k=1}^{\bar{k}} (\bar{p}^k - \bar{c} + \bar{b})$ and $\psi \in [\mu/2, 3\mu/4)$. Our sticky policy associated with parameters μ, ν, and ψ will ensure that, once $\tilde{V}_{t-1}^{k_t} \geq \max_{k \neq k_t} \tilde{V}_{t-1}^{k}$ for some $k_t = 1, 2, ..., \bar{k}$ has happened, k_t will keep on being chosen for periods $t, t+1, ..., t+t'-1$ as long as there is no interruption from learning and $\tilde{V}_{t+\tau-1}^{k_t} \geq \max_{k \neq k_t} \tilde{V}_{t+\tau-1}^{k} - \nu/(t+\tau-1)^{3\mu/4-\psi}$ for $\tau = 1, 2, ..., t'$. Before spelling out any other details about these policies, the following would already verify that t' can be quite large just because of this one common sticky property.

Proposition 13 *Suppose t is large enough so that*

$$\left(\left(\frac{t}{\bar{k}} \right)^{\mu} - 1 \right)^{1/4} \geq \left[2 \cdot \max_{k=1}^{\bar{k}} (\bar{p}^k - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2) \cdot \frac{t^{3\mu/4-\psi}}{\nu} \right] \vee \bar{d}.$$

Then, for t' as large as some $C^{Prop13} \cdot \nu \cdot \left((t/\bar{k})^{\mu} - 1 \right)^{3/4} / t^{3\mu/4-\psi}$ where C^{Prop13} is a positive constant, any sticky policy would ensure that $k_t = k_{t+1} = \cdots = k_{t+t'-1}$ for any $k_t = 1, 2, ..., \bar{k}$ as long as $\tilde{V}_{t-1}^{k_t} \geq \max_{k \neq k_t} \tilde{V}_{t-1}^{k}$ and there is no learning in periods $t, t+1, ..., t+t'-1$.

Proofs of this section, except that for Theorem 4, have been put into Appendix 11. We now supply more details about our stickier variant LwD$'(\mu, \nu, \psi)$ of LwD(μ).

First, a redefinition of virtual learning periods is needed. Let

$$G_{\mu,\nu} = \frac{4 \cdot \bar{k}^{3\mu/4}}{C^{Prop13} \cdot \nu} + 1, \tag{94}$$

where C^{Prop13} is a constant that fits Proposition 13. Also, let $I_{\mu,\nu,\psi} \geq \bar{k}$ be the smallest integer such that both $\lceil (1 + I_{\mu,\nu,\psi})^{1/(1-\psi)} / G_{\mu,\nu}^{1/(1-\psi)} \rceil \geq 1$ and $\lceil (2 + I_{\mu,\nu,\psi})^{1/(1-\psi)} / G_{\mu,\nu}^{1/(1-\psi)} \rceil - \lceil (1 + I_{\mu,\nu,\psi})^{1/(1-\psi)} / G_{\mu,\nu}^{1/(1-\psi)} \rceil \geq 2$. Now redefine $s'_i = \lceil (i + I_{\mu,\nu,\psi})^{1/(1-\psi)} / G_{\mu,\nu}^{1/(1-\psi)} \rceil$ for $i = 1, 2, ...$ as virtual learning periods. We have omitted expressing the dependence of the s'_i's on (μ, ν, ψ) for simplicity. Again, define $\mathcal{N}'_{t,0}$ through (81), as the number of virtual learning episodes by time t. The policy LwD$'(\mu, \nu, \psi)$ favors incumbent price choices in most of the *doing* periods. In particular, this new stickier policy shares the same *learning* periods as LwD(μ). Also, it behaves the same as the original one except in *doing* periods t satisfying

$$\text{neither } \ m_{t-1} = 0 \quad \text{nor } \ t - 1 = s'_i \ \text{ for some } i. \tag{95}$$

For such a period, the only difference lies in replacing the original step 1.2 with the following:

1.2′. let price choice k^* be a maximizer of $\tilde{V}_{t-1}^k$ from $k = 1, 2, ..., \bar{k}$. If

$$\tilde{V}_{t-1}^{k^*} \geq \tilde{V}_{t-1}^{k_{t-1}} + \frac{\nu}{t^{3\mu/4-\psi}},$$

let $k_t = k^*$; otherwise, let $k_t = k_{t-1}$.

Certainly, we will have $k_t = k_{t-1}$ when the incumbent choice k_{t-1} happens to be k^*. Otherwise, now there is a $\nu/t^{3\mu/4-\psi}$-sized threshold for the profit estimate to cross before the price choice switches from the incumbent k_{t-1} to the new k^*. Note that LwD(μ) could somehow be understood as LwD$'(\mu, 0, \psi')$ for any $\psi' \in [\mu/2, 3\mu/4)$, though Proposition 13 would not necessarily apply in view of its requirement on ν's

range. On the flip side, Propositions 3 and 4 are hinged on the schedule of *learning* epochs and not affected by the change in some other periods. So they will remain applicable to the new policy LwD$'(\mu,\nu,\psi)$.

Let $L'(t)=\{s'_1,s'_2,...,s'_{\mathcal{N}'_{t,0}}\}$ be the set of virtual learning periods up to t. For $m\in\mathcal{M}(t)$, consider the combined set $L''(m,t)=L(m,t)\cup L'(t)$, where $L(m,t)$ is still the set of actual learning periods as specified by LwD(μ) under m. We can write $L''(m,t)$ as $\{s''_1(m),s''_2(m),...,s''_{\mathcal{N}''_{t,0}(m)}(m)\}$ with $1\le s''_1(m)<s''_2(m)<\cdots<s''_{\mathcal{N}''_{t,0}(m)}(m)\le t$ and each $s''_i(m)$ being either some $s_j(m)$ or some s'_l or both. We still have (82) for the new combined learning periods. Now the condition (95) for step 1.2$'$ to be executed in a *doing* period t is that it not be some $s''_i(m)+1$ under the sequence m.

For $R^{T1}_{\mathbf{f}}(\mathbf{k},\mathbf{y})$ defined at (62), the decomposition at (65) can be kept intact without any change on T_1 of (66) or T_3 of (68). On the other hand, we can replace (67) with

$$T_2=\sum_{k\neq k^*_{\mathbf{f}}}\sum_{t=1}^{T}\delta V^k_{\mathbf{f}}\cdot\mathbb{P}_{\mathbf{f}}\left[m_t=1\text{ and }\max_{k'\neq k}\tilde{V}^{k'}_{t-1}\le\tilde{V}^k_{t-1}+\frac{\nu}{t^{3\mu/4-\psi}}\right]. \tag{96}$$

The extra term $\nu/t^{3\mu/4-\psi}$ comes from the fact that sometimes a price choice can have its estimate worse off by as much as this amount and yet still keep its place. Without keeping track of the incumbent prices, the above is an overestimate. But that is allowed for an upper bound. Indeed, in the same vein of (69),

$$T_2\le\sum_{k\neq k^*_{\mathbf{f}}}\sum_{t=1}^{T}\delta V^k_{\mathbf{f}}\cdot\mathbb{P}_{\mathbf{f}}\left[\tilde{V}^{k^*_{\mathbf{f}}}_{t-1}\le\tilde{V}^k_{t-1}+\frac{\nu}{t^{3\mu/4-\psi}}\right]. \tag{97}$$

We have the following adaptation of Proposition 7.

Proposition 14 *Let* $(\mathbf{k},\mathbf{y})$ *be the policy generated from LwD$'(\mu,\nu,\psi)$. Then, there*

are positive constants $A^{Prop14}_{\mu,\nu,\psi}$ *and* $B^{Prop14}_{\mu,\nu,\psi}$ *such that*

$$\sup_{\mathbf{f}\in(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}} R^{T1}_{\mathbf{f}}(\mathbf{k},\mathbf{y}) \leq A^{Prop14}_{\mu,\nu,\psi} + B^{Prop14}_{\mu,\nu,\psi} \cdot T^{\mu\vee(1-3\mu/4+\psi)}.$$

Any choice with $\mu = 4/5$ *and* $\psi = 2/5$ *will achieve an* $O(T^{4/5})$*-bound.*

Propositions 7 and 14 both have $T^{4/5}$-sized bounds. So the LwD$'(\mu,\nu,\psi)$ policies do not give much away in terms of performances when items are perishable.

We now bound $R^{T2}(\mathbf{k},\mathbf{y})$ defined at (63), the task for which the new policies are designed. As in Proposition 11, the frequencies at which the new combined learning periods $s''_i(m)$ arise are also constrained in both directions.

Proposition 15 *We have the following useful inequalities:*

$$\mathcal{N}''_{t,0}(m) \leq H_{\mu,\nu} \cdot t^{\mu\vee(1-\psi)}, \tag{98}$$

for some constant $H_{\mu,\nu}$ *which is above* $G_{\mu,\nu}$*;*

$$s''_i(m) \leq \left(\frac{1}{G_{\mu,\nu}^{1/(1-\psi)}}\right) \cdot (i + I_{\mu,\nu,\psi})^{1/(1-\psi)} + 1; \tag{99}$$

$$s''_{i+1}(m) - s''_i(m) \leq \left[\left(\frac{4}{G_{\mu,\nu}}\right) \cdot (s''_i(m))^{\psi} + 1\right] \vee \left\lceil \frac{(1 + I_{\mu,\nu,\psi})^{1/(1-\psi)}}{G_{\mu,\nu}^{1/(1-\psi)}} \right\rceil. \tag{100}$$

In Proposition 15, (98), (99), and (100) correspond, respectively, to (83), (85), and (86) in Proposition 11. However, the counterpart to the earlier (84) is not needed presently. We can still decompose $R^{T2}(\mathbf{k},\mathbf{y})$ in the fashion of (88) to (91). Then, by leveraging the fact that $\tilde{V}^{k_t}_{t-1} \geq \max_{k\neq k_t} \tilde{V}^{k}_{t-1}$ for a *doing* period t which happens to be some $s''_i(m)+1$, as well as Propositions 13 and 15 in similar manners in which we used Propositions 8 to 11 in the proof of Proposition 12, we can come to the following counterpart to the latter result.

Proposition 16 *There are positive constants* $A^{Prop16}_{\mu,\nu,\psi}$, $B^{Prop16}_{\mu,\nu,\psi}$, *and* $C^{Prop16}_{\mu,\nu,\psi}$, *such that the LwD′*$(\mu,\nu,\psi)$ *policy* $(\mathbf{k},\mathbf{y})$ *would satisfy, for any* $\mathbf{f}\in(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}$,

$$R^{T2}_{\mathbf{f}}(\mathbf{k},\mathbf{y})\leq A^{Prop16}_{\mu,\nu,\psi}+B^{Prop16}_{\mu,\nu,\psi}\cdot T^{\mu\vee(1-\psi)}+C^{Prop16}_{\mu,\nu,\psi}\cdot T^{(2-\psi)\cdot(\mu\vee(1-\psi))/(2-2\psi)}\cdot(\ln T)^{5/2}.$$

When combining Propositions 14 and 16, we obtain a complete bound.

Theorem 3 *Let* $(\mathbf{k},\mathbf{y})$ *be the policy generated from following LwD′*(μ,ν,ψ). *Then, there are constants* $A^{Them3}_{\mu,\nu,\psi}$, $B^{Them3}_{\mu,\nu,\psi}$, *and* $C^{Them3}_{\mu,\nu,\psi}$, *such that* $\sup_{\mathbf{f}\in(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}}R^{T}_{\mathbf{f}}(\mathbf{y},\mathbf{k})$ *is below*

$$A^{Them3}_{\mu,\nu,\psi}+B^{Them3}_{\mu,\nu,\psi}\cdot T^{\mu\vee(1-\psi)\vee(1-3\mu/4+\psi)}+C^{Them3}_{\mu,\nu,\psi}\cdot T^{(2-\psi)\cdot(\mu\vee(1-\psi))/(2-2\psi)}\cdot(\ln T)^{5/2}.$$

Also, the choice $\mu=2/3$ *and* $\psi=1/3$ *will achieve an* $O(T^{5/6}\cdot(\ln T)^{5/2})$*-bound.*

Our analyses in Sections 6 through 8 have now culminated at Theorem 3's $T^{5/6}$-sized upper bound for the T-period regret in joint inventory-price control. It provides a performance guarantee to the LwD′(μ,ν,ψ) policies, stickier variants of the LwD(μ) ones that use profit estimates $\tilde{V}^k_{t-1}$ for pricing decisions and newsvendor-induced levels $y^*_{\tilde{f}^k_{t-1}}\wedge\bar{d}$ for ordering decisions. The need to handle items' nonperishability seems to us to have contributed the most to the difficulty of the entire undertaking.

This is more so in joint control than pure control, because now the underlying demand pattern switches from time to time. We have designed the stickier variants just to ensure that prices can keep being adopted over sufficiently long sequences of periods for the pure-control nonperishability-related bound (76) to be useful.

A slightly less thornier issue is the unboundedness of the demand support. Even though $\mathcal{F}_2(\bar{m},\bar{s})$ has restricted on the mean and standard deviation of the potential random demand under any price $\bar{p}^k$, and has thus induced the bound $\bar{m}/(1-\beta)$, as shown in (12), on the newsvendor ordering level, it has nevertheless allowed the

actual demand level to have all natural-number realizations. Major effort has been spent on the task to accommodate this justifiable generality. On the flip side, though, some of our intermediate bounds could be improved if the demand level were known beforehand to never exceed a level, say still $\bar{d}$. Let

$$\mathcal{F}_\infty(\bar{d}) \equiv \left\{ f \equiv (f(d))_{d\in\mathbb{N}} \in \mathcal{F}_0 : \sum_{d=0}^{\bar{d}} f(d) = 1 \right\}. \tag{101}$$

Clearly, $\mathcal{F}_\infty(\bar{d}) \subset \mathcal{F}_2(\bar{d},\bar{d}) \subset \mathcal{F}_1(\bar{d}) \subset \mathcal{F}_0$. When the demand-distribution vector $\mathbf{f} \equiv (f^k)_{k=1,2,\ldots,\bar{k}}$ is known to come from some $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ rather than merely some $(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}$, we would be able to improve the perishability-only bounds in both Proposition 7 for LwD(μ) and Proposition 14 for LwD$'(\mu,\nu,\psi)$ from $T^{4/5}$- to $T^{2/3}$-sized, and the combined bound in Theorem 2 when one price leads the pack by a clear margin from $T^{11/14}$- to $T^{3/4}$-sized. Interestingly, Theorem 3's ultimate $T^{5/6}$-sized bound has so far resisted improvements. Due to space limitations, we refrain from going into the details about the aforementioned performance improvements made at the expense of model generalities.

Finally, we touch lightly on the issue of lower bound, that about how fast the regret must grow when using even the best available adaptive policy. Without loss of generality, we suppose that $\bar{p}^1 < \bar{p}^2 < \cdots < \bar{p}^{\bar{k}}$. Our best effort so far has achieved a result of $\Omega(T^{1/2})$ when, either because $\bar{b}+\bar{c}-\bar{p}^{\bar{k}} \leq 0$ or because $\bar{b}+\bar{c}-\bar{p}^{\bar{k}} > 0$ but $\bar{d}$ is large enough,

$$(\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d}-1) > (\bar{b}+\bar{c}-\bar{p}^{\bar{k}}) \cdot (1-\beta). \tag{102}$$

Theorem 4 *Under (102), there is a constant A^{Them4} such that for any adaptive policy $(\mathbf{k},\mathbf{y})$,*

$$\sup_{\mathbf{f}\in(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}} R^T_{\mathbf{f}}(\mathbf{k},\mathbf{y}) \geq A^{Them4} \cdot T^{1/2}.$$

Also, this is true even if we relax the requirement (3).

This bound has not taken advantage of "more adverse" demand distributions, say those from $\mathcal{F}_2(\bar{d},\bar{d}) \setminus \mathcal{F}_\infty(\bar{d})$; moreover, it is no tighter than what Besbes and Muharremoglu [6] could achieve for pure inventory control. Nevertheless, in the hope that our current derivation can be improved upon to reach a tighter bound, we have presented details in Section 3 of Appendix 12. There remains a sizable gap between Theorem 3's $T^{5/6}$-sized upper bound and the current $T^{1/2}$-sized lower bound. Besides working to tighten Theorem 4, future research might also look beyond the techniques involving (76) for clues to improve Proposition 16 and then Theorem 3 concerning nonperishable items.

Theorems 2 and 3, as well as to a lesser degree, Propositions 6, 7, and 14 for the perishable-item special case, are complementary to the upper bounds achieved by Chen, Chao, and Ahn [14]. Dealing with continuous demands that enjoy specific relations with prices, the earlier work obtained $T^{1/2}$-sized bounds. We, on the other hand, have treated the discrete-item case where the finest tuning of ordering decisions are impossible. In addition, we have allowed the unknown demand-distribution vector $\mathbf{f}$ to come from virtually anywhere in $(\mathcal{F}_2(\bar{m},\bar{s}))^{\bar{k}}$ in case of the $T^{11/14}$-sized bounds and literally everywhere in case of the $T^{5/6}$-sized bounds. This helps us to largely shrug off the perils of model mis-specification.

The current bounds involving discrete items are also compatible with known results. When items are perishable, ours could indeed be understood as a multi-armed bandit problem with each bandit (k,y) for $k = 1,2,...,\bar{k}$ and $y = 0,1,...,\bar{d}$ earning, on average, $V_{f^k}(\bar{p}^k,y)$ as defined in (37). For this classical problem, certain upper confidence bound (UCB) policies were shown by Auer, Cesa-Bianchi, and Fischer [2] to enjoy $(\ln T)$-sized bounds. However, these bounds have coefficients that grow to $+\infty$ as the gaps $\delta V_{f^k}(\bar{p}^k,y) \equiv V^*_{\mathbf{f}} - V_{f^k}(\bar{p}^k,y)$ between non-optimal choices (k,y) and the optimal ones tend to 0^+. In the simulation study to be presented in

Section 9, we will test UCB-inspired policies along with others including our own, while these gaps have no predetermined bounds and items might be nonperishable.

9 Simulation Study

We use a simulation study to gain more insights on regret growth trends. For pure inventory control, we test both our newsvendor-based policy defined through (13) and (14) and the SA-based one proposed by Huh and Rusmevichientong [24]. Both policies are known to have regret growth rates of the $t^{1/2}$-size. Our policy appears to have an edge in terms of smaller coefficients. This is most likely attributable to its more thorough utilization of the historical demand information. Since all we need here is confirmation on the dependability of the newsvendor-based ordering, we omit presenting details of this simulation study here.

In the study concerning joint inventory-price control, we therefore use the newsvendor-based policy as a default choice when it comes to ordering. Our main purpose here is to determine where the growth rate of the optimal regret stands against the backdrops of the $T^{5/6}$-sized upper bound of Theorem 3 and the $T^{1/2}$-sized lower bound of Theorem 4. In addition, we want to identify competitive policies. Recall that each LwD(μ) could be understood as an LwD$'(\mu, 0, \psi)$ one. Besides the general LwD$'(\mu, \nu, \psi)$ policies, we also test random policies which we call rLwD(υ, ω), where υ and ω are positive constants. In every period t, a particular rLwD(υ, ω) adopts the price choice k that maximizes

$$\tilde{V}_{t-1}^{k} + \upsilon \cdot \frac{|Z|}{(\mathcal{N}_{t-1}^{k} + 1)^{\omega}}, \tag{103}$$

where $\tilde{V}_{t-1}^{k}$ as defined through (47) is an estimate about the price $\bar{p}^{k}$'s profitability based primarily on the empirical distribution under it, Z is sampled from the standard Normal distribution. This policy also strives to balance between exploration and exploitation, albeit in a fashion different than that employed by any LwD$'(\mu, \nu, \psi)$.

We also test policies that are inspired by the UCB ones for the multi-armed bandit

problem; see Auer, Cesa-Bianchi, and Fischer [2]. Depending on whether each (k, y)-combination or each k is viewed as a bandit, we consider two versions. Let $\mathcal{N}_{t-1}^{k,y}$ be the number of times that price choice k and intended order-up-to level y have been chosen by the end of period $t-1$. Also, let $\hat{V}_t^{k,y}$ be the average profit actually experienced under the (k, y)-choice:

$$\hat{V}_{t-1}^{k,y} = \frac{\sum_{s=1}^{t-1} \mathbf{1}(p_s = \bar{p}^k \text{ and } \hat{y}_s = y) \cdot v(p_s, y_s, d_s)}{\mathcal{N}_{t-1}^{k,y}}, \tag{104}$$

where each y_s may be greater than the intended $\hat{y}_s$ due to items' nonperishability, each d_s is the period-s realized demand, and $v(p_s, y_s, d_s)$ is defined at (36). Our UCB1 policy just chooses in each period t the (k, y)-pair that maximizes

$$\hat{V}_{t-1}^{k,y} + \sqrt{\frac{2 \cdot \ln(t-1)}{\mathcal{N}_{t-1}^{k,y}}}. \tag{105}$$

Our UCB2 policy treats each price as an arm, all the while using the newsvendor-based policy for ordering. Recall that $\mathcal{N}_{t-1}^{k}$ is the number of times the price index k has been chosen by time $t-1$ and $\hat{V}_{t-1}^{k}$ as defined through (53) is the average profit actually experienced under the choice. In every period t, the UCB2 policy will choice k that maximizes

$$\hat{V}_{t-1}^{k} + \sqrt{\frac{2 \cdot \ln(t-1)}{\mathcal{N}_{t-1}^{k}}}. \tag{106}$$

Although our theoretical study has allowed demand-distribution vectors $\mathbf{f} \equiv (f^k)_{k=1,2,\ldots,\bar{k}}$ to come from some $(\mathcal{F}_2(\bar{m}, \bar{s}))^{\bar{k}}$, the current simulation study has to restrict the vectors to some $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ due to computers' limitations. To compare policies, we randomly generate some M number of demand-distribution vectors $\mathbf{f}$ in $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ uniformly, and at each selected vector $\mathbf{f}$, randomly generate some L number of demand-vector sample paths. Let k_t^a be the price choice and y_t^a the

order-up-to level in period t under a given policy a for a particular demand-vector sample path. Due to (37), (40), and (41), we use the following as an approximation to the policy's regret on demand-distribution vector $\mathbf{f}$ by time t:

$$r_{\mathbf{f}}^{a,t} = V_{\mathbf{f}}^* \cdot t - \text{AVG}\left\{\sum_{s=1}^{t}\left[(\bar{p}^{k_s^a} - \bar{c}) \cdot d_s - \bar{h} \cdot (y_s^a - d_s)^+ - \bar{b} \cdot (d_s - y_s^a)^+\right]\right\}, \quad (107)$$

where $V_{\mathbf{f}}^*$ is defined at (39) and AVG stands for an average over the L demand-vector paths.

For any $\alpha \in (0,1)$, we let $R_\alpha^{a,t}$ be the conditional value at risk at the α-quantile of the M regrets $r_{\mathbf{f}}^{a,t}$. When $M = 1,000$, $R_{95\%}^{a,t}$ would stand for the average of the top 50 highest $r_{\mathbf{f}}^{a,t}$ values, where each is the regret of policy a by time t, out of the 1,000 randomly generated $\mathbf{f}$'s. Also, $R_{0\%}^{a,t}$ would be the average of all the $r_{\mathbf{f}}^{a,t}$'s of $\mathbf{f}$'s sampled from all over $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$. When M and L both approach $+\infty$ and α approaches 100%, $R_\alpha^{a,t}$ will approach the worst regret of policy a over demand-distribution vectors in $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$. Since L is finite, each $r_{\mathbf{f}}^{a,t}$ is merely an approximation of the true regret at $\mathbf{f}$. Moreover, the finiteness of M means that the $\mathbf{f}$ in $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ generating the worst regret will most likely be missed. Nevertheless, the $R_\alpha^{a,t}$ values with α close to 100% will yield insights about regrets.

At this stage, we fix $M = 1,000$ and $L = 200$. Using the $R_{\alpha=99\%}^{a,t}$ values for $t = 1, 2, ..., T = 5,000$ under various combinations of the other parameters $\bar{d}$, $\bar{k}$, $(\bar{p}^k)_{k=1,2,...,\bar{k}}$, $\bar{c}$, $\bar{h}$, and $\bar{b}$, we examine various policies a where each a represents either some LwD$'(\mu, \nu, \psi)$, some rLwD(υ, ω), UCB1, or UCB2. For the LwD$'(\mu, \nu, \psi)$ policies, we have tried $\mu = 1/2$, 2/3, 4/5, $\nu = 0$, 10, 100, and $\psi = \mu/2$, $3\mu/4$; also, for the rLwD(υ, ω) policies, we have tried υ =100, 500, 1000, 2000 and ω =1/2, 2/3, 1, and 3/2. Although having offered help to our theoretical development in Section 8, we find the general LwD$'(\mu, \nu, \psi)$ policies with $\nu > 0$ do not provide substantial improvements over the corresponding LwD(μ) ones. Hence,

we can just as well revert back to the simpler LwD(μ) policies. Besides, the performance of LwD(4/5) is not particularly impressive. Among the rLwD(υ, ω) policies, on the other hand, we find the one with $(\upsilon, \omega) = (2000, 1)$ to be particularly competitive. On the other hand, UCB2 fares far worse than UCB1.

From now on, let us narrow down to four policies: LwD(1/2), LwD(2/3), rLwD(2000, 1), and UCB1. For either of the first two policies, we have also tested the variants inspired by Burnetas and Smith [13], in which the $\tilde{V}^k_{t-1}$ defined at (47) is replaced in step 1.2 by $\hat{V}^k_{t-1}$ defined at (53). However, the changes do not much improve performances.

To focus our study, we fix $\bar{d} = 20$, $\bar{k} = 2$, $\bar{p}^1 = 80$, $\bar{p}^2 = 100$, $\bar{c} = 50$, $\bar{h} = 1$, and $\bar{b} = 2$ for the time being. At various t points up to $T = 20,000$, we compare the $R^{a,t}_{\alpha=99\%}$ values among the different policies. With $M = 1,000$, each $R^{a,t}_{99\%}$ captures the average of the 10 worst regrets $r^{a,t}_{\mathbf{f}}$ as computed in (107). An α even closer to 100% would certainly produce results that reflect the true worst regret more faithfully. However, we observe that average of a sufficient number of demand-distribution vectors is needed for our regret trend to be smooth. In addition, constraints on computational resources have prevented us from pursuing even larger M values. Thus, we settle with $M = 1,000$ and $\alpha = 99\%$.

Now in Figures 1 to 3, we present results on $R^{a,t}_{99\%}$ for the four policies at various time points with, respectively, the horizontal axis being scaled at $t^{5/6}$, $t^{1/2}$, and $t^{2/3}$. In Figure 1 where $t^{5/6}$ serves as the horizontal axis, the growth rates of regrets of all policies are downward-sloping, and in Figure 2 where $t^{1/2}$ serves as the same, those rates are all upward-sloping. Meanwhile, in Figure 3 where $t^{2/3}$ serves as the horizontal axis, all regrets grow almost linearly. These suggest that the growth rate of the regret of any "reasonable policy" is in the vicinity of $T^{2/3}$. So far, after much effort, we have not found a policy whose regret growth rate can be significantly slower than the $T^{2/3}$-pace. In the long run, LwD(1/2) performs the best among the

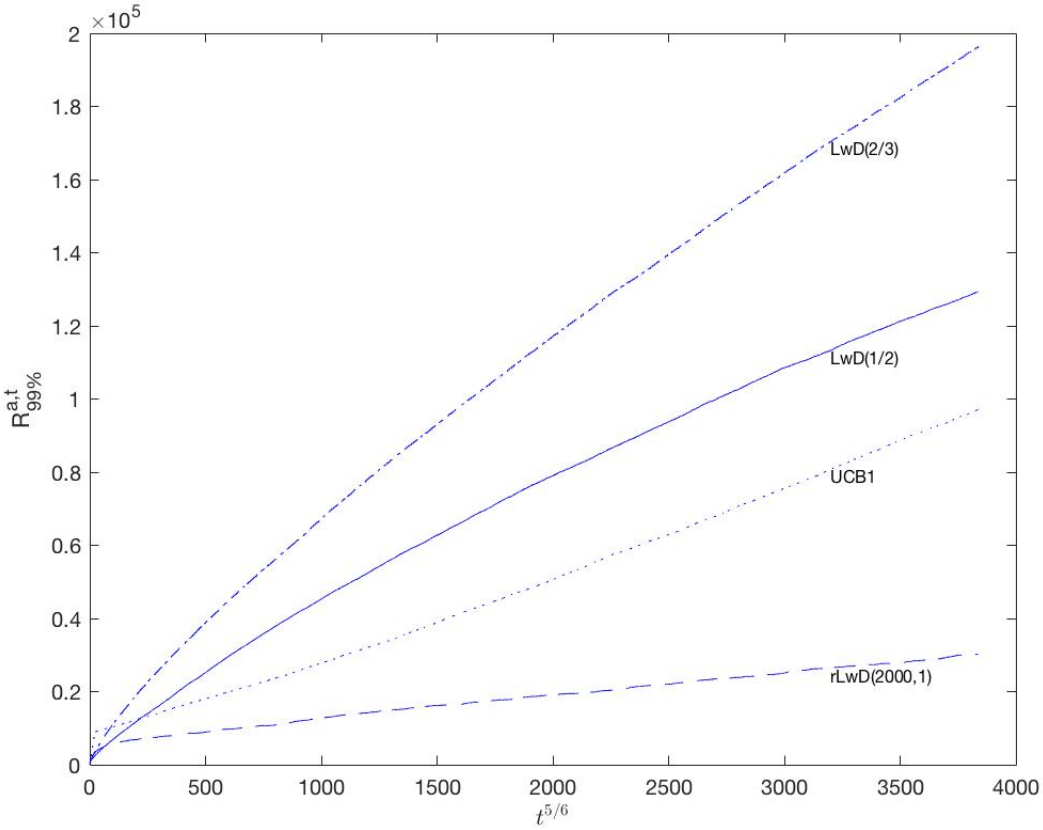

Figure 1: $R^{a,t}_{99\%}$ Values when Horizontal Scale is $t^{5/6}$

policies that we can analyze. It is slightly outperformed by UCB1 which is in turn outperformed by rLwD$(2000, 1)$.

For the most promising policy rLwD$(2000, 1)$, we choose $M = 10,000$ and continue to simulate the regrets $R^{2,t}_{99.8\%}$ (20 out of 10,000 is 0.2%) up to $t = 20,000$. We then conduct a linear regression analysis in the form of

$$\ln\left(R^{2,t}_{99.8\%}\right) = a + b \cdot \ln t, \tag{108}$$

for $t = 2,001$ to 20,000. With the R-square at 99.6% and the p-value practically zero, the least-squares estimate for the slope b turns out to be 0.662, which is very close to 2/3. Hence, neither Theorem 3 nor Theorem 4 seems to have had the final

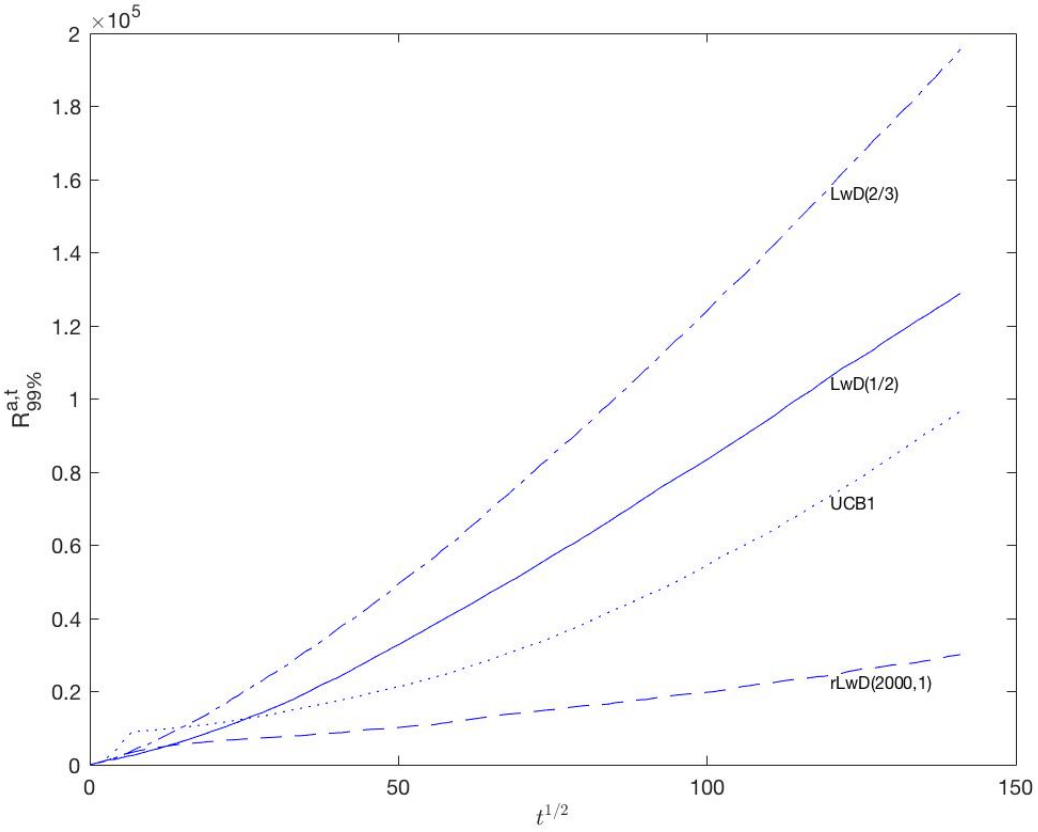

Figure 2: $R^{a,t}_{99\%}$ Values when Horizontal Scale is $t^{1/2}$

say yet. Since being worse than 99.8% of distributions is not the worst yet, the regret growth of rLwD$(2000, 1)$ could well be $T^{2/3}$-sized. On the other hand, there still might be more competitive policies. We tend to conjecture that the signature regret growth rate for joint inventory-price control is somewhere between $T^{3/5}$- and $T^{2/3}$-sized; this sets it apart from pure inventory control, whose signature rate is $T^{1/2}$-sized.

Finally, we focus on the case where T is in the hundreds rather than tens of thousands. Instead of better understandings of regret growth trends, the objective here is more about assessments of practical performances. After all, policies are most likely given tens or at the best hundreds of periods in real applications. At $T = 200$, and still $M = 2,000$, $L = 200$, and $\bar{d} = 20$, we can afford to go beyond

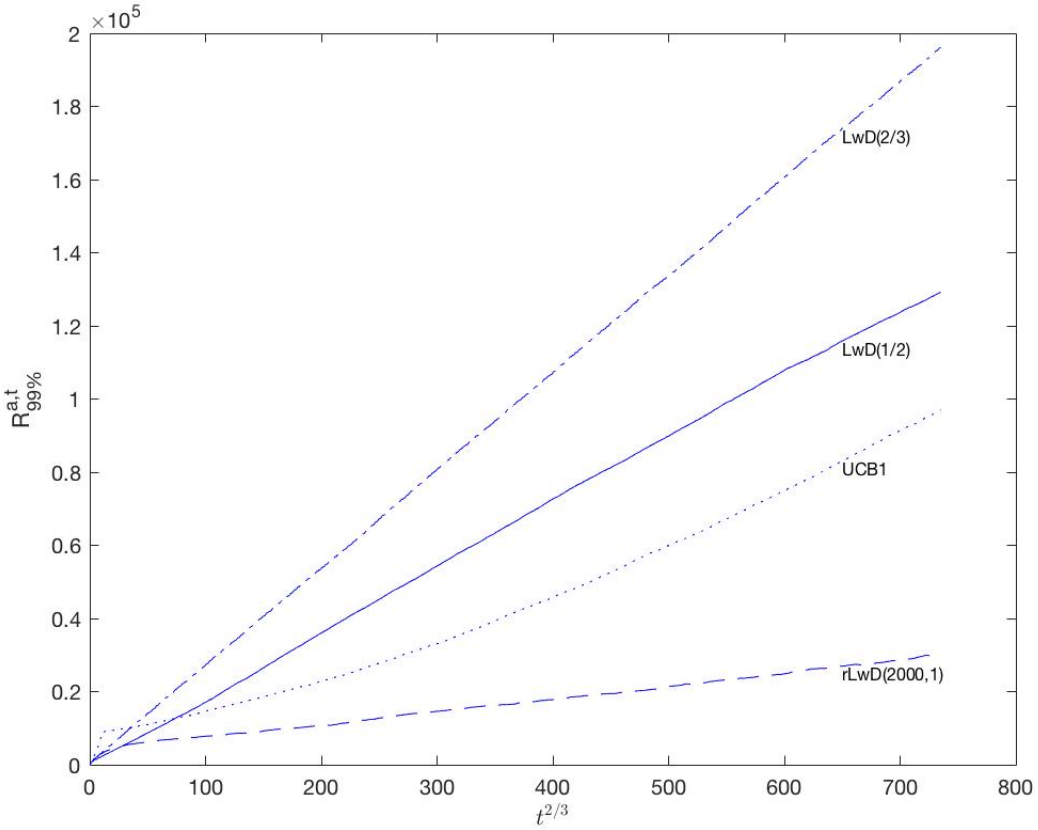

Figure 3: $R^{a,t}_{99\%}$ Values when Horizontal Scale is $t^{2/3}$

$\bar{k} \geq 2$. Let there be $\bar{k} = 5$ price choices. We now generate the $\bar{p}^k$'s uniformly from $[50, 100]$, $\bar{c}$ uniformly from $[30, \min_{k=1}^{\bar{k}} \bar{p}^k]$, $\bar{h}$ uniformly from $[0, 2]$, and $\bar{b}$ uniformly from $[0, 5]$. For $C = 100$ cases of the these randomly generated instances $((\bar{p}^k)_{k=1,2,...,\bar{k}}, \bar{c}, \bar{h}, \bar{b})$, we compute for each policy a the value $R^{a,T=200}_{\alpha=99\%}$. When the cases are ordered according to the performances of LwD(1/2) on the horizontal axis, we obtain Figure 4.

For the current case involving 5 price choices and a terminal time T that is set at the more practical value of 200, Figure 4 tells us that the analyzable policy LwD(1/2) actually stands out as the best-performing one. On the other hand, UCB1 is the least impressive here even though its large-T performance was outstanding. With each (k, y)-pair treated as an arm, it has 105 arms to try in barely 200 periods.

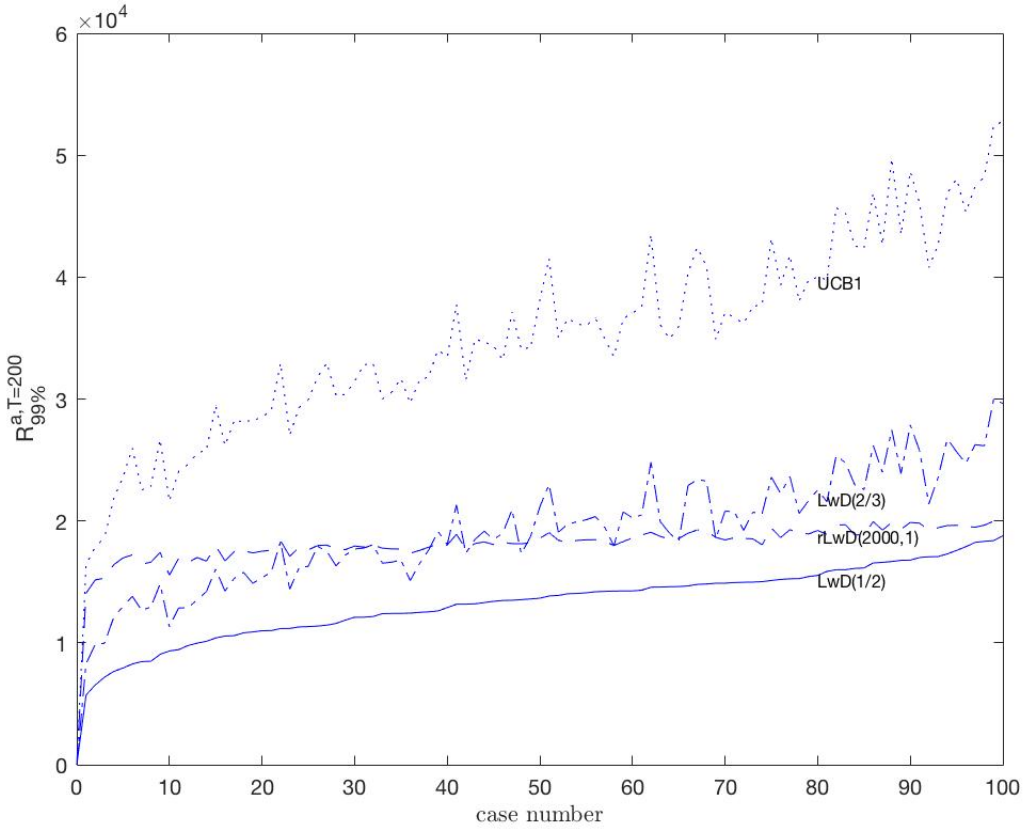

Figure 4: $R_{99\%}^{a,200}$ Values at Various Instances

There is no way that the policy can learn every arm well enough in such a short period of time. In contrast, the three other policies appear to have learned faster.

10 Concluding Remarks

We have worked on both pure inventory and joint inventory-price controls involving discrete nonperishable items and unknown demand. For the former, we contribute in the case where demand is completely unknown. For the latter joint control case, we have proposed LwD(μ) policies and their variants that build on the newsvendor-based ordering policy and a balance between learning/exploration and doing/exploitation. The nonperishability issue here is dealt with using a bound developed for pure control. Our emphasis on the case with completely thorough ambiguity in demand can help users to avoid model mis-specification. This, however, may have come at the expense of potentially less desirable bounds. Certainly, more await to be done. The newsvendor-based policy requires higher observability of historical demand levels than other policies, say the SA-based one. This has compromised its suitability for situations involving demand censoring. We speculate that the latter's advent to our setting might propel the use of the Kaplan-Meier estimators over empirical demand distributions. When pricing is involved, the gap between upper and lower bounds need to be narrowed. Furthermore, the complete removal of any relation between price and demand might have overshot. For instance, it is utterly reasonable that the demand distribution, though otherwise unknown, be decreasing in price in the stochastic sense. It will thus be interesting to know how such knowledge will help to tighten the regret bounds.

11 Appendices

Proofs of Section 4

Proof of Proposition 1: For any demand distributions $f \in \mathcal{F}_1(\bar{m})$ and $g \in \mathcal{F}_0$, as well as order-up-to level $y^0 \equiv y_g^* \wedge \bar{d}$, note that $Q_f(y^0) - Q_f(y_f^*)$ is equal to

$$[Q_f(y^0)-Q_g(y^0)]+[Q_g(y^0)-Q_g(y_g^*)]+[Q_g(y_g^*)-Q_g(y_f^*)]+[Q_g(y_f^*)-Q_f(y_f^*)]. \quad (11.109)$$

The first and fourth terms can be made small when f and g are close, the second term can be made small when y^0 and y_g^* are close, and the third term is always negative due to y_g^*'s optimality when the underlying demand distribution is g. Let us investigate how small the first and fourth terms can be. For the first and fourth terms, we have from (4) that

$$\begin{aligned}
&[Q_f(y^0) - Q_g(y^0)] + [Q_g(y_f^*) - Q_f(y_f^*)] \\
&\quad = \bar{h} \cdot \textstyle\sum_{d=0}^{y^0-1}[F_f(d) - F_g(d)] + \bar{b} \cdot \sum_{d=y^0}^{+\infty}[F_g(d) - F_f(d)] \\
&\qquad\quad +\bar{h} \cdot \textstyle\sum_{d=0}^{y_f^*-1}[F_g(d) - F_f(d)] + \bar{b} \cdot \sum_{d=y_f^*}^{+\infty}[F_f(d) - F_g(d)] \\
&\quad \leq (\bar{h} + \bar{b}) \cdot \textstyle\sum_{d=y^0\wedge y_f^*}^{y^0\vee y_f^*-1} |F_f(d) - F_g(d)| \leq (\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \max_{d=0}^{\bar{d}-1} |F_f(d) - F_g(d)|,
\end{aligned} \quad (11.110)$$

where the second inequality is due to the fact that both y^0 and y_f^* are in the range of $0, 1, ..., \bar{d}$. The sum will be small when f and g are close enough and $\bar{d}$ is small enough so that $\max_{d=0}^{\bar{d}-1} |F_f(d) - F_g(d)|$ is small. The second term will be 0 when $\bar{d}$ is large enough so that $y_g^* \leq \bar{d}$. Regardless of whether $\max_{d=0}^{\bar{d}-1} |F_f(d) - F_g(d)|$ is small, we still have

$$Q_f(y^0)-Q_f(y_f^*) = \bar{h}\cdot\sum_{d=0}^{y^0-1} F_f(d)+\bar{b}\cdot\sum_{d=y^0}^{+\infty}(1-F_f(d))-\bar{h}\cdot\sum_{d=0}^{y_f^*-1} F_f(d)-\bar{b}\cdot\sum_{d=y_f^*}^{+\infty}(1-F_f(d)), \quad (11.111)$$

which is below $\bar{h}\bar{d}+\bar{b}\bar{d}$ for a similar reason pertinent to the range of both y^0 and y_f^*. Summarizing (11.110) and (11.111), we obtain the first conclusion of the proposition.

Consider $R_f^{T1}(\mathbf{y})$ defined at (16). By (13), we have

$$R_f^{T1}(\mathbf{y}) = \sum_{t=1}^{T} \left\{ \mathbb{E}_f[Q_f(y^*_{\hat{f}_{t-1}} \wedge \bar{d})] - Q_f(y_f^*) \right\}. \tag{11.112}$$

Let ε_t be a sequence of positive constants. We then see that $R_f^{T1}(\mathbf{y})$ is below

$$\begin{aligned}
&\textstyle\sum_{t=1}^{T} \{ \mathbb{P}_f[\max_{d=0}^{\bar{d}-1} |F_f(d) - \hat{F}_{t-1}(d)| < \varepsilon_t \text{ and } y^*_{\hat{f}_{t-1}} \le \bar{d}] \times \\
&\quad \times \mathbb{E}_f[Q_f(y^*_{\hat{f}_{t-1}} \wedge \bar{d}) - Q_f(y_f^*)| \max_{d=0}^{\bar{d}-1} |F_f(d) - \hat{F}_{t-1}(d)| < \varepsilon_t \text{ and } y^*_{\hat{f}_{t-1}} \le \bar{d}] \\
&\quad + \mathbb{P}_f[\max_{d=0}^{\bar{d}-1} |F_f(d) - \hat{F}_{t-1}(d)| \ge \varepsilon_t \text{ or } y^*_{\hat{f}_{t-1}} \ge \bar{d}+1] \times \\
&\quad \times \mathbb{E}_f[Q_f(y^*_{\hat{f}_{t-1}} \wedge \bar{d}) - Q_f(y_f^*)| \max_{d=0}^{\bar{d}-1} |F_f(d) - \hat{F}_{t-1}(d)| \ge \varepsilon_t \text{ or } y^*_{\hat{f}_{t-1}} \ge \bar{d}+1] \} \\
&\le (\bar{h}\bar{d}+\bar{b}\bar{d}) \cdot \textstyle\sum_{t=1}^{T} [\varepsilon_t + 2\bar{d} \cdot \exp(-2\varepsilon_t^2 \cdot (t-1)) + 2 \cdot \exp(-(1-\beta)^2 \cdot (t-1)/2)],
\end{aligned} \tag{11.113}$$

where the inequality comes from (19), (26), (11.110), and (11.111). Suppose $\varepsilon_1 = 0$ and $\varepsilon_t = (\ln t/(t-1))^{1/2}$ for $t = 2, 3, ..., T$. Then, after plugging this into (11.113), we get

$$R_f^{T1}(\mathbf{y}) \le (\bar{h}\bar{d}+\bar{b}\bar{d}) \cdot [T_1 + (2\bar{d}+2) \cdot T_2 + 2T_3], \tag{11.114}$$

where

$$T_1 = \sum_{t=1}^{T-1} \frac{(\ln(t+1))^{1/2}}{t^{1/2}}, \qquad T_2 = \sum_{t=1}^{T} \frac{1}{t^2}, \qquad \text{and} \qquad T_3 = \sum_{t=1}^{T} \exp(-(1-\beta)^2 \cdot (t-1)/2). \tag{11.115}$$

Clearly,

$$T_1 \le (\ln T)^{1/2} \cdot \int_0^T \frac{1}{t^{1/2}} \cdot dt = 2T^{1/2} \cdot (\ln T)^{1/2}, \tag{11.116}$$

while T_2 and T_3 are both bounded by constants. ∎

Proof of Proposition 2: Let $\gamma = 1 - f(0)$. We divide the proof into two cases, with respectively, $\gamma \in [(1-\beta)/2, 1]$ and $\gamma \in [0, (1-\beta)/2)$.

Consider the first case with $\gamma \in [(1-\beta)/2, 1]$. For any positive integer τ_T, we show how $\mathbb{P}_f[S_{i+1} - S_i - 1 \geq \tau_T + 1]$ can be bounded. By the definition of the S_i's around (31),

$$\hat{y}_{S_{i+1}-1} \leq \hat{y}_{S_i} - D_{S_i} - D_{S_i+1} - \cdots - D_{S_{i+1}-2} - 1. \tag{11.117}$$

Since both $\hat{y}_{S_i}$ and $\hat{y}_{S_{i+1}-1}$ are between 0 and $\bar{d}$, the above necessitates that

$$D_{S_i} + D_{S_i+1} + \cdots + D_{S_{i+1}-2} \leq \bar{d} - 1. \tag{11.118}$$

This is only possible when there are at least $S_{i+1} - S_i - \bar{d}$ zeros among the $S_{i+1} - S_i - 1$ demand levels $D_{S_i}, D_{S_i+1}, ..., D_{S_{i+1}-2}$. When $S_{i+1} - S_i - 1 = \tau + 1 \geq \bar{d} - 1$, the latter event's chance under f with $f(0) = 1 - \gamma$ is, by the binomial formula,

$$\sum_{k=\tau-\bar{d}+2}^{\tau+1} \frac{(\tau+1)!}{k! \cdot (\tau+1-k)!} \cdot (1-\gamma)^k \cdot \gamma^{\tau+1-k} < (\tau+1)^{\bar{d}} \cdot (1-\gamma)^{\tau-\bar{d}+2} \cdot (1+\gamma+\cdots+\gamma^{\bar{d}}), \tag{11.119}$$

which is less than $(\tau+1)^{\bar{d}} \cdot (1-\gamma)^{\tau-\bar{d}+1}$. There exists $\theta_\gamma = \bar{d} - 1, \bar{d}, ...$ such that when $\tau \geq \theta_\gamma$, the aforementioned term will decrease with τ. For $\tau_T \geq \theta_\gamma$, we can thus deduce that

$$\mathbb{P}_f[S_{i+1} - S_i - 1 \geq \tau_T + 1] < (\tau_T + 1)^{\bar{d}} \cdot (1-\gamma)^{\tau_T - \bar{d}+1}. \tag{11.120}$$

The summands in (32) are bounded. Indeed, suppose $y, z = 0, 1, ..., \bar{d}$ are such that

$y \leq z$. Then, following (4),

$$Q_f(z) - Q_f(y) = (\bar{h} + \bar{b}) \cdot \sum_{d=y}^{z-1} F_f(d) - \bar{b} \cdot (z - y) < \bar{h}\bar{d}. \tag{11.121}$$

So by (29) and (11.120),

$$\begin{aligned} R_f^{T2}(\mathbf{y}) \leq \textstyle\sum_{t=3}^{T} \sum_{s=2\vee(t-\tau_T)}^{t-1} \mathbb{E}_f[\mid Q_f(\hat{y}_s - D_s - \cdots - D_{t-1}) - Q_f(\hat{y}_t) \mid \times \\ \times \mathbf{1}(\hat{y}_t \leq \hat{y}_s - D_s - \cdots - D_{t-1} - 1)] \\ + \bar{h}\bar{d} \cdot (T-2) \cdot (\tau_T + 1)^{\bar{d}} \cdot (1-\gamma)^{\tau_T - \bar{d}+1}. \end{aligned} \tag{11.122}$$

The above right-hand side can be written as

$$\sum_{\tau=1}^{\tau_T} R_f^{T2,\tau}(\mathbf{y}) + \bar{h}\bar{d} \cdot (T-2) \cdot (\tau_T + 1)^{\,\bar{d}} \cdot (1-\gamma)^{\tau_T - \bar{d}+1}, \tag{11.123}$$

where for $\tau = 1, 2, ..., \tau_T$,

$$\begin{aligned} R_f^{T2,\tau}(\mathbf{y}) = \textstyle\sum_{t=\tau+2}^{T} \mathbb{E}_f[\mid Q_f(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1}) - Q_f(\hat{y}_t) \mid \times \\ \times \mathbf{1}(\hat{y}_t \leq \hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1)]. \end{aligned} \tag{11.124}$$

By (9) and (13), we have $\hat{y}_t \leq \hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1$ only if

$$\hat{F}_{t-\tau-1}(\hat{y}_{t-\tau} - 1) < \beta \leq \hat{F}_{t-1}(\hat{y}_t) \leq \hat{F}_{t-1}(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1). \tag{11.125}$$

Also, due to the nature of the empirical distribution as illustrated in (10),

$$\hat{F}_{t-1}(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1) \leq \hat{F}_{t-1}(\hat{y}_{t-\tau} - 1) \leq \hat{F}_{t-\tau-1}(\hat{y}_{t-\tau} - 1) + \frac{\tau}{t-\tau}. \tag{11.126}$$

Therefore, $\hat{y}_t \leq \hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1$ only if

$$\beta \leq \hat{F}_{t-1}(\hat{y}_t) \leq \hat{F}_{t-1}(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1) < \beta + \frac{\tau}{t-\tau} \leq \beta + \frac{\tau_T}{t-\tau}, \tag{11.127}$$

an inequality alluded to earlier in (33). On the other hand, (4) has that, for $y \leq z$,

$$Q_f(z) - Q_f(y) = \bar{h} \cdot \sum_{d=y}^{z-1} F_f(d) - \bar{b} \cdot \sum_{d=y}^{z-1} (1 - F_f(d)) = (\bar{h} + \bar{b}) \cdot \sum_{d=y}^{z-1} (F_f(d) - \beta). \quad (11.128)$$

Now by (11.124), (11.127), and (11.128),

$$R_f^{T2,\tau}(\mathbf{y}) \leq (\bar{h} + \bar{b}) \cdot \sum_{t=\tau+2}^{T} \mathbb{E}_f[Z_t], \quad (11.129)$$

where

$$\begin{aligned} Z_t = & \textstyle\sum_{d=\hat{y}_t}^{\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1} \mid F_f(d) - \beta \mid \times \\ & \times \mathbf{1}(\beta \leq \hat{F}_{t-1}(\hat{y}_t) \leq \hat{F}_{t-1}(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1) < \beta + \tau_T/(t-\tau)). \end{aligned} \quad (11.130)$$

Due partially to the limited ranges of $\hat{y}_t$ and $\hat{y}_{t-\tau}$,

$$Z_t \leq \bar{d}; \quad (11.131)$$

in addition,

$$\begin{aligned} Z_t \leq & \textstyle\sum_{d=\hat{y}_t}^{\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1} (\mid F_f(d) - \hat{F}_{t-1}(d) \mid + \mid \hat{F}_{t-1}(d) - \beta \mid) \times \\ & \times \mathbf{1}(\beta \leq \hat{F}_{t-1}(\hat{y}_t) \leq \hat{F}_{t-1}(\hat{y}_{t-\tau} - D_{t-\tau} - \cdots - D_{t-1} - 1) < \beta + \tau_T/(t-\tau)) \\ \leq & \bar{d} \cdot [\delta_V(f, \hat{f}_{t-1}, \bar{d}) + \tau_T/(t-\tau)], \end{aligned} \quad (11.132)$$

where $\delta_V(f, \hat{f}_{t-1}, \bar{d})$ still stands for $\max_{d=0}^{\bar{d}-1} |\hat{F}_{t-1}(d) - F_f(d)|$. By (11.129), for a sequence ε_t,

$$\begin{aligned} R_f^{T2,\tau}(\mathbf{y}) \leq & (\bar{h} + \bar{b}) \cdot \textstyle\sum_{t=\tau+2}^{T} \mathbb{E}_f[Z_t | \delta_V(f, \hat{f}_{t-1}, \bar{d}) \leq \varepsilon_t] \cdot \mathbb{P}_f[\delta_V(f, \hat{f}_{t-1}, \bar{d}) \leq \varepsilon_t] \\ & + (\bar{h} + \bar{b}) \cdot \textstyle\sum_{t=\tau+2}^{T} \mathbb{E}_f[Z_t | \delta_V(f, \hat{f}_{t-1}, \bar{d}) > \varepsilon_t] \cdot \mathbb{P}_f[\delta_V(f, \hat{f}_{t-1}, \bar{d}) > \varepsilon_t], \end{aligned} \quad (11.133)$$

which by (19), (11.131), and (11.132), is less than

$$(\bar{h}\bar{d}+\bar{b}\bar{d})\cdot\sum_{t=1}^{T}\left[\varepsilon_t+2\bar{d}\cdot\exp\left(-2\varepsilon_t^2\cdot(t-1)\right)+\frac{\tau_T}{t}\right]. \tag{11.134}$$

The situation we face is very similar to (11.113) except for the τ_T/t-term. So as in Proposition 1, there are constants C'', D'', and E'' such that

$$R_f^{T2,\tau}(\mathbf{y})\leq C''+D''\cdot T^{1/2}\cdot(\ln T)^{1/2}+E''\tau_T\cdot\ln T. \tag{11.135}$$

The E''-term stems from the τ_T/t-term in (11.134). In view of (11.122) and (11.123), $R_f^{T2}(\mathbf{y})$ is below

$$C''\tau_T+D''\tau_T\cdot T^{1/2}\cdot(\ln T)^{1/2}+E''\tau_T^2\cdot\ln T+\bar{h}\bar{d}\cdot(T-2)\cdot(\tau_T+1)^{\bar{d}}\cdot(1-\gamma)^{\tau_T-\bar{d}+1}, \tag{11.136}$$

when τ_T is above the θ_γ defined right after (11.119). Otherwise, we have almost the same inequality, albeit with the last term replaced by $\bar{h}\bar{d}\cdot(T-2)$. Choose τ_T appropriately, say $\tau_T=\lfloor\ln T/\ln(1/(1-\gamma))\rfloor$. Then, as long as T is large enough, say greater than some T_γ^0, we can ensure that τ_T is above θ_γ. Very importantly, just because $\gamma\in(0,1]$, we can make sure that the last term, regardless whether τ_T is below or above θ_γ, is always bounded from above by a positive constant F_γ''. Thus, $R_f^{T2}(\mathbf{y})$ is less than

$$\frac{C''\cdot\ln T}{\ln(1/(1-\gamma))}+\frac{D''\cdot T^{1/2}\cdot(\ln T)^{3/2}}{\ln(1/(1-\gamma))}+E''\cdot\left(\frac{1}{\ln(1-\gamma)}\right)^2\cdot(\ln T)^3+F_\gamma''\cdot(\ln T)^{\bar{d}}. \tag{11.137}$$

However, as long as T is large enough, the $T^{1/2}\cdot(\ln T)^{3/2}$-sized term will dominate all other terms. A constant term can certainly cover the case when T is not that

large. Therefore, positive constants C''_γ and D''_γ exist for the intended inequality

$$R_f^{T2}(\mathbf{y}) \leq C''_\gamma + D''_\gamma \cdot T^{1/2} \cdot (\ln T)^{3/2}. \tag{11.138}$$

Since $(C''_{(1-\beta)/2}, D''_{(1-\beta)/2})$ can be used as (C''_γ, D''_γ) for cases with $\gamma \geq (1-\beta)/2 > 0$, we can have the intended bound, namely,

$$R_f^{T2}(\mathbf{y}) \leq A^{Prop2} + B^{Prop2} \cdot T^{1/2} \cdot (\ln T)^{3/2}, \tag{11.139}$$

as long as γ stays above $(1-\beta)/2$.

We now turn to the second case with $\gamma \in [0, (1-\beta)/2)$. From (32), $R_f^{T2}(\mathbf{y})$ is equal to

$$\begin{aligned}
&\textstyle\sum_{t=3}^{T} \sum_{s=2}^{t-1} \mathbb{E}_f[Q_f(\hat{y}_s - D_s - \cdots - D_{t-1}) - Q_f(\hat{y}_t) | L(t) = s] \cdot \mathbb{P}_f[L(t) = s] \\
&\leq \bar{h}\bar{d} \cdot \textstyle\sum_{t=3}^{T} \sum_{s=2}^{t-1} \mathbb{E}_f[\mathbf{1}(\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1) | L(t) = s] \cdot \mathbb{P}_f[L(t) = s] \\
&= \bar{h}\bar{d} \cdot \textstyle\sum_{t=3}^{T} \sum_{s=2}^{t-1} \mathbb{P}_f[\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1 \text{ and } L(t) = s] \\
&\leq \bar{h}\bar{d} \cdot \textstyle\sum_{t=3}^{T} \sum_{s=2}^{t-1} \mathbb{P}_f[\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1],
\end{aligned} \tag{11.140}$$

where the first inequality is due partially to

$0 \leq \hat{y}_t \leq \hat{y}_{L(t)} - D_{L(t)} - \cdots - D_{t-1} - 1 \leq \bar{d}$; see (29); also note that

$Q_f(z) - Q_f(y) \leq \bar{h}\bar{d} \cdot \mathbf{1}(z \geq y+1)$ for $0 \leq y \leq z \leq \bar{d}$; see (11.121). But

$$\mathbb{P}_f[\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1] \leq \mathbb{P}_f[\hat{y}_s \geq 1] \wedge \sum_{d=1}^{\bar{d}} \mathbb{P}_f[\hat{y}_s \geq d] \cdot \mathbb{P}_f[D_s + \cdots + D_{t-1} \leq d-1]. \tag{11.141}$$

Meanwhile, by (13) and the current range of γ,

$$\mathbb{P}_f[\hat{y}_s \geq 1] = \mathbb{P}_f\left[\hat{F}_{s-1}(0) < \beta\right] \leq \mathbb{P}_f\left[\delta_V(f, \hat{f}_{s-1}, \bar{d}) > 1 - \beta - \gamma\right], \tag{11.142}$$

which, due to (19), is below $2\bar{d} \cdot \exp(-2 \cdot (1-\beta-\gamma)^2 \cdot (s-1))$. Thus,

$$\mathbb{P}_f[\hat{y}_s \geq 1] \leq 2\bar{d} \cdot \exp\left(-2 \cdot (1-\beta-\gamma)^2 \cdot (s-1)\right). \tag{11.143}$$

Now we deal with the second term in (11.141). For $d = 1, 2, ..., \bar{d}$, let $\gamma_d = 1 - F_f(d-1)$. Our setup is such that $0 \leq \gamma_{\bar{d}} \leq \gamma_{\bar{d}-1} \leq \cdots \leq \gamma_1 = \gamma < (1-\beta)/2$. Again due to (13),

$$\mathbb{P}_f[\hat{y}_s \geq d] = \mathbb{P}_f\left[\hat{F}_{s-1}(d-1) < \beta\right] = \mathbb{P}_f\left[\sum_{\tau=1}^{s-1} \mathbf{1}(D_\tau \geq d) > (1-\beta) \cdot (s-1)\right]. \tag{11.144}$$

Note that $\mathbf{1}(D_1 \geq d), \mathbf{1}(D_2 \geq d), ..., \mathbf{1}(D_{s-1} \geq d)$ are independent Bernoulli random variables with mean γ_d, and hence $\sum_{\tau=1}^{s-1} \mathbf{1}(D_\tau \geq d)$ is a Binomial random variable with mean $\gamma_d \cdot (s-1)$. So by Markov's inequality, the rightmost term in (11.144) is below

$$\frac{\mathbb{E}_f[\sum_{\tau=1}^{s-1} \mathbf{1}(D_\tau \geq d)]}{(1-\beta) \cdot (s-1)} = \frac{\gamma_d}{1-\beta}. \tag{11.145}$$

Therefore,

$$\mathbb{P}_f[\hat{y}_s \geq d] \leq \frac{\gamma_d}{1-\beta}. \tag{11.146}$$

Also, it is easy to see that

$$\mathbb{P}_f[D_s + \cdots + D_{t-1} \leq d-1] \leq (1-\gamma_d)^{t-s}. \tag{11.147}$$

Combining (11.141), (11.143), (11.146), and (11.147), we can conclude that the term

$\sum_{s=2}^{t-1} \mathbb{P}_f[\hat{y}_s - D_s - \cdots - D_{t-1} \geq 1]$ is below

$$\sum_{s=2}^{t-1}\{[2\bar{d}\cdot\exp(-2\cdot(1-\beta-\gamma)^2\cdot(s-1))]\wedge[\sum_{d=1}^{\bar{d}}(\gamma_d/(1-\beta))\cdot(1-\gamma_d)^{t-s}]\}. \quad (11.148)$$

Consider $a(\gamma,\tau)\equiv 2\bar{d}\cdot\exp(-2\cdot(1-\beta-\gamma)^2\cdot(\tau-1))$. There exists a $t_0\geq 1$ such that for any $t\geq t_0$,

$$a\left(\frac{1-\beta}{2},t\right)=2\bar{d}\cdot\exp\left(-\frac{(1-\beta)^2}{2}\cdot(t-1)\right)<\exp\left(-\frac{(1-\beta)^2\cdot t}{4}\right). \quad (11.149)$$

Note also that $a(\gamma,s)<a((1-\beta)/2,s)$ for $\gamma\in(0,(1-\beta)/2)$. Next, consider $b(\gamma',\tau)\equiv\gamma'\cdot(1-\gamma')^\tau$. Note that

$$\frac{\partial b(\gamma',\tau)}{\partial\gamma'}=(1-\gamma')^{\tau-1}\cdot[1-(\tau+1)\cdot\gamma'], \quad (11.150)$$

and

$$\frac{\partial^2 b(\gamma',\tau)}{\partial(\gamma')^2}=-\tau\cdot(1-\gamma')^{\tau-2}\cdot[2-(\tau+1)\cdot\gamma']. \quad (11.151)$$

So the b-maximizing γ' is $\gamma^*_\tau=1/(\tau+1)$. Plugging back, we have

$$b(\gamma^*_\tau,\tau)=\frac{1}{\tau+1}\cdot(1-\frac{1}{\tau+1})^\tau=\frac{1}{\tau+1}\cdot\frac{1}{(1+1/\tau)^\tau}. \quad (11.152)$$

Note that $\lim_{\tau\to+\infty}(1+1/\tau)^\tau=e$, the natural logarithmic base which is above 2. So when τ is large enough, say greater than some t_1, the above will be below $1/(2\tau+2)$. For $T\geq 2\cdot(t_0+t_1)^2$, the upper bound in (11.148) is further bounded by a constant plus

$$\begin{aligned}\sum_{t=2\cdot(t_0+t_1)^2+1}^{T}[&\sum_{s=2}^{\lfloor t^{1/2}/2\rfloor}\sum_{d=1}^{\bar{d}}(\gamma_d/(1-\beta))\cdot(1-\gamma_d)^{t-s}\\&+\sum_{s=\lfloor t^{1/2}/2\rfloor+1}^{t-1}2\bar{d}\cdot\exp(-2\cdot(1-\beta-\gamma)^2\cdot(s-1)],\end{aligned} \quad (11.153)$$

which, according to the above from (11.149) to (11.152), is below

$$\sum_{t=2\cdot(t_0+t_1)^2+1}^{T}\left[\frac{\bar{d}}{1-\beta}\cdot\sum_{s=2}^{\lfloor t^{1/2}/2\rfloor}\frac{1}{2t-2s+2}+\sum_{s=\lfloor t^{1/2}/2\rfloor+1}^{t-1}\exp\left(-\frac{(1-\beta)^2\cdot s}{4}\right)\right]. \tag{11.154}$$

But this is smaller than

$$\sum_{t=2\cdot(t_0+t_1)^2+1}^{T}\left[\frac{\bar{d}}{2\cdot(1-\beta)\cdot t^{1/2}}+\frac{4}{(1-\beta)^2}\cdot\exp\left(-\frac{(1-\beta)^2\cdot t^{1/2}}{8}\right)\right], \tag{11.155}$$

which has a constant-plus-$T^{1/2}$ bound. So, there are positive constants E'' and F'' such that

$$R_f^{T2}(\mathbf{y})\leq E''+F''\cdot T^{1/2}, \tag{11.156}$$

for any $\gamma\in(0,(1-\beta)/2)$. Now between (11.139) and (11.156), only the former has to be used when T is made large enough. We therefore have the intended bound. ∎

Proofs of Section 6

Proof of Proposition 3: Fix some $k=1,2,...,\bar{k}$ and $t=1,2,....$ If $\mathcal{N}_{s,0}^k=\mathcal{N}_{s-1,0}^k+1$ never occurred for $s=1,2,...,t$, we can conclude that $\mathcal{N}_{t,0}^k=0$ from the policy's initialization. Otherwise, let $s=1,2,...,t$ be the latest time for the update (48) to occur. Note this must have coincided with $m_s=0$ and $k_s=\kappa_{s-1}(1)=k$. To have triggered this in the policy, it must follow that $\mathcal{N}_{s-1}^k<(s/\bar{k})^\mu$. Thus,

$$\mathcal{N}_{t,0}^k=\mathcal{N}_{s,0}^k=\mathcal{N}_{s-1,0}^k+1\leq\mathcal{N}_{s-1}^k+1<\left(\frac{s}{\bar{k}}\right)^\mu+1\leq\left(\frac{t}{\bar{k}}\right)^\mu+1. \tag{11.157}$$

For either case, we see that the desired inequality is valid. ∎

Proof of Proposition 4: We first use induction to prove that, for $t = 0, 1, ...,$

$$\mathcal{N}_t^{\kappa_t(k)} \geq \left(\frac{t+k}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., \bar{k}. \tag{11.158}$$

For $t = 0$, we have

$$\mathcal{N}_0^{\kappa_0(k)} = \mathcal{N}_0^k = 0 = \left(\frac{\bar{k}}{\bar{k}}\right)^{\mu} - 1 \geq \left(\frac{k}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., \bar{k}, \tag{11.159}$$

which is exactly (11.158) at $t = 0$. Now suppose (11.158) is true for $t - 1$. That is,

$$\mathcal{N}_{t-1}^{\kappa_{t-1}(k)} \geq \left(\frac{t-1+k}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., \bar{k}. \tag{11.160}$$

To show that (11.158) is true, we discuss whether $\mathcal{N}_{t-1}^{\kappa_{t-1}(1)} < (t/\bar{k})^{\mu}$ and hence $m_t = 0$ and $l_t = 1$, or $\mathcal{N}_{t-1}^{\kappa_{t-1}(1)} \geq (t/\bar{k})^{\mu}$ and hence $m_t = 1$. For the former case, we have

$$\mathcal{N}_t^{\kappa_{t-1}(1)} = \mathcal{N}_{t-1}^{\kappa_{t-1}(1)} + 1 \geq \left(\frac{t}{\bar{k}}\right)^{\mu} \geq \left(\frac{t+\bar{k}}{\bar{k}}\right)^{\mu} - 1, \tag{11.161}$$

where the equality comes from (48), the first inequality comes from (11.160) when applied $k = 1$, and the second inequality is due to the fact that $x^{\mu} + 1 \geq (x+1)^{\mu}$. Now,

$$\mathcal{N}_t^{\kappa_t(k)} = \mathcal{N}_t^{\kappa_{t-1}(k+1)} = \mathcal{N}_{t-1}^{\kappa_{t-1}(k+1)} \geq \left(\frac{t+k}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., j_t - 1, \tag{11.162}$$

where the first equality is due to (12.8), the second equality is due to (50), and the inequality is due to (11.160); for $k = j_t$,

$$\mathcal{N}_t^{\kappa_t(j_t)} = \mathcal{N}_t^{\kappa_{t-1}(1)} \geq \left(\frac{t+\bar{k}}{\bar{k}}\right)^{\mu} - 1 \geq \left(\frac{t+j_t}{\bar{k}}\right)^{\mu} - 1, \tag{11.163}$$

where the equality is due to the assignment that $\kappa_t(j_t) = \kappa_{t-1}(l_t) = \kappa_{t-1}(1)$ as carried out between (12.8) and (12.9), the first ineqaulity comes from (11.161), and the second inequality comes from the fact that $j_t \leq \bar{k}$; also, for $k = j_t + 1, j_t + 2, ..., \bar{k}$,

$$\mathcal{N}_t^{\kappa_t(k)} = \mathcal{N}_t^{\kappa_{t-1}(k)} \geq \mathcal{N}_t^{\kappa_{t-1}(1)} \geq \left(\frac{t+\bar{k}}{\bar{k}}\right)^{\mu} - 1 \geq \left(\frac{t+k}{\bar{k}}\right)^{\mu} - 1, \tag{11.164}$$

where the equality comes from (12.9), the first inequality is due to the definition of j_t with respect to $l_t = 1$, the second inequality comes from (11.161), and the third inequality is due to the fact that $k \leq \bar{k}$. For the latter case, for any $k = 1, 2, ..., \bar{k}$,

$$\mathcal{N}_t^k \geq \mathcal{N}_{t-1}^k \geq \mathcal{N}_{t-1}^{\kappa_{t-1}(1)} \geq \left(\frac{t}{\bar{k}}\right)^{\mu} \geq \left(\frac{t+\bar{k}}{\bar{k}}\right)^{\mu} - 1 \geq \left(\frac{t+k}{\bar{k}}\right)^{\mu} - 1, \tag{11.165}$$

where the first inequality comes from (48) to (50), the second inequality comes from (54) at $t-1$, the third inequality stems from the definition of this case with $m_t = 1$, the fourth and last inequalities come from reasons already stated. When combining (11.162) to (11.165), we can derive that κ_t satisfies (11.158).

Combining (54) and (11.158) at $t-1$, we have

$$\mathcal{N}_{t-1}^k \geq \mathcal{N}_{t-1}^{\kappa_{t-1}(1)} \geq \left(\frac{t}{\bar{k}}\right)^{\mu} - 1, \qquad \forall k = 1, 2, ..., \bar{k}, \tag{11.166}$$

which is the desired inequality. ∎

Proof of Proposition 5: By (37) and (47),

$$|\tilde{V}_{t-1}^k - V_{f^k}^k| \leq (\bar{p}^k - \bar{c}) \cdot |\mathbb{E}_{\tilde{f}_{t-1}^k}[D] - \mathbb{E}_{f^k}[D]| + |Q_{\tilde{f}_{t-1}^k}(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d}) - Q_{f^k}(y^*_{f^k})|. \tag{11.167}$$

For convenience, let us use $\tilde{F}_{t-1}^k$ for $F_{\tilde{f}_{t-1}^k}$. By $\tilde{f}_{t-1}^k$'s definition around (46), $\tilde{F}_{t-1}^k(d) = 1$ for $d = \tilde{d}_{t-1}^k, \tilde{d}_{t-1}^k + 1, ...$ Thus, for the first term on the right-hand side

of (11.167),

$$\begin{aligned}|\mathbb{E}_{\tilde{f}^k_{t-1}}[D] - \mathbb{E}_{f^k}[D]| &= |\sum_{d=0}^{+\infty}(1-\tilde{F}^k_{t-1}(d)) - \sum_{d=0}^{+\infty}(1-F_{f^k}(d))| \\ &\leq |\sum_{d=0}^{\tilde{d}^k_{t-1}-1}[F_{f^k}(d) - \tilde{F}^k_{t-1}(d)]| + \sum_{d=\tilde{d}^k_{t-1}}^{+\infty}(1-F_{f^k}(d)) \\ &\leq \delta_W(f^k, \tilde{f}^k_{t-1}, \tilde{d}^k_{t-1}) + (\bar{m}^2+\bar{s}^2)/(2\tilde{d}^k_{t-1}+1),\end{aligned} \tag{11.168}$$

where the equality is from the definition at (1), the first inequality is from the observation just made on $\tilde{F}^k_{t-1}$, and the second inequality is by both δ_W's definition at (55) and the inequality at (59). As for the second term on the right-hand side of (11.167), note it is below

$$|Q_{\tilde{f}^k_{t-1}}(y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}) - Q_{f^k}(y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d})| + |Q_{f^k}(y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}) - Q_{f^k}(y^*_{f^k})|. \tag{11.169}$$

For the first term of (11.169), note that for $y \equiv y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d} = 0, 1, ..., \bar{d}$,

$$\begin{aligned}|Q_{\tilde{f}^k_{t-1}}(y) - Q_{f^k}(y)| &= |\bar{h}\cdot\sum_{d=0}^{y-1}(\tilde{F}^k_{t-1}(d) - F_{f^k}(d)) + \bar{b}\cdot\sum_{d=y}^{+\infty}(F_{f^k}(d) - \tilde{F}^k_{t-1}(d))| \\ &= |(\bar{h}+\bar{b})\cdot\sum_{d=0}^{y-1}(\tilde{F}^k_{t-1}(d) - F_{f^k}(d)) + \bar{b}\cdot\sum_{d=0}^{+\infty}(F_{f^k}(d) - \tilde{F}^k_{t-1}(d))| \\ &\leq (\bar{h}\bar{d}+\bar{b}\bar{d})\cdot\max_{d=0}^{\bar{d}-1}|\tilde{F}^k_{t-1}(d) - F_{f^k}(d)| + \bar{b}\cdot|\sum_{d=0}^{\tilde{d}^k_{t-1}-1}(F_{f^k}(d) - \tilde{F}^k_{t-1}(d))| \\ &\qquad + \bar{b}\cdot\sum_{d=\tilde{d}^k_{t-1}}^{+\infty}(1-F_{f^k}(d)) \\ &\leq (\bar{h}\bar{d}+\bar{b}\bar{d})\cdot\delta_V(f^k, \hat{f}^k_{t-1}, \bar{d}) + \bar{b}\cdot\delta_W(f^k, \hat{f}^k_{t-1}, \tilde{d}^k_{t-1}) + \bar{b}\cdot(\bar{m}^2+\bar{s}^2)/(2\tilde{d}^k_{t-1}+1),\end{aligned} \tag{11.170}$$

where the first equality is due to (4), the second equality is through a regrouping, the first inequality relies on the limited range of y and the earlier observation about $\tilde{F}^k_{t-1}$, and the second inequality depends on the fact that $\tilde{f}^k_{t-1}$ is the same as $\hat{f}^k_{t-1}$ for d levels up to $\tilde{d}^k_{t-1}-1$, as well as the definitions of δ_V and δ_W, along with (59). For the second term of (11.169), first note (52). Thus, we can use Proposition 1 to derive that

$$|Q_{f^k}(y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}) - Q_{f^k}(y^*_{f^k})| \leq (\bar{h}\bar{d}+\bar{b}\bar{d})\cdot\left[\delta_V(f^k, \hat{f}^k_{t-1}, \bar{d}) + \mathbf{1}(y^*_{\hat{f}^k_{t-1}} \geq \bar{d}+1)\right]. \tag{11.171}$$

Combining (11.167) to (11.171), we obtain

$$|\tilde{V}^k_{t-1} - V^k_{f^k}| \leq (\bar{p}^k - \bar{c} + \bar{b}) \cdot \delta_W(f^k, \hat{f}^k_{t-1}, \tilde{d}^k_{t-1}) + (2\bar{h}\bar{d} + 2\bar{b}\bar{d}) \cdot \delta_V(f^k, \hat{f}^k_{t-1}, \bar{d}) + (\bar{p}^k - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}^k_{t-1} + 1) + (\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \mathbf{1}\left(y^*_{\hat{f}^k_{t-1}} \geq \bar{d} + 1\right). \tag{11.172}$$

Thus, when $(\bar{p}^k - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}^k_{t-1} + 1) \leq \varepsilon/3$, we will have $|\tilde{V}^k_{t-1} - V^k_{f^k}| \geq \varepsilon$ only if i) $(\bar{p}^k - \bar{c} + \bar{b}) \cdot \delta_W(f^k, \hat{f}^k_{t-1}, \tilde{d}^k_{t-1}) \geq \varepsilon/3$, ii) $(2\bar{h}\bar{d} + 2\bar{b}\bar{d}) \cdot \delta_V(f^k, \hat{f}^k_{t-1}, \bar{d}) \geq \varepsilon/3$, or iii) $y^*_{\hat{f}^k_{t-1}} \geq \bar{d} + 1$. Therefore, when $\tilde{d}^k_{t-1} \geq (3\bar{p}^k - 3\bar{c} + 3\bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)/(2\varepsilon)$,

$$\mathbb{P}_f\left[|\tilde{V}^k_{t-1} - V^k_{f^k}| \geq \varepsilon\right] \leq \mathbb{P}_f\left[\delta_W(f^k, \hat{f}^k_{t-1}, \tilde{d}^k_{t-1}) \geq \varepsilon/(3\bar{p}^k - 3\bar{c} + 3\bar{b})\right] + \mathbb{P}_f\left[\delta_V(f^k, \hat{f}^k_{t-1}, \bar{d}) \geq \varepsilon/(6\bar{h}\bar{d} + 6\bar{b}\bar{d})\right] + \mathbb{P}_f\left[y^*_{\hat{f}^k_{t-1}} \geq \bar{d} + 1\right]. \tag{11.173}$$

By (19), (26), and (56), this will result in

$$\mathbb{P}_f\left[|\tilde{V}^k_{t-1} - V^k_{f^k}| \geq \varepsilon\right] \leq 2 \cdot \exp\left(-A'' \cdot \varepsilon^2 \cdot \mathcal{N}^k_{t-1}/(\tilde{d}^k_{t-1})^2\right) + 2\bar{d} \cdot \exp\left(-B'' \cdot \varepsilon^2 \cdot \mathcal{N}^k_{t-1}\right) + 2 \cdot \exp\left(-(1-\beta)^2 \cdot \mathcal{N}^k_{t-1}/2\right), \tag{11.174}$$

where A'' and B'' are positive constants. Note $\mathcal{N}^k_{t-1}$ is involved because, as indicated in (46), it instead of $t-1$ is the number of times that demand under the price choice k has been observed by the end of period $t-1$. For $\tilde{d}^k_{t-1}$ defined at (45), both the lower-bounding requirement on it in front of (11.173) and $\tilde{d}^k_{t-1} = (\mathcal{N}^k_{t-1})^{1/4}$ can be simultaneously satisfied when $\mathcal{N}^k_{t-1} \geq \left((C''/\varepsilon) \vee \bar{d}\right)^4$ for some positive constant C''. As Proposition 4 states that $\mathcal{N}^k_{t-1} \geq (t/\bar{k})^\mu - 1$, this can further be guaranteed when

$$t \geq \bar{k} \cdot \left[\left(\frac{C''}{\varepsilon} \vee \bar{d}\right)^4 + 1\right]^{1/\mu}. \tag{11.175}$$

For some positive constants D'' and E'', the right-hand side of (11.175) is less than

$D'' + E''/\varepsilon^{4/\mu}$. In any event, when (11.175) occurs, (11.174) will lead to

$$\mathbb{P}_f\left[|\tilde{V}^k_{t-1} - V^k_{f^k}| \geq \varepsilon\right] \leq 2 \cdot \exp\left(-A'' \cdot \varepsilon^2 \cdot (\mathcal{N}^k_{t-1})^{1/2}\right) + 2\bar{d} \cdot \exp\left(-B'' \cdot \varepsilon^2 \cdot \mathcal{N}^k_{t-1}\right) + 2 \cdot \exp\left(-(1-\beta)^2 \cdot \mathcal{N}^k_{t-1}/2\right). \tag{11.176}$$

By Proposition 4, this is further below

$$2 \cdot \exp\left(-A'' \cdot \varepsilon^2 \cdot ((t/\bar{k})^\mu - 1)^{1/2}\right) + 2\bar{d} \cdot \exp\left(-B'' \cdot \varepsilon^2 \cdot ((t/\bar{k})^\mu - 1)\right) + 2 \cdot \exp\left(-(1-\beta)^2 \cdot ((t/\bar{k})^\mu - 1)/2\right). \tag{11.177}$$

When t is greater than some constant F'', though, $(t/\bar{k})^\mu - 1$ will be greater than $(t/\bar{k})^\mu/2$ and $A'' \cdot t^{\mu/2}$ will be below $B'' \cdot t^\mu$. Combine this with the observation below (11.175), and we can reach our first conclusion. This upper bound is certainly below

$$A^{Prop\ 5} \cdot \exp\left(-B^{Prop5} \cdot (\varepsilon^2 \wedge 1) \cdot t^{\mu/2}\right) + 2 \cdot \exp\left(-C^{Prop5} \cdot t^\mu\right). \tag{11.178}$$

When t is greater than a constant, $B^{Prop5} \cdot (\varepsilon^2 \wedge 1) \cdot t^{\mu/2}$ will be less than $C^{Prop5} \cdot t^\mu$. Then, the first term will dominate. This will give rise to the second expression for the upper bound. ∎

Proof of Proposition 6: To bound T_1 at (66), note from Proposition 3 that

$$\mathcal{N}^k_{T,0} < \frac{1}{\bar{k}^\mu} \cdot T^\mu + 1. \tag{11.179}$$

Since $\delta V^k_{\mathbf{f}}$ is bounded from above by $\max_{k=1}^{\bar{k}} \bar{p}^k \cdot \bar{d}$, (66) will lead to (72).

To bound T_2 at (67), we make the simplifying assumption that $k^*_{\mathbf{f}} = 1$. Note that

$$\mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \tilde{V}^k_{t-1}\right] \leq \mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \frac{V^1_{f^1} + V^k_{f^k}}{2}\right] + \mathbb{P}_{\mathbf{f}}\left[\frac{V^1_{f^1} + V^k_{f^k}}{2} \leq \tilde{V}^k_{t-1}\right]. \tag{11.180}$$

But by the definition of $\delta V_{\mathbf{f}}^k$ in (64),

$$\mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^1 \leq \frac{V_{f^1}^1 + V_{f^k}^k}{2}\right] = \mathbb{P}_{\mathbf{f}}\left[V_{f^1}^1 - \tilde{V}_{t-1}^1 \geq \frac{\delta V_{\mathbf{f}}^k}{2}\right], \tag{11.181}$$

$$\mathbb{P}_{\mathbf{f}}\left[\frac{V_{f^1}^1 + V_{f^k}^k}{2} \leq \tilde{V}_{t-1}^k\right] = \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k - V_{f^k}^k \geq \frac{\delta V_{\mathbf{f}}^k}{2}\right]. \tag{11.182}$$

Now utilizing Proposition 5 while noting that any probability, especially those corresponding small t-values, is below 1, we see that $\sum_{t=1}^T \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^1 \leq \tilde{V}_{t-1}^k\right]$ is below

$$I'' + J''/(\delta V_{\mathbf{f}}^k)^{4/\mu} + K'' \cdot \sum_{t=1}^T \exp\left(-L'' \cdot (\delta V_{\mathbf{f}}^k)^2 \cdot t^{\mu/2}\right) + 2 \cdot \sum_{t=1}^T \exp\left(-M'' \cdot t^{\mu}\right), \tag{11.183}$$

for some positive constants I'', J'', K'', L'', and M''. For $a > 0$ and $b \in [1/4, 1)$, note that

$$\sum_{t=1}^T \exp(-a \cdot t^b) \leq \int_0^\infty \exp(-a \cdot t^b) \cdot dt = \frac{1}{a^{1/b} \cdot b} \cdot \int_0^\infty y^{1/b-1} \cdot \exp(-y) \cdot dy, \tag{11.184}$$

which is below some positive constant say J'' times $1/(a^{1/b} \cdot b)$. Now (11.183) will entail (73). By (69) and the fact that the $\delta V_{\mathbf{f}}^k$'s are bounded from above by $\max_{k=1}^{\bar{k}} \bar{p}^k \cdot \bar{d}$ while from below by the given $\delta > 0$, we can obtain (74). This can be translated into

$$T_2 \leq W''_\delta, \tag{11.185}$$

for some δ-related constant W''_δ that satisfies $\lim_{\delta \to 0^+} W''_\delta = +\infty$.

To bound T_3 at (68), note that $\hat{y}_t$ there is equal to $y^*_{\tilde{f}^k_{t-1}} \wedge \bar{d}$ and hence $y^*_{\hat{f}^k_{t-1}} \wedge \bar{d}$ by (51) and (52). Further due to (37) and (38),

$$V_{f^k}^k - V_{f^k}\left(\bar{p}^k, y^*_{\hat{f}^k_{t-1}} \wedge \bar{d}\right) = Q_{f^k}\left(y^*_{\hat{f}^k_{t-1}} \wedge \bar{d}\right) - Q_{f^k}\left(y^*_{f^k}\right). \tag{11.186}$$

According to (13), the pure control uses the same ordering policy. So (68) can be written as

$$T_3 = \sum_{k=1}^{\bar{k}} \mathbb{E}_{\mathbf{f}} \left[R_{f^k}^{\mathcal{N}_T^k,1}(\mathbf{y}) \right], \tag{11.187}$$

where each $R_{f^k}^{\mathcal{N}_T^k,1}(\mathbf{y})$ is as defined in (16), with $\mathcal{N}_T^k$ here replacing T there and f^k here replacing f there. Due to Proposition 1, we have (75).

Combining (65), (72), (11.185), and (75), we can obtain

$$R_{\mathbf{f}}^{T1}(\mathbf{k},\mathbf{y}) \leq A^{Prop6}{}_{\delta} + B^{Prop6} \cdot T^{\mu} + C^{Prop6} \cdot T^{1/2} \cdot (\ln T)^{1/2}, \tag{11.188}$$

for some positive constants $A^{Prop6}{}_{\delta}$, B^{Prop6}, and C^{Prop6}. When $\mu = 1/2$, the term involving B^{Prop6} is also not necessary, and the regret is at its lowest growth rate. ∎

Proof of Proposition 7: We still have (72) of Proposition 6 for bounding T_1. To bound T_2, we can still resort to (69) and Proposition 6's (73). Hence,

$$T_2 \leq \sum_{k=2}^{\bar{k}} \left\{ \left(L'' \cdot \delta V_{\mathbf{f}}^k + \frac{M''}{(\delta V_{\mathbf{f}}^k)^{\,4/\mu - 1}} \right) \bigwedge \left(\delta V_{\mathbf{f}}^k \cdot T \right) \right\}. \tag{11.189}$$

The cases where $T \leq L''$ can be covered by a constant. Suppose $T \geq L'' + 1$. Then, on the right-hand side, the maximum at each k is achieved at $\delta V_{\mathbf{f}}^k = (M''/(T - L''))^{\mu/4}$. Therefore, for some positive constants A'' and B'',

$$T_2 \leq A'' + B'' \cdot T^{1-\mu/4}. \tag{11.190}$$

On the other hand, we can still use (75) of Proposition 6 to bound T_3.

Combining (65), (72), (75), and (11.190), while noting that

$\mu \vee (1-\mu/4) \geq 4/5 > 1/2$ for $\mu \in [1/2, 1)$, we can obtain

$$R_{\mathbf{f}}^{T1}(\mathbf{k}, \mathbf{y}) \leq A^{Prop7} + B^{Prop7} \cdot T^{\mu \vee (1-\mu/4)}, \tag{11.191}$$

for some positive μ-independent constants A^{Prop7} and B^{Prop7}. The choice for $\mu \in [1/2, 1)$ that ensures the slowest guaranteed growth rate for $R_{\mathbf{f}}^{T1}(\mathbf{k}, \mathbf{y})$ is certainly 4/5. ∎

Proofs of Section 7

Proof of Proposition 8: Note that $\mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^1 \leq \max_{k=2}^{\bar{k}} \tilde{V}_{t-1}^k + \delta V_{\mathbf{f}}^*/2 | M(t) = m\right]$ is below

$$\mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^1 \leq V_{f^1}^1 - \frac{\delta V_{\mathbf{f}}^*}{4} | M(t) = m\right] + \mathbb{P}_{\mathbf{f}}\left[V_{f^1}^1 - \frac{3\delta V_{\mathbf{f}}^*}{4} \leq \max_{k=2}^{\bar{k}} \tilde{V}_{t-1}^k | M(t) = m\right], \tag{11.192}$$

which is further below

$$\mathbb{P}_{\mathbf{f}}\left[V_{f^1}^1 - \tilde{V}_{t-1}^1 \geq \frac{\delta V_{\mathbf{f}}^*}{4} | M(t) = m\right] + \sum_{k=2}^{\bar{k}} \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k \geq V_{f^1}^1 - \frac{3\delta V_{\mathbf{f}}^*}{4} | M(t) = m\right]. \tag{11.193}$$

But by the definition of $\delta V_{\mathbf{f}}^k$ in (64) and that of $\delta V_{\mathbf{f}}^*$ in (70), for $k = 2, ..., \bar{k}$,

$$\begin{aligned} &\mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k \geq V_{f^1}^1 - 3\delta V_{\mathbf{f}}^*/4 | M(t) = m\right] \leq \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k \geq V_{f^1}^1 - 3\delta V_{\mathbf{f}}^k/4 | M(t)\right] \\ &= \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k - V_{f^k}^k \geq \delta V_{\mathbf{f}}^k/4 | M(t) = m\right] \leq \mathbb{P}_{\mathbf{f}}\left[\tilde{V}_{t-1}^k - V_{f^k}^k \geq \delta V_{\mathbf{f}}^*/4 | M(t) = m\right]. \end{aligned} \tag{11.194}$$

It can be checked that both Propositions 4 and 5 are valid when the concerned probabilities are replaced with probabilities conditioned on $M(t)$ being any particular m. So by Proposition 5, the earlier probability before (11.192) will be

below

$$A'' \cdot \exp\left(-B'' \cdot (\delta V_{\mathbf{f}}^* \wedge 1)^2 \cdot t^{\mu/2}\right), \tag{11.195}$$

when t is greater than $C'' + D''/(\delta V_{\mathbf{f}}^*)^{4/\mu}$; here, A'', B'', C'', and D'' are all positive constants. ∎

Proof of Proposition 9: By (37) and (47),

$$|\tilde{V}_{t'-1}^k - \tilde{V}_{t-1}^k| \leq (\bar{p}^k - \bar{c}) \cdot |\mathbb{E}_{\tilde{f}_{t'-1}^k}[D] - \mathbb{E}_{\tilde{f}_{t-1}^k}[D]| + |Q_{\tilde{f}_{t'-1}^k}(y^*_{\tilde{f}_{t'-1}^k} \wedge \bar{d}) - Q_{\tilde{f}_{t-1}^k}(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d})|. \tag{11.196}$$

Like in the proof of Proposition 5, let us use $\tilde{F}_{s-1}^k$ for $F_{\tilde{f}_{s-1}^k}$. By $\tilde{f}_{s-1}^k$'s definition around (46), $\tilde{F}_{s-1}^k(d) = 1$ for $d = \tilde{d}_{s-1}^k, \tilde{d}_{s-1}^k + 1, ...$ Thus, for the first term on the right-hand side of (11.196),

$$\begin{aligned} |\mathbb{E}_{\tilde{f}_{t'-1}^k}[D] - \mathbb{E}_{\tilde{f}_{t-1}^k}[D]| &= |\textstyle\sum_{d=0}^{+\infty}(1 - \tilde{F}_{t'-1}^k(d)) - \sum_{d=0}^{+\infty}(1 - \tilde{F}_{t-1}^k(d))| \\ &\leq |\textstyle\sum_{d=0}^{\tilde{d}_{t-1}^k - 1}[\tilde{F}_{t'-1}^k(d) - \tilde{F}_{t-1}^k(d)]| + \sum_{d=\tilde{d}_{t-1}^k}^{\tilde{d}_{t'-1}^k}(1 - \tilde{F}_{t'-1}^k(d)) \\ &\leq \delta_W(\tilde{f}_{t'-1}^k, \tilde{f}_{t-1}^k, \tilde{d}_{t-1}^k) + (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}_{t-1}^k + 1), \end{aligned} \tag{11.197}$$

where the equality is from the definition at (1), the first inequality is from the observation just made on $\tilde{F}_{t-1}^k$, and the second inequality is by δ_W's definition and (59). Since $Q_f(\cdot)$ as defined at (4) is convex for any $f \in \mathcal{F}^0$, we must have

$$Q_{\tilde{f}_{t'-1}^k}\left(y^*_{\tilde{f}_{t'-1}^k} \wedge \bar{d}\right) \leq Q_{\tilde{f}_{t'-1}^k}\left(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d}\right), \qquad Q_{\tilde{f}_{t-1}^k}\left(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d}\right) \leq Q_{\tilde{f}_{t-1}^k}\left(y^*_{\tilde{f}_{t'-1}^k} \wedge \bar{d}\right). \tag{11.198}$$

Thus, the second term on the right-hand side of (11.196) is below

$$|Q_{\tilde{f}_{t'-1}^k}(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d}) - Q_{\tilde{f}_{t-1}^k}(y^*_{\tilde{f}_{t-1}^k} \wedge \bar{d})| \vee |Q_{\tilde{f}_{t-1}^k}(y^*_{\tilde{f}_{t'-1}^k} \wedge \bar{d}) - Q_{\tilde{f}_{t'-1}^k}(y^*_{\tilde{f}_{t'-1}^k} \wedge \bar{d})|. \tag{11.199}$$

But for any $y = 0, 1, ..., \bar{d}$, the term $|Q_{\tilde{f}^k_{t'-1}}(y) - Q_{\tilde{f}^k_{t-1}}(y)|$ is equal to

$$\begin{aligned}&|\bar{h} \cdot \textstyle\sum_{d=0}^{y-1}(\tilde{F}^k_{t'-1}(d) - \tilde{F}^k_{t-1}(d)) + \bar{b} \cdot \sum_{d=y}^{+\infty}(\tilde{F}^k_{t-1}(d) - \tilde{F}^k_{t'-1}(d))| \\ &\quad = |(\bar{h} + \bar{b}) \cdot \textstyle\sum_{d=0}^{y-1}(\tilde{F}^k_{t'-1}(d) - \tilde{F}^k_{t-1}(d)) + \bar{b} \cdot \sum_{d=0}^{+\infty}(\tilde{F}^k_{t-1}(d) - \tilde{F}^k_{t'-1}(d))|,\end{aligned} \tag{11.200}$$

where the first equality is due to (4) and the second equality is through a regrouping. But the above is further less than

$$\begin{aligned}&(\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \max_{d=0}^{\bar{d}-1} |\tilde{F}^k_{t'-1}(d) - \tilde{F}^k_{t-1}(d)| \\ &\qquad + \bar{b} \cdot |\textstyle\sum_{d=0}^{\tilde{d}^k_{t-1}-1}(\tilde{F}^k_{t-1}(d) - \tilde{F}^k_{t'-1}(d))| + \bar{b} \cdot \sum_{d=\tilde{d}^k_{t-1}}^{\tilde{d}^k_{t'-1}}(1 - \tilde{F}^k_{t'-1}(d)) \\ &\leq (\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \delta_V(\tilde{f}^k_{t'-1}, \tilde{f}^k_{t-1}, \bar{d}) + \bar{b} \cdot \delta_W(\tilde{f}^k_{t-1}, \tilde{f}^k_{t'-1}, \tilde{d}^k_{t-1}) \\ &\qquad + \bar{b} \cdot (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}^k_{t-1} + 1),\end{aligned} \tag{11.201}$$

where the first inequality relies on the limited range of y and the earlier observation about any $\tilde{F}^k_{s-1}$, and the second inequality depends on the definitions of δ_V and δ_W, as well as (59). Combining (11.196) to (11.201), we can obtain the first claim.

To prove the second claim, note that (46) entails

$$\tilde{F}^k_{t+\tau-1}(d) = \frac{\sum_{s=1}^{t+\tau-1} \mathbf{1}(p_s = \bar{p}^k \text{ and } d_s \leq d)}{\mathcal{N}^k_{t+\tau-1}}, \tag{11.202}$$

for $\tau = 0, 1, ...$ and $d = 0, 1, ..., \tilde{d}^k_{t-1} - 1$. Therefore,

$$\mathcal{N}^k_{t-1} \cdot \tilde{F}^k_{t-1}(d) = \sum_{s=1}^{t-1} \mathbf{1}(p_s = p^k \text{ and } d_s \leq d) \leq \sum_{s=1}^{t+\tau-1} \mathbf{1}(p_s = p^k \text{ and } d_s \leq d), \tag{11.203}$$

which is equal to $\mathcal{N}^k_{t+\tau-1} \cdot \tilde{F}^k_{t+\tau-1}(d)$. Also,

$$\sum_{s=1}^{t+\tau-1} \mathbf{1}(p_s = p^k \text{ and } d_s \leq d) \leq \sum_{s=1}^{t-1} \mathbf{1}(p_s = p^k \text{ and } d_s \leq d) + \tau = \mathcal{N}^k_{t-1} \cdot \tilde{F}^k_{t-1}(d) + \tau, \tag{11.204}$$

At the same time, the hypothesis that $\mathcal{N}^k_{t+\tau-1} = \mathcal{N}^k_{t-1} + \tau$ will lead to

$$\mathcal{N}^k_{t+\tau-1} \cdot \tilde{F}^k_{t+\tau-1} = (\mathcal{N}^k_{t-1} + \tau) \cdot \tilde{F}^k_{t+\tau-1} \leq \mathcal{N}^k_{t-1} \cdot \tilde{F}^k_{t+\tau-1} + \tau. \tag{11.205}$$

Combine (11.203) to (11.205), and we can obtain

$$-\frac{\tau}{\mathcal{N}^k_{t-1}} \cdot \left(1 - \tilde{F}^k_{t+\tau-1}(d)\right) \leq \tilde{F}^k_{t-1}(d) - \tilde{F}^k_{t+\tau-1}(d) \leq \frac{\tau}{\mathcal{N}^k_{t-1}} \cdot \tilde{F}^k_{t+\tau-1}(d), \tag{11.206}$$

and hence $|\tilde{F}^k_{t-1} - \tilde{F}^k_{t+\tau-1}(d)| \leq \tau/\mathcal{N}^k_{t-1}$. This has helped us achieve the second claim. ∎

Proof of Propostion 10: Due to the first claim of Proposition 9, we have

$$\begin{aligned} |\tilde{V}^1_{t+\tau-1} - \tilde{V}^1_{t-1}| \leq\ & (\bar{h}\bar{d} + \bar{b}\bar{d}) \cdot \delta_V(\tilde{f}^1_{t-1}, \tilde{f}^1_{t+\tau-1}, \bar{d}) \\ & + (\bar{p}^1 - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)/(2\tilde{d}^1_{t-1} + 1) + (\bar{p}^1 - \bar{c} + \bar{b}) \cdot \delta_W(\tilde{f}^1_{t-1}, \hat{f}^1_{t+\tau-1}, \tilde{d}^1_{t-1}), \end{aligned} \tag{11.207}$$

as long as $\mathcal{N}^1_{t+\tau-1} = \mathcal{N}^1_{t-1} + \tau$; that is, as long as $k_t = k_{t+1} = \cdots = k_{t+\tau-1} = 1$. Note it has been hypothesized that $\tilde{V}^1_{t-1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{t-1} + \delta V^*_{\mathbf{f}}/2$ and there is no interruption from learning. For $\tau = 1$, we already have $k_{t+\tau-1} = 1$ and $\mathcal{N}^1_{t+\tau-1} = \mathcal{N}^1_{t-1} + \tau$. Suppose this is true for an arbitrary τ. Then, due to the nature of the LwD(μ) policy, we will have $k_{t+\tau} = 1$ and hence $\mathcal{N}^1_{t+\tau} = \mathcal{N}^1_{t-1} + \tau + 1$ when the right-hand side of (11.207) is below $\delta V^*_{\mathbf{f}}/2$.

While keeping this process going without being interrupted by learning in the periods $t, t+1, ..., t+t'-1$, we will achieve $k_t = k_{t+1} = \cdots = k_{t+t'-1} = 1$. So the key lies in whether we can keep the right-hand side of (11.207) below $\delta V^*_{\mathbf{f}}/2$. Meanwhile, this can be maintained as long as for some positive constants A'', B'', and C'',

$$\delta_V(\tilde{f}^1_{t-1}, \tilde{f}^1_{t+\tau-1}, \bar{d}) \leq A'' \cdot \delta V^*_{\mathbf{f}}, \qquad \forall \tau = 1, ..., t', \tag{11.208}$$

$$\tilde{d}^1_{t-1} \geq B''/\delta V^*_{\mathbf{f}}, \text{ and}$$

$$\delta_W(\hat{f}^1_{t-1}, \hat{f}^1_{t+\tau-1}, d^1_{t-1}) \leq C'' \cdot \delta V^*_{\mathbf{f}}, \qquad \forall \tau = 1, ..., t'. \tag{11.209}$$

By (18), the requirement (11.208) can be achieved if

$$|\tilde{F}^1_{t-1}(d) - \tilde{F}^1_{t+\tau-1}(d)| \leq A'' \cdot \delta V^*_{\mathbf{f}}, \qquad \forall d = 0, 1, ..., \bar{d}-1,\ \tau = 1, ..., t'; \tag{11.210}$$

also, with (45), we can guarantee the requirement on $\tilde{d}^1_{t-1}$ when

$$\mathcal{N}^1_{t-1} \geq \frac{D''}{(\delta V^*_{\mathbf{f}})^4}, \tag{11.211}$$

for some constant D''; in addition, due to (55), we can obtain a guarantee for (11.209) as

$$|\sum_{d=0}^{\tilde{d}^1_{t-1}-1} [\tilde{F}^1_{t-1}(d) - \tilde{F}^1_{t+\tau-1}(d)]| \leq C'' \cdot \delta V^*_{\mathbf{f}}, \qquad \forall \tau = 1, ..., t'. \tag{11.212}$$

In periods t to $t + t' - 1$, price 1 is constantly being adopted; hence, $\mathcal{N}^1_{t+\tau-1} = \mathcal{N}^1_{t-1} + \tau$ for $\tau = 1, ..., t'$. So by the second claim of Proposition 9, the requirement (11.210) would be true if $\tau \leq A'' \cdot \delta V^*_{\mathbf{f}} \cdot \mathcal{N}^1_{t-1}$. Due to Proposition 4, this in turn can be guaranteed when $\tau \leq A'' \cdot \delta V^*_{\mathbf{f}} \cdot ((t/\bar{k})^\mu - 1)$. Due to the same proposition, (11.211) would be guaranteed by $t \geq \bar{k} \cdot (D''/(\delta V^*_{\mathbf{f}})^4 + 1)^{1/\mu}$. Meanwhile, (11.212) would be true if $\tau \leq C'' \cdot \delta V^*_{\mathbf{f}} \cdot \mathcal{N}^1_{t-1}/\tilde{d}^1_{t-1}$. By (45), this can be guaranteed by $\mathcal{N}^1_{t-1} \geq \bar{d}^4$ and $\tau \leq C'' \cdot \delta V^*_{\mathbf{f}} \cdot (\mathcal{N}^1_{t-1})^{3/4}$. In the end, all of these can be guaranteed by $t \geq A^{Prop\ 10} + B^{Prop\ 10}/(\delta V^*_{\mathbf{f}})^{4/\mu}$ and $\tau \leq C^{Prop\ 10} \cdot \delta V^*_{\mathbf{f}} \cdot \left((t/\bar{k})^\mu - 1\right)^{3/4}$ for some positive constants A^{Prop10}, B^{Prop10}, and C^{Prop10}. ∎

Proof of Proposition 11: Since $s'_i \geq (i + I_{\mu,\delta})^{1/\mu}/G_{\mu,\delta}^{1/\mu}$, (81) will lead to

$$\frac{(\mathcal{N}'_{t,0} + I_{\mu,\delta})^{1/\mu}}{G_{\mu,\delta}^{1/\mu}} \leq s'_{\mathcal{N}'_{t,0}} \leq t, \tag{11.213}$$

and hence

$$\mathcal{N}'_{t,0} \leq G_{\mu,\delta} \cdot t^{\mu} - I_{\mu,\delta}. \tag{11.214}$$

Due to (77), (82), and (11.214), as well as the facts that $I_{\mu,\delta} \geq \bar{k}$, we can ensure (83); for instance, we could let $H_{\mu,\delta} = G_{\mu,\delta} + \bar{k}^{1-\mu}$.

For $t = s''_i(m)$ at any given i, (83) would lead to

$$i = \mathcal{N}''_{t,0}(m) \leq H_{\mu,\delta} \cdot t^{\mu} = H_{\mu,\delta} \cdot (s''_i(m))^{\mu}, \tag{11.215}$$

which is just (84). We have (85) because $s''_i(m) \leq s'_i$ and $s'_i \leq (i + I_{\mu,\delta})^{1/\mu}/G_{\mu,\delta}^{1/\mu} + 1$.

For any $x \in (0, 1)$, we have from Taylor expansion that

$$(1+x)^{1/\mu} = 1 + \left(\frac{1}{\mu}\right) \cdot x + \frac{1}{2} \cdot \left(\frac{1}{\mu}\right) \cdot \left(\frac{1}{\mu} - 1\right) \cdot x^2 + \sum_{k=1}^{+\infty} (-T_{1k} + T_{2k}), \tag{11.216}$$

where for $k = 1, 2, ...,$

$$T_{1k} = \left(\frac{1}{\mu}\right) \cdot \left(\frac{1}{\mu} - 1\right) \cdot \frac{1}{(2k+1)!} \cdot \prod_{j=2}^{2k} \left(j - \frac{1}{\mu}\right) \cdot x^{2k+1}, \tag{11.217}$$

and

$$T_{2k} = \left(\frac{1}{\mu}\right) \cdot \left(\frac{1}{\mu} - 1\right) \cdot \frac{1}{(2k+2)!} \cdot \prod_{j=2}^{2k+1} \left(j - \frac{1}{\mu}\right) \cdot x^{2k+2}. \tag{11.218}$$

If $\mu = 1/2$, we have $T_{1k} = T_{2k} = 0$ throughout. Suppose $\mu \in (1/2, 1)$. Then,

$$\frac{T_{2k}}{T_{1k}} = \frac{2k + 1 - 1/\mu}{2k + 2} \cdot x < 1. \tag{11.219}$$

Either way, we can conclude from (11.216) that

$$(1+x)^{1/\mu} - 1 \leq \left(\frac{1}{\mu}\right) \cdot x + \frac{1}{2} \cdot \left(\frac{1}{\mu}\right) \cdot \left(\frac{1}{\mu} - 1\right) \cdot x^2 < \left(\frac{1}{\mu^2}\right) \cdot x \leq 4x. \tag{11.220}$$

For $y > 1$, we then have

$$(y+1)^{1/\mu} - y^{1/\mu} = y^{1/\mu} \cdot \left[\left(1 + \frac{1}{y}\right)^{1/\mu} - 1\right] < 4 \cdot y^{1/\mu - 1}. \tag{11.221}$$

By the definition that $s'_i = \lceil (i + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \rceil$, it follows that

$$s'_{i+1} - s'_i \leq \left[\frac{(i + I_{\mu,\delta} + 1)^{1/\mu}}{G_{\mu,\delta}^{1/\mu}} - \frac{(i + I_{\mu,\delta})^{1/\mu}}{G_{\mu,\delta}^{1/\mu}}\right] + 1 < \frac{4 \cdot (i + I_{\mu,\delta})^{1/\mu - 1}}{G_{\mu,\delta}^{1/\mu}} + 1. \tag{11.222}$$

Since $(i + I_{\mu,\delta})^{1/\mu} / G_{\mu,\delta}^{1/\mu} \leq s'_i$ and hence $i \leq G_{\mu,\delta} \cdot (s'_i)^{\mu} - I_{\mu,\delta}$, the above would entail

$$s'_{i+1} - s'_i \leq \left(\frac{4}{G_{\mu,\delta}}\right) \cdot (s'_i)^{1-\mu} + 1. \tag{11.223}$$

For any i with $s''_{i+1}(m) \geq s'_1 + 1$, we can identify j such that $s'_j \leq s''_i(m)$ and $s'_{j+1} \geq s''_{i+1}(m)$. So due to (11.223),

$$s''_{i+1}(m) - s''_i(m) \leq s'_{j+1} - s'_j \leq \left(\frac{4}{G_{\mu,\delta}}\right) \cdot (s'_j)^{1-\mu} + 1, \tag{11.224}$$

and hence (86) as by choice, $s'_j \leq s''_i(m)$. Otherwise, we still have $s''_{i+1}(m) - s''_i(m) \leq s'_1$. ∎

Proof of Proposition 12: Due to (83) and the fact that $r_f^1(x)$ is bounded by a

constant, (89) would lead to

$$T_1 \leq A'' \cdot \mathbb{E}_{\mathbf{f}}[\mathcal{N}''_{T,0}] \leq B''_{\mu,\delta} \cdot T^{\mu}, \tag{11.225}$$

for some constants A'' and $B''_{\mu,\delta}$. Since $B''_{\mu,\delta} \geq A'' \cdot H_{\mu,\delta}$, we know $\lim_{\delta \to 0^+} B''_{\mu,\delta} = +\infty$. In view of Proposition 8 as well as the facts that regrets are positive and each $r_f^t(x)$ is bounded by a constant times t, (92) will lead to

$$\begin{aligned}\eta_2(m) \leq C'' \cdot \textstyle\sum_{i=1}^{\mathcal{N}''_{T,0}(m)} [\mathbf{1}\left(s''_i(m) \leq A^{Prop8} + B^{Prop8}/\delta^{4/\mu}\right) \vee \\ \vee \exp\left(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot (s''_i(m)+1)^{\mu/2}\right)] \cdot \left(s''_{i+1}(m) - s''_i(m) - 1\right),\end{aligned} \tag{11.226}$$

where C'' is a positive constant. By (86), this is below

$$\begin{aligned}C'' \cdot (A^{Prop8} + B^{Prop8}/\delta^{4/\mu}) \times \\ \times (4/G_{\mu,\delta}) \cdot \left[(A^{Prop8} + B^{Prop8}/\delta^{4/\mu})^{1-\mu} \vee \lceil (1+I_{\mu,\delta})^{1/\mu}/G_{\mu,\delta}^{1/\mu} \rceil\right] + \lambda(m),\end{aligned} \tag{11.227}$$

where the term before T_3 caps the case where the $s''_i(m)$'s are not large enough and

$$\lambda(m) = C'' \cdot \sum_{i=1}^{\mathcal{N}''_{T,0}(m)} \exp\left(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot (s''_i(m)+1)^{\mu/2}\right) \cdot \left(s''_{i+1}(m) - s''_i(m) - 1\right). \tag{11.228}$$

By (83) and (86), we can see that $\lambda(m)$ is below

$$\left(\frac{4C''}{G_{\mu,\delta}}\right) \cdot \sum_{i=1}^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil} \exp\left(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot (s''_i(m)+1)^{\mu/2}\right) \cdot (s''_i(m)+1)^{1-\mu} + D''_{\mu,\delta}, \tag{11.229}$$

where $D''_{\mu,\delta}$ covers the case when the $s''_i(m)$'s are not large enough. Note $x^{1-\mu} \cdot \exp(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot x^{\mu/2})$ is decreasing in x when the latter is large enough.

So in view of (84),

$$\lambda(m) \leq \left(\frac{4C''}{G_{\mu,\delta}}\right) \cdot \sum_{i=1}^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil} \exp\left(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot (i/H_{\mu,\delta})^{1/2}\right) \cdot (i/H_{\mu,\delta})^{(1-\mu)/\mu} + E''_{\mu,\delta}, \tag{11.230}$$

where $E''_{\mu,\delta}$ is a constant that grows with $1/\delta$ which accounts for the occasion when $s''_i(m)$ is not large enough. Therefore,

$$\lambda(m) \leq \left(\frac{4C''}{G_{\mu,\delta}}\right) \cdot \int_0^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil / H_{\mu,\delta}} \exp\left(-C^{Prop8} \cdot (\delta^2 \wedge 1) \cdot x^{1/2}\right) \cdot x^{(1-\mu)/\mu} \cdot dx + F''_{\mu,\delta}, \tag{11.231}$$

where $F''_{\mu,\delta}$ is another constant that has to grow with $1/\delta$. Since the integral is bounded by $[2/(C^{Prop8} \cdot (\delta^2 \wedge 1))^{2/\mu}] \cdot \int_0^{+\infty} y^{(2-\mu)/\mu} \cdot \exp(-y) \cdot dy$, there is a constant $G''_{\mu,\delta}$ that grows with $1/\delta$, such that $\lambda(m) \leq G''_{\mu,\delta}$. Combine (11.227) with this, and we obtain

$$\eta_2(m) \leq J''_{\mu,\delta}, \tag{11.232}$$

for some constant $J''_{\mu,\delta}$ that grows with $1/\delta$.

Meanwhile, (93) will lead to

$$\begin{aligned} \zeta_2(m) \leq \sum_{i=1}^{\mathcal{N}''_{T,0}(m)} \mathbb{E}_{\mathbf{f}}[\sum_{j=1}^{N_i(m)} r_{f^{K_{i,j}(m)}}^{U_{i,j+1}(m)-U_{i,j}(m)}(X_{i,j}(m))| \\ |M(T) = m \text{ and } \tilde{V}^1_{s''_i(m)+1} > \max_{k=2}^{\bar{k}} \tilde{V}^k_{s''_i(m)+1} + \delta V^*_{\mathbf{f}}/2]. \end{aligned} \tag{11.233}$$

By the way in which $G_{\mu,\delta}$ is specified in (80) and the facts that $\mu \geq 4/7$, we would have

$$\begin{aligned} (4/G_{\mu,\delta}) \cdot (s''_i(m))^{1-\mu} &\leq (C^{Prop10} \cdot \delta/\bar{k}^{3\mu/4}) \cdot [(s''_i(m)+1)^{\mu} - \bar{k}^{\mu}]^{3/4} \\ &\leq (C^{Prop10} \cdot \delta V^*_{\mathbf{f}}/\bar{k}^{3\mu/4}) \cdot [(s''_i(m)+1)^{\mu} - \bar{k}^{\mu}]^{3/4}, \end{aligned} \tag{11.234}$$

as long as $s_i''(m)$ is large enough. When this happens, (86) will then lead to

$$s_{i+1}''(m) - s_i''(m) - 1 \leq C^{Prop10} \cdot \delta V_{\mathbf{f}}^* \cdot \left(\frac{(s_i''(m)+1)^{\mu}}{\bar{k}^{\mu}} - 1 \right)^{3/4}. \tag{11.235}$$

Due to Proposition 10, we would have $N_i(m) = 1$ and $K_{i,1}(m) = 1$ as long as $s_i''(m)$ is large enough. Just like in (11.227), the contribution by terms with small $s_i''(m)$'s to the right-hand side of (11.233) is bounded. Thus, (11.233) would lead to

$$\begin{aligned} \zeta_2(m) \leq \sum_{i=1}^{\mathcal{N}_{T,0}''(m)} \mathbb{E}_{\mathbf{f}}[r_{f^1}^{s_{i+1}''(m)-s_i''(m)-1}(X_{s_i''(m)+1})| \\ |M(T) = m \text{ and } \tilde{V}_{s_i''(m)+1}^1 > \max_{k=2}^{\bar{k}} \tilde{V}_{s_i''(m)+1}^k + \delta V_{\mathbf{f}}^*/2] + K_{\mu,\delta}'', \end{aligned} \tag{11.236}$$

where $K_{\mu,\delta}''$ is a constant that grows with $1/\delta$. But by (76), for $w(t) \equiv t^{1/2} \cdot (\ln t)^{3/2}$,

$$\zeta_2(m) \leq \sum_{i=1}^{\mathcal{N}_{T,0}''(m)} \left[A^{Prop2} + B^{Prop2} \cdot w\left(s_{i+1}''(m) - s_i''(m) - 1 \right) \right] + K_{\mu,\delta}''. \tag{11.237}$$

With the help of (83), we can then come to

$$\zeta_2(m) \leq K_{\mu,\delta}'' + A^{Prop2} H_{\mu,\delta} \cdot T^{\mu} + B^{Prop2} \cdot \gamma_2(m), \tag{11.238}$$

where $\gamma_2(m) = \sum_{i=1}^{\mathcal{N}_{T,0}''(m)} w\left(s_{i+1}''(m) - s_i''(m) - 1 \right)$. The key now lies in bounding $\gamma_2(m)$.

Due to (83), (86), and the monotonicity of $w(\cdot)$,

$$\gamma_2(m) \leq \sum_{i=1}^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil} w\left(\frac{4}{G_{\mu,\delta}} \cdot (s_i''(m))^{1-\mu} \right), \tag{11.239}$$

which by (85) leads further to

$$\gamma_2(m) \leq \sum_{i=1}^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil} w\left(\frac{4}{G_{\mu,\delta}} \cdot \left(\frac{1}{G_{\mu,\delta}^{1/\mu}} \cdot (i + I_{\mu,\delta})^{1/\mu} + 1 \right)^{1-\mu} \right). \tag{11.240}$$

When i is large, the term inside $w(\cdot)$ will be below $M''_{\mu,\delta} \cdot i^{1/\mu-1}$ for some $M''_{\mu,\delta} \geq 1$.

So

$$\gamma_2(m) \leq N''_{\mu,\delta} + \sum_{i=1}^{\lceil H_{\mu,\delta} \cdot T^{\mu} \rceil} w\left(M''_{\mu,\delta} \cdot i^{1/\mu-1}\right) \leq N''_{\mu,\delta} + \int_1^{H_{\mu,\delta} \cdot T^{\mu}+2} w\left(M''_{\mu,\delta} \cdot x^{1/\mu-1}\right) \cdot dx, \tag{11.241}$$

where $N''_{\mu,\delta}$ is a constant which safeguards against the occasion when i is not yet large enough. If we let $y = \ln(M''_{\mu,\delta} \cdot x^{1/\mu-1}) \geq 0$, the integral in (11.241) would become

$$\frac{\mu}{(1-\mu) \cdot (M''_{\mu,\delta})^{\mu/(1-\mu)}} \cdot \int_0^{\ln\left(M''_{\mu,\delta} \cdot (H_{\mu,\delta} \cdot T^{\mu}+2)^{1/\mu-1}\right)} \exp\left(\frac{(1+\mu) \cdot y}{2-2\mu}\right) \cdot y^{3/2} \cdot dy. \tag{11.242}$$

After letting z equal $(1+\mu) \cdot y/(2-2\mu)$, the above is equal to

$$O''_{\mu,\delta} \cdot \int_0^{(1+\mu) \cdot \ln\left(M''_{\mu,\delta} \cdot (H_{\mu,\delta} \cdot T^{\mu}+2)^{1/\mu-1}\right)/(2-2\mu)} \exp(z) \cdot z^{3/2} \cdot dz, \tag{11.243}$$

for some constant $O''_{\mu,\delta}$. Through integration by parts, the above is further bounded by a term proportional to

$$u\left(\frac{(1+\mu) \cdot \ln\left(M''_{\mu,\delta} \cdot (H_{\mu,\delta} \cdot T^{\mu}+2)^{1/\mu-1}\right)}{2-2\mu}\right), \tag{11.244}$$

for $u(z) \equiv \exp(z) \cdot z^{5/2}$. Thus, for large enough constants $P''_{\mu,\delta}$ and $Q''_{\mu,\delta}$, we have

$$\gamma_2(m) \leq P''_{\mu,\delta} + Q''_{\mu,\delta} \cdot T^{(1+\mu)/2} \cdot (\ln T)^{5/2}. \tag{11.245}$$

In view of (11.238) and (11.245),

$$\zeta_2(m) \leq K''_{\mu,\delta} + B^{Prop2} P''_{\mu,\delta} + A^{Prop2} H_{\mu,\delta} \cdot T^{\mu} + B^{Prop2} Q''_{\mu,\delta} \cdot T^{(1+\mu)/2} \cdot (\ln T)^{5/2}. \tag{11.246}$$

Combining (88) to (93), (11.225), (11.232), (11.246), and the facts that

$\sum_{m\in\mathcal{M}(T)}\mathbb{P}_{\mathbf{f}}[M(T)=m]=1$, $\mu\geq 4/7$, and $\theta_2(m)=\eta_2(m)+\zeta_2(m)$, we can obtain the desired result for $R_{\mathbf{f}}^{T2}(\mathbf{k},\mathbf{y})$, with $A^{Prop12}_{\ \mu,\delta}=J''_{\mu,\delta}+K''_{\mu,\delta}+B^{Prop2}P''_{\mu,\delta}$, $B^{Prop12}_{\ \mu,\delta}=B''_{\mu,\delta}+A^{Prop2}H_{\mu,\delta}$, and $C^{Prop12}_{\ \mu,\delta}=B^{Prop2}Q''_{\mu,\delta}$.

Since both $B''_{\mu,\delta}$ and $H_{\mu,\delta}$ grow to $+\infty$ when δ approaches 0^+, we know $\lim_{\delta\to 0^+}B^{Prop12}_{\ \mu,\delta}=+\infty$. From (11.244), we know it not necessary that $Q''_{\mu,\delta}>2\cdot(M''_{\mu,\delta})^{(1+\mu)/(2-2\mu)}\cdot H_{\mu,\delta}^{\ (1+\mu)/(2\mu)}$. Due to (11.240) and (11.241), it is also not necessary that $M''_{\mu,\delta}>8/G_{\mu,\delta}^{\ 1/\mu}$. Since $H_{\mu,\delta}$ can be made less than $2G_{\mu,\delta}$, these translate into it being possible for $C^{Prop12}_{\ \mu,\delta}$ to be less than a δ-independent constant times $1/G_{\mu,\delta}^{\ (1+\mu)/(2-2\mu)}$. As $\lim_{\delta\to 0^+}G_{\mu,\delta}=+\infty$, it is possible that $\lim_{\delta\to 0^+}C^{Prop12}_{\ \mu,\delta}=0$. ∎

Proofs of Section 8

Proof of Propostion 13: Without loss of generality, suppose $k_t=1$. Due to the first claim of Proposition 9,

$$|\tilde{V}^1_{t+\tau-1}-\tilde{V}^1_{t-1}|\leq(\bar{h}\bar{d}+\bar{b}\bar{d})\cdot\delta_V(\tilde{f}^1_{t-1},\tilde{f}^1_{t+\tau-1},\bar{d})$$
$$+(\bar{p}^1-\bar{c}+\bar{b})\cdot(\bar{m}^2+\bar{s}^2)/(2\tilde{d}^1_{t-1}+1)+(\bar{p}^1-\bar{c}+\bar{b})\cdot\delta_W(\tilde{f}^1_{t-1},\hat{f}^1_{t+\tau-1},\tilde{d}^1_{t-1}), \tag{11.247}$$

as long as $\mathcal{N}^1_{t+\tau-1}=\mathcal{N}^1_{t-1}+\tau$; that is, as long as $k_t=k_{t+1}=\cdots=k_{t+\tau-1}=1$. Note it has been hypothesized that $\tilde{V}^1_{t-1}\geq\max_{k=2}^{\bar{k}}\tilde{V}^k_{t-1}$ and there is no interruption from learning. For $\tau=1$, we already have $k_{t+\tau-1}=1$ and $\mathcal{N}^1_{t+\tau-1}=\mathcal{N}^1_{t-1}+\tau$. Suppose this is true for an arbitrary τ. Then, due to the nature of the sticky policy, we will have $k_{t+\tau}=1$ and hence $\mathcal{N}^1_{t+\tau}=\mathcal{N}^1_{t-1}+\tau+1$ when the right-hand side of (11.247) is below $\nu/(t+\tau-1)^{3\mu/4-\psi}$.

While keeping this process going without being interrupted by learning in the periods $t,t+1,...,t+t'-1$, we will achieve $k_t=k_{t+1}=\cdots=k_{t+t'-1}=1$. So the key lies in whether we can keep the right-hand side of (11.247) below $\nu/(t+\tau-1)^{3\mu/4-\psi}$.

Meanwhile, this can be maintained as long as for some positive constant A'',

$$\delta_V(\tilde{f}^1_{t-1}, \tilde{f}^1_{t+\tau-1}, \bar{d}) \leq A'' \cdot \frac{\nu}{t^{3\mu/4-\psi}} \leq 2A'' \cdot \frac{\nu}{(t+t'-1)^{3\mu/4-\psi}}, \qquad \forall \tau = 1, ..., t', \tag{11.248}$$

$\tilde{d}^1_{t-1} \geq 2 \cdot (\bar{p}^1 - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2) \cdot t^{3\mu/4-\psi}/\nu \geq (\bar{p}^1 - \bar{c} + \bar{b}) \cdot (t+t'-1)^{3\mu/4-\psi}/\nu$, and

$$\delta_W(\hat{f}^1_{t-1}, \hat{f}^1_{t+\tau-1}, d^1_{t-1}) \leq A'' \cdot \frac{\nu}{t^{3\mu/4-\psi}} \leq 2A'' \cdot \frac{\nu}{(t+t'-1)^{3\mu/4-\psi}}, \qquad \forall \tau = 1, ..., t'. \tag{11.249}$$

By (18), the requirement (11.248) can be achieved if

$$|\tilde{F}^1_{t-1}(d) - \tilde{F}^1_{t+\tau-1}(d)| \leq A'' \cdot \frac{\nu}{t^{3\mu/4-\psi}}, \qquad \forall d = 0, 1, ..., \bar{d}-1, \ \tau = 1, ..., t'; \tag{11.250}$$

also, with (45), we can guarantee the requirement on $\tilde{d}^1_{t-1}$ when

$$(\mathcal{N}^1_{t-1})^{1/4} \geq 2 \cdot (\bar{p}^1 - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2) \cdot \frac{t^{3\mu/4-\psi}}{\nu}; \tag{11.251}$$

in addition, by (55), we can obtain the guarantee for (11.249) as

$$|\sum_{d=0}^{\tilde{d}^1_{t-1}-1} [\tilde{F}^1_{t-1}(d) - \tilde{F}^1_{t+\tau-1}(d)]| \leq A'' \cdot \frac{\nu}{t^{3\mu/4-\psi}}, \qquad \forall \tau = 1, ..., t'. \tag{11.252}$$

By the second claim of Proposition 9, the requirement (11.250) would be true if $\tau \leq A'' \cdot \nu \cdot \mathcal{N}^1_{t-1}/t^{3\mu/4-\psi}$. Due to Proposition 4, this in turn can be guaranteed when $\tau \leq A'' \cdot \nu \cdot ((t/\bar{k})^\mu - 1)/t^{3\mu/4-\psi}$. Due to the same proposition, (11.251) would be guaranteed by our choices that $\psi \geq \mu/2$, $\nu > 2\bar{k}^{\mu/4} \cdot (\bar{p}^1 - \bar{c} + \bar{b}) \cdot (\bar{m}^2 + \bar{s}^2)$, and t be large enough. Meanwhile, (11.252) would be true if $\tau \leq A'' \cdot \nu \cdot \mathcal{N}^1_{t-1}/(\tilde{d}^1_{t-1} \cdot t^{3\mu/4-\psi})$. By (45), this can be guaranteed by $\mathcal{N}^1_{t-1} \geq \bar{d}^4$ and $\tau \leq A'' \cdot \nu \cdot (\mathcal{N}^1_{t-1})^{3/4}/t^{3\mu/4-\psi}$. Due to Proposition 4, this can in turn be guaranteed when $(t/\bar{k})^\mu - 1 \geq \bar{d}^4$ and $\tau \leq A'' \cdot \nu \cdot \left((t/\bar{k})^\mu - 1\right)^{3/4}/t^{3\mu/4-\psi}$. ∎

Proof of Proposition 14: We still have (72) of Proposition 6 for bounding T_1. To bound T_2 at (96), we suppose $k^*_{\mathbf{f}} = 1$. Note that $\mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \tilde{V}^k_{t-1} + \nu/t^{3\mu/4-\psi}\right]$ is below

$$\mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \frac{V^1_{f^1} + V^k_{f^k}}{2} + \frac{\nu}{2\ t^{3\mu/4-\psi}}\right] + \mathbb{P}_{\mathbf{f}}\left[\frac{V^1_{f^1} + V^k_{f^k}}{2} - \frac{\nu}{2\ t^{3\mu/4-\psi}} \leq \tilde{V}^k_{t-1}\right]. \quad (11.253)$$

But by the definition of $\delta V^k_{\mathbf{f}}$ in (64),

$$\mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \frac{V^1_{f^1} + V^k_{f^k}}{2} + \frac{\nu}{2\ t^{3\mu/4-\psi}}\right] = \mathbb{P}_{\mathbf{f}}\left[V^1_{f^1} - \tilde{V}^1_{t-1} \geq \frac{\delta V^k_{\mathbf{f}}}{2} - \frac{\nu}{2\ t^{3\mu/4-\psi}}\right], \quad (11.254)$$

$$\mathbb{P}_{\mathbf{f}}\left[\frac{V^1_{f^1} + V^k_{f^k}}{2} - \frac{\nu}{2\ t^{3\mu/4-\psi}} \leq \tilde{V}^k_{t-1}\right] = \mathbb{P}_{\mathbf{f}}\left[\tilde{V}^k_{t-1} - V^k_{f^k} \geq \frac{\delta V^k_{\mathbf{f}}}{2} - \frac{\nu}{2\ t^{3\mu/4-\psi}}\right]. \quad (11.255)$$

Let $t^0_{\mu,\nu,\psi} \equiv \lfloor \nu^{4/(3\mu-4\psi)}/(\delta V^k_{\mathbf{f}})^{4/(3\mu-4\psi)} \rfloor$. Note that $\delta V^k_{\mathbf{f}}/2 - \nu/(2t^{3\mu/4-\psi}) \geq 0$ will be true when $t \geq t^0_{\mu,\nu,\psi} + 1$. Thus, by (11.253) to (11.255),
$\sum_{t=1}^T \mathbb{P}_{\mathbf{f}}\left[\tilde{V}^1_{t-1} \leq \tilde{V}^k_{t-1} + \nu/t^{3\mu/4-\psi}\right]$ is below

$$\begin{aligned}\nu^{4/(3\mu-4\psi)}/(\delta V^k_{\mathbf{f}})^{4/(3\mu-4\psi)} &+ \textstyle\sum_{t=t^0_{\mu,\nu,\psi}+1}^T \mathbb{P}_{\mathbf{f}}\left[V^1_{f^1} - \tilde{V}^1_{t-1} \geq \delta V^k_{\mathbf{f}}/2 - \nu/(2t^{3\mu/4-\psi})\right] \\ &+ \textstyle\sum_{t=t^0_{\mu,\nu,\psi}+1}^T \mathbb{P}_{\mathbf{f}}\left[\tilde{V}^k_{t-1} - \tilde{V}^k_{f^k} \geq \delta V^k_{\mathbf{f}}/2 - \nu/(2t^{3\mu/4-\psi})\right].\end{aligned} \quad (11.256)$$

Let $t^1_{\mu,\nu,\psi} \equiv \lfloor (2\nu)^{4/(3\mu-4\psi)}/(\delta V^k_{\mathbf{f}})^{4/(3\mu-4\psi)} \rfloor$. We have $\delta V^k_{\mathbf{f}}/2 - \nu/(2t^{3\mu/4-\psi}) \geq \delta V^k_{\mathbf{f}}/4$ when $t \geq t^1_{\mu,\nu,\psi}$. Hence, the above is further below

$$\frac{2\cdot(2\nu)^{4/(3\mu-4\psi)}}{(\delta V^k_{\mathbf{f}})^{4/(3\mu-4\psi)}} + \sum_{t=t^1_{\mu,\nu,\psi}+1}^{T}\left\{\mathbb{P}_{\mathbf{f}}\left[V^1_{f^1} - \tilde{V}^1_{t-1} \geq \frac{\delta V^k_{\mathbf{f}}}{4}\right] + \mathbb{P}_{\mathbf{f}}\left[\tilde{V}^k_{t-1} - \tilde{V}^k_{f^k} \geq \frac{\delta V^k_{\mathbf{f}}}{4}\right]\right\}. \quad (11.257)$$

Now utilizing Proposition 5 while noting that any probability is below 1, we see that

$\sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}_{t-1}^{1} \leq \tilde{V}_{t-1}^{k} + \nu / t^{3\mu/4-\psi} \right]$ is below

$$A'' + B''/(\delta V_{\mathbf{f}}^{k})^{4/\mu} + C''_{\mu,\nu,\psi}/(\delta V_{\mathbf{f}}^{k})^{4/(3\mu-4\psi)} + D'' \cdot \sum_{t=1}^{T} \exp\left(-E'' \cdot ((\delta V_{\mathbf{f}}^{k}/4) \wedge 1)^{2} \cdot t^{\mu/2}\right), \tag{11.258}$$

for some positive constants A'', B'', $C''_{\mu,\nu,\psi}$, D'', and E''. For $a > 0$ and $b \in [1/4, 1)$, note that

$$\sum_{t=1}^{T} \exp(-a \cdot t^{b}) \leq \int_{0}^{\infty} \exp(-a \cdot t^{b}) \cdot dt = \frac{1}{a^{1/b} \cdot b} \cdot \int_{0}^{\infty} y^{1/b-1} \cdot \exp(-y) \cdot dy, \tag{11.259}$$

which is below some positive constant times $1/(a^{1/b} \cdot b)$. Now (11.258) will entail

$$\sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}_{t-1}^{1} \leq \tilde{V}_{t-1}^{k} + \frac{\nu}{t^{3\mu/4-\psi}} \right] \leq A'' + \frac{F''}{(\delta V_{\mathbf{f}}^{k} \wedge 4)^{4/\mu}} + \frac{C''_{\mu,\nu,\psi}}{(\delta V_{\mathbf{f}}^{k})^{4/(3\mu-4\psi)}}, \tag{11.260}$$

where F'' is another positive constant. Since the case where $\delta V_{\mathbf{f}}^{k} > 4$ can be covered by a constant and $4/(3\mu - 4\psi) \geq 4/\mu$ due to the fact that $\psi \geq \mu/2$, we further have

$$\sum_{t=1}^{T} \mathbb{P}_{\mathbf{f}} \left[\tilde{V}_{t-1}^{1} \leq \tilde{V}_{t-1}^{k} + \frac{\nu}{t^{3\mu/4-\psi}} \right] \leq G''_{\mu} + \frac{H''_{\mu,\nu,\psi}}{(\delta V_{\mathbf{f}}^{k})^{4/(3\mu-4\psi)}}, \tag{11.261}$$

for some positive constants G''_{μ} and $H''_{\mu,\nu,\psi}$. Now by (97), we have

$$T_2 \leq \sum_{k=2}^{\bar{k}} \left\{ \left(G''_{\mu} \cdot \delta V_{\mathbf{f}}^{k} + \frac{H''_{\mu,\nu,\psi}}{(\delta V_{\mathbf{f}}^{k})^{\,4/(3\mu-4\psi)-1}} \right) \bigwedge \left(\delta V_{\mathbf{f}}^{k} \cdot T \right) \right\}. \tag{11.262}$$

The cases where $T \leq G''_{\mu}$ can be covered by a constant. Suppose $T \geq G''_{\mu} + 1$. Then, on the right-hand side, the maximum at each k is achieved at $\delta V_{\mathbf{f}}^{k} = (H''_{\mu,\mu,\psi}/(T - G''_{\mu}))^{3\mu/4-\psi}$. Therefore, for some positive constants $I''_{\mu,\nu,\psi}$ and $J''_{\mu,\nu,\psi}$,

$$T_2 \leq I''_{\mu,\nu,\psi} + J''_{\mu,\nu,\psi} \cdot T^{1-3\mu/4+\psi}. \tag{11.263}$$

On the other hand, we can still use (75) of Proposition 6 to bound T_3. Combining (65), (72), (75), and (11.263), while noting that $\mu \vee (1 - 3\mu/4 + \psi) \geq \mu \vee (1 - \mu/4) > 4/5 \geq 1/2$ for $\mu \in [1/2, 1)$ and $\psi \in [\mu/2, 3\mu/4)$, we can obtain

$$R_{\mathbf{f}}^{T1}(\mathbf{k}, \mathbf{y}) \leq A^{Prop14}_{\mu,\nu,\psi} + B^{Prop14}_{\mu,\nu,\psi} \cdot T^{\mu \vee (1-3\mu/4+\psi)}, \tag{11.264}$$

for some positive constants $A^{Prop14}_{\mu,\nu,\psi}$ and $B^{Prop14}_{\mu,\nu,\delta}$. The parameters that ensure the slowest guaranteed growth rate for $R_{\mathbf{f}}^{T1}(\mathbf{k}, \mathbf{y})$ certainly satisfy $\mu = 4/5$ and $\psi = 2/5$. ∎

Proof of Proposition 15: Since $s'_i \geq (i + I_{\mu,\nu,\psi})^{1/(1-\psi)} / G_{\mu,\nu}^{1/(1-\psi)}$, (81) will lead to

$$\frac{(\mathcal{N}'_{t,0} + I_{\mu,\nu,\psi})^{1/(1-\psi)}}{G_{\mu,\nu}^{1/(1-\psi)}} \leq s'_{\mathcal{N}'_{t,0}} \leq t, \tag{11.265}$$

and hence

$$\mathcal{N}'_{t,0} \leq G_{\mu,\nu} \cdot t^{1-\psi} - I_{\mu,\nu,\psi}. \tag{11.266}$$

Due to (77), (82), and (11.266), as well as the fact that $I_{\mu,\nu,\psi} \geq \bar{k}$, we can ensure (98); for instance, we could let $H_{\mu,\nu} = G_{\mu,\nu} + \bar{k}^{1-\mu}$. By treating $1 - \psi$ here as μ in Proposition 11, we can obtain (99) and (100) as we did (85) and (86) in the proof of the earlier proposition. ∎

Proof of Proposition 16: Due to (98) and the fact that $r_f^1(x)$ is bounded by a constant, (89) would lead to

$$T_1 \leq A'' \cdot \mathbb{E}_{\mathbf{f}}[\mathcal{N}''_{T,0}] \leq B''_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)}, \tag{11.267}$$

for some constants A'' and $B''_{\mu,\nu}$.

By the way in which $G_{\mu,\nu}$ is specified in (94), we would have

$$\left(\frac{4}{G_{\mu,\nu}}\right)\cdot(s_i''(m))^{\psi} \leq C^{Prop13}\cdot\left(\left(\frac{s_i''(m)+1}{\bar{k}}\right)^{\mu}-1\right)^{3/4}\cdot\frac{\nu}{(s_i''(m)+1)^{3\mu/4-\psi}}, \quad (11.268)$$

as long as $s_i''(m)$ is large enough. When this happens, (100) will then lead to

$$s_{i+1}''(m)-s_i''(m)-1 \leq C^{Prop13}\cdot\left(\left(\frac{s_i''(m)+1}{\bar{k}}\right)^{\mu}-1\right)^{3/4}\cdot\frac{\nu}{(s_i''(m)+1)^{3\mu/4-\psi}}. \quad (11.269)$$

Due to Proposition 13 and the fact that $k_{s_i''(m)+1}$ achieves the maximum $\tilde{V}_{t-1}^{k}$ among $k=1,...,\bar{k}$, we now have $N_i(m)=1$ and $K_{i,1}(m)=1$. Thus, (91) would lead to

$$\theta_2(m) \leq \sum_{i=1}^{\mathcal{N}_{T,0}''(m)} \mathbb{E}_{\mathbf{f}}\left[r^{s_{i+1}''(m)-s_i''(m)-1}_{f^{k_{s_i''(m)+1}}}(X_{s_i''(m)+1})|M(T)=m\right]+K_{\mu,\nu,\psi}'', \quad (11.270)$$

where $K_{\mu,\nu,\psi}''$ is a constant. But by (76), we further have, for $w(t)=t^{1/2}\cdot(\ln t)^{3/2}$,

$$\theta_2(m) \leq \sum_{i=1}^{\mathcal{N}_{T,0}''(m)} \left[A^{Prop2}+B^{Prop2}\cdot w\left(s_{i+1}''(m)-s_i''(m)-1\right)\right]+K_{\mu,\nu,\psi}''. \quad (11.271)$$

With the help of (98), we can come to

$$\theta_2(m) \leq K_{\mu,\nu,\psi}''+A^{Prop2}H_{\mu,\nu}\cdot T^{\mu\vee(1-\psi)}+B^{Prop2}\cdot\gamma_2(m), \quad (11.272)$$

where $\gamma_2(m)=\sum_{i=1}^{\mathcal{N}_{T,0}''(m)} w\left(s_{i+1}''(m)-s_i''(m)-1\right)$.

Due to (98), (100), and the monotonicity of $w(\cdot)$,

$$\gamma_2(m) \leq \sum_{i=1}^{\lceil H_{\mu,\nu}\cdot T^{\mu\vee(1-\psi)}\rceil} w\left(\frac{4}{G_{\mu,\nu}}\cdot(s_i''(m))^{\psi}\right), \quad (11.273)$$

which by (99) leads further to

$$\gamma_2(m) \leq \sum_{i=1}^{\lceil H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} \rceil} w \left(\frac{4}{G_{\mu,\nu}} \cdot \left(\frac{1}{G_{\mu,\nu}^{1/(1-\psi)}} \cdot (i + I_{\mu,\nu,\psi})^{1/(1-\psi)} + 1 \right)^{\psi} \right). \quad (11.274)$$

When i is large, the term inside $w(\cdot)$ will be below $M''_{\mu,\nu,\psi} \cdot i^{\psi/(1-\psi)}$ for some constant $M''_{\mu,\nu,\psi} \geq 1$. Therefore, for some constant $N''_{\mu,\nu,\psi}$,

$$\begin{aligned} \gamma_2(m) &\leq N''_{\mu,\nu,\psi} + \textstyle\sum_{i=1}^{\lceil H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} \rceil} w \left(M''_{\mu,\nu,\psi} \cdot i^{\psi/(1-\psi)} \right) \\ &\leq N''_{\mu,\nu,\psi} + \textstyle\int_1^{H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} + 2} w \left(M''_{\mu,\nu,\psi} \cdot x^{\psi/(1-\psi)} \right) \cdot dx, \end{aligned} \quad (11.275)$$

If we let $y = \ln(M''_{\mu,\nu,\psi} \cdot x^{\psi/(1-\psi)}) \geq 0$, the integral in (11.275) would become

$$\frac{1-\psi}{\psi \cdot (M''_{\mu,\nu,\psi})^{1/\psi - 1}} \cdot \int_0^{\ln\left(M''_{\mu,\nu,\psi} \cdot (H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} + 2)^{\psi/(1-\psi)} \right)} \exp\left(\frac{(2-\psi) \cdot y}{2\psi} \right) \cdot y^{3/2} \cdot dy. \quad (11.276)$$

After letting z equal $(2-\psi) \cdot y/(2\psi)$, the above is equal to

$$O''_{\mu,\nu,\psi} \cdot \int_0^{(2-\psi) \cdot \ln\left(M''_{\mu,\nu,\psi} \cdot (H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} + 2)^{\psi/(1-\psi)} \right)/(2\psi)} \exp(z) \cdot z^{3/2} \cdot dz, \quad (11.277)$$

for some constant $O''_{\mu,\nu,\psi}$. Through integration by parts, the above is further bounded by a term proportional to

$$u\left(\frac{(2-\psi) \cdot \ln\left(M''_{\mu,\nu,\psi} \cdot (H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} + 2)^{\psi/(1-\psi)} \right)}{2\psi} \right), \quad (11.278)$$

for $u(z) \equiv \exp(z) \cdot z^{5/2}$. Thus, for large enough constants $P''_{\mu,\nu,\psi}$ and $Q''_{\mu,\nu,\psi}$, we have

$$\gamma_2(m) \leq P''_{\mu,\nu,\psi} + Q''_{\mu,\nu,\psi} \cdot T^{(2-\psi) \cdot (\mu \vee (1-\psi))/(2-2\psi)} \cdot (\ln T)^{5/2}. \quad (11.279)$$

In view of (11.272) and (11.279),

$$\theta_2(m) \leq K''_{\mu,\nu,\psi} + B^{Prop2} P''_{\mu,\nu,\psi} + A^{Prop2} H_{\mu,\nu} \cdot T^{\mu \vee (1-\psi)} + B^{Prop2} Q''_{\mu,\nu,\psi} \cdot T^{(2-\psi)\cdot(\mu \vee (1-\psi))/(2-2\psi)} \cdot (\ln T)^{5/2}. \tag{11.280}$$

Combining (88) to (91), (11.267), (11.280), and the fact that $\sum_{m \in \mathcal{M}(T)} \mathbb{P}_{\mathbf{f}}[M(T) = m] = 1$, we can obtain the desired result for $R_{\mathbf{f}}^{T2}(\mathbf{k}, \mathbf{y})$, with $A^{Prop16}_{\mu,\nu,\psi} = K''_{\mu,\nu,\psi} + B^{Prop2} P''_{\mu,\nu,\psi}$, $B^{Prop16}_{\mu,\nu,\psi} = B''_{\mu,\nu} + A^{Prop2} H_{\mu,\nu}$, and $C^{Prop16}_{\mu,\nu,\psi} = B^{Prop2} Q''_{\mu,\nu,\psi}$. ∎

12 Supplementary Materials

1 Discussion of the Four Cases

Let production cost be linear at a unit rate $\bar{c}$. We suppose that, at the end of the terminal period T, the firm will gain $\bar{c}x$ if there are x items left and will pay, on top of backlogging costs that may apply, $\bar{c}x$ if there are still x items owed to customers. For the backlogging case involving nonperishable items, recall that $\bar{h}$ is the holding cost rate and $\bar{b}$ the backlogging cost rate. To ensure positive production quantities $y_1 - 0$ as well as $y_t - (y_{t-1} - d_{t-1})$ for $t = 2, 3, ..., T$, we maintain that (3) be true.

Now note the identity

$$\sum_{t=1}^{T} d_t = (y_1 - 0) + \sum_{t=2}^{T}(y_t - y_{t-1} + d_{t-1}) - (y_T - d_T), \tag{12.1}$$

where the right-hand side sums up the T periods of production quantities less the final leftover or negative backlogged quantity. Thus, we can summarize the firm's total cost as

$$\bar{c} \cdot \sum_{t=1}^{T} d_t + \bar{h} \cdot \sum_{t=1}^{T}(y_t - d_t)^+ + \bar{b} \cdot \sum_{t=1}^{T}(d_t - y_t)^+. \tag{12.2}$$

Since the first term in (12.2) is not affected by the decision sequence $(y_1, y_2, ..., y_T)$, we can focus on the latter inventory-related cost term $\sum_{t=1}^{T} q(y_t, d_t)$ with $q(y, d)$ defined at (2).

Now suppose each item left over at the end of a period is worth some $\bar{s} < \bar{c}$ to the firm. However, it will no longer be available in the next period. Suppose any unit unsatisfied demand still incurs a penalty of $\bar{b}$ per period. For the terminal period T, let the firm be additionally charged $\bar{c}$ per unit owed. Here, the production quantities $y_1 - 0$ as well as $y_t + (d_{t-1} - y_{t-1})^+$ for $t = 2, 3, ..., T$ are always positive. The total

cost to the firm will be

$$\bar{c}\cdot\{(y_1-0)+\sum_{t=2}^{T}[y_t+(d_{t-1}-y_{t-1})^+]+(d_T-y_T)^+\}-\bar{s}\cdot\sum_{t=1}^{T}(y_t-d_t)^+ +\bar{b}\cdot\sum_{t=1}^{T}(d_t-y_t)^+. \tag{12.3}$$

In view of (12.1), the first term in (12.3) above will be

$$\bar{c}\cdot\sum_{t=1}^{T}d_t+\bar{c}\cdot[\sum_{t=2}^{T}(y_{t-1}-d_{t-1})^+ +(y_T-d_T)^+]=\bar{c}\cdot\sum_{t=1}^{T}d_t+\bar{c}\cdot\sum_{t=1}^{T}(y_t-d_t)^+. \tag{12.4}$$

We will get back to (12.2) when identifying $\bar{h}$ with $\bar{c}-\bar{s}$.

For the lost sales case involving nonperishable items, we can assume the same unit production cost $\bar{c}$ and holding cost rate $\bar{h}$. On the other hand, let $\bar{l}>\bar{c}$ be the cost of not satisfying a unit demand in any of the periods $1,...,T$. Again, let each item leftover at the end of period T be worth $\bar{c}$. To ensure positive production quantities y_1-0 as well as $y_t-(y_{t-1}-d_{t-1})^+$ for $t=2,3,...,T$, we effectively require (3) due to the positivity of the y_t's. Then, the total cost will be

$$\bar{c}\cdot\{(y_1-0)+\sum_{t=2}^{T}[y_t-(y_{t-1}-d_{t-1})^+]-(y_T-d_T)^+\}+\bar{h}\cdot\sum_{t=1}^{T}(y_t-d_t)^+ +\bar{l}\cdot\sum_{t=1}^{T}(d_t-y_t)^+. \tag{12.5}$$

But in view of (12.1), the above (12.5) will be the same as (12.2) when we identify $\bar{b}$ with $\bar{l}-\bar{c}$.

For the repeated-newsvendor case, i.e., the case where items are perishable and unsatisfied demands are lost, the starting inventory level in every period is 0 and each y_t will be both the ordering and order-up-to level for period t. Suppose each item left over at the end of a period will earn the firm $\bar{s}<\bar{c}$ and as in the lost sales case, each unsatisfied demand unit will cost the firm $\bar{l}>\bar{c}$. Then the total cost will

be

$$\bar{c} \cdot \sum_{t=1}^{T} y_t - \bar{s} \cdot \sum_{t=1}^{T} (y_t - d_t)^+ + \bar{l} \cdot \sum_{t=1}^{T} (d_t - y_t)^+. \tag{12.6}$$

Again, (12.6) will be the same as (12.2) when we identify $\bar{h}$ with $\bar{c} - \bar{s}$ and $\bar{b}$ with $\bar{l} - \bar{c}$.

Effectively, all cases enjoy the same cost expression (12.2), albeit with different interpretations for the strictly positive constants $\bar{h}$ and $\bar{b}$. Also, whether items are nonperishable or not depends on whether or not (3) is enforced.

2 Details on Maintaining (54)

First, it can let $\kappa_0(k) = k$ for $k = 1, 2, ..., \bar{k}$. Due to the initialization where $\mathcal{N}_0^k = 0$ for every k, (54) is true for $t = 0$. Next, suppose (54) is true for $t - 1$ at the beginning of period t. Now step 0.2 can be rewritten as the following:

0.2. also, let price choice $k_t = \kappa_{t-1}(l_t)$ with $l_t = 1$;

meanwhile, step 1.2 can be rewritten as the following:

1.2. also, let price choice $k_t = \kappa_{t-1}(l_t)$, where l_t is any member of $\text{argmax}_{l=1,2,...,\bar{k}} \tilde{V}_{t-1}^{\kappa_{t-1}(l)}$.

The policy can use the following to obtain κ_t for period t. First, let

$$\kappa_t(k) = \kappa_{t-1}(k), \qquad \forall k = 1, 2, ..., l_t - 1. \tag{12.7}$$

Then, as k traverses rightward from l_t to $\bar{k} - 1$, let

$$\kappa_t(k) = \kappa_{t-1}(k+1), \qquad \text{whenever } \mathcal{N}_t^{\kappa_{t-1}(l_t)} > \mathcal{N}_t^{\kappa_{t-1}(k+1)}. \tag{12.8}$$

Denote by j_t the smallest $k = l_t, l_t + 1, ..., \bar{k} - 1$ such that $\mathcal{N}_t^{\kappa_{t-1}(l_t)} \leq \mathcal{N}_t^{\kappa_{t-1}(k+1)}$, while making $j_t = \bar{k}$ when the inequality is unsatisfiable. Now let $\kappa_t(j_t) = \kappa_{t-1}(l_t)$,

and

$$\kappa_t(k) = \kappa_{t-1}(k), \qquad \forall k = j_t + 1, j_t + 2, ..., \bar{k}. \tag{12.9}$$

Due to (48), (49), and (50), as well as (54) at $t-1$,

$$\mathcal{N}_t^{\kappa_{t-1}(1)} \leq \mathcal{N}_t^{\kappa_{t-1}(2)} \leq \cdots \leq \mathcal{N}_t^{\kappa_{t-1}(l_t-1)}, \qquad \mathcal{N}_t^{\kappa_{t-1}(l_t+1)} \leq \mathcal{N}_t^{\kappa_{t-1}(l_t+2)} \leq \cdots \leq \mathcal{N}_t^{\bar{k}}; \tag{12.10}$$

also, $\mathcal{N}_t^{\kappa_{t-1}(l_t)} = \mathcal{N}_{t-1}^{\kappa_{t-1}(l_t)} + 1$. Hence, (12.7) to (12.9) will guarantee κ_t's satisfaction of (54).

3 Details Revolving around Theorem 4

We can construct a demand-distribution vector $\mathbf{f} \equiv (f^k)_{k=1,2,...,\bar{k}} \in (\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ such that $0 = V_{f^1}^1 = \cdots = V_{f^{\bar{k}-1}}^{\bar{k}-1} < V_{f^{\bar{k}}}^{\bar{k}}$. For two demand distributions $f_{a,T}^{\bar{k}}$ and $f_{b,T}^{\bar{k}}$ that become ever closer to $f^{\bar{k}}$ as $T \to +\infty$, consider two vectors $\mathbf{f}_{a,T} \equiv (f^1, ..., f^{\bar{k}-1}, f_{a,T}^{\bar{k}})$ and $\mathbf{f}_{b,T} \equiv (f^1, ..., f^{\bar{k}-1}, f_{b,T}^{\bar{k}})$. By exploiting the difficulty of distinguishing the two distributions, as expressed through Theorem 2.2 (iii) of Tsybakov [36], we can establish the $T^{1/2}$-sized bound.

However, at least for the case where

$$\bar{p}^1 - \bar{c} > \bar{b}, \tag{12.11}$$

there is hope that a three- rather than two-scenario treatment might help tighten up the bound further. The three scenarios $\mathbf{f}_{a,T}$, $\mathbf{f}_{b,T}$, and $\mathbf{f}_{c,T}$ can be constructed differently from the above. A policy's ability to tell apart scenarios $\mathbf{f}_{a,T}$ and $\mathbf{f}_{b,T}$ can be shown to depend on $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$, the average number of times that $\bar{p}^1$ is adopted when applying the policy to scenario $\mathbf{f}_{b,T}$. The smaller $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$ is, the tighter the regret's lower bound will be. We design $\mathbf{f}_{c,T}$ such that a large $\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ will be impossible when the regret is small. Yet, also by ensuring that $\mathbf{f}_{c,T}$ is close to $\mathbf{f}_{b,T}$,

we can potentially make $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$ small as well. Unfortunately, so far a good enough bound for $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1] - \mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ has not been found. The current $T^{1/2}$-sized bound has only utilized the crude $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1] \leq T$.

In the hope that our proof technique can be further improved, we choose to present details for the case satisfying (12.11). At each horizon length T, we consider three demand-distribution vectors $\mathbf{f}_{a,T}$, $\mathbf{f}_{b,T}$, and $\mathbf{f}_{c,T}$ from $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ that differ only under the first price choice. Effectively, let $\mathbf{f}_{a,T} = (f_{a,T}^1, f^2, f^3, ..., f^{\bar{k}})$, $\mathbf{f}_{b,T} = (f_{b,T}^1, f^2, f^3, ..., f^{\bar{k}})$, and $\mathbf{f}_{c,T} = (f_{c,T}^1, f^2, f^3, ..., f^{\bar{k}})$. For some small constants ϵ_T and η_T that go down to 0 at different speeds as $T \to +\infty$, we make sure that $\mathrm{D}_{KL}(f_{b,T}^1 || f_{a,T}^1)$ is proportional to ϵ_T^2 and $\mathrm{D}_{KL}(f_{c,T}^1 || f_{b,T}^1)$ is proportional to η_T^2. We need to consider only deterministic policies $(\mathbf{k}, \mathbf{y})$, so that each time-t price choice k_t is a function of historical demand observation $\mathbf{d}_{[1,t-1]}$ and so is each ordering decision y_t. This is because the performance of any randomized policy is the average of deterministic policies.

Under any fixed deterministic policy $(\mathbf{k}, \mathbf{y})$, we can ensure that the regret $R_{\mathbf{f}_{a,T}}^T(\mathbf{k}, \mathbf{y})$ of applying the policy to scenario $\mathbf{f}_{a,T}$ is more than a constant times of ϵ_T multiplying $\sum_{t=1}^T \mathbb{P}_{\hat{f}_{a,T}}[k_t \neq 1 \text{ or } y_t \neq 1]$, the average number of times either price choice is not 1 or ordering decision is not 1; we can be sure that the regret $R_{\mathbf{f}_{b,T}}^T(\mathbf{k}, \mathbf{y})$ of applying the policy to scenario $\mathbf{f}_{b,T}$ is more than a constant times of ϵ_T multiplying $\sum_{t=1}^T \mathbb{P}_{\hat{f}_{b,T}}[k_t = 1 \text{ and } y_t = 1]$, the average number of times both price choice is 1 and ordering decision is 1; also, we can be sure that the regret $R_{\mathbf{f}_{c,T}}^T(\mathbf{k}, \mathbf{y})$ of applying the policy to scenario $\mathbf{f}_{c,T}$ is more than a constant times of η_T multiplying $\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$, the average number of times the price choice is 1.

Since "either $k_t \neq 1$ or $y_t \neq 1$" and "both $k_t = 1$ and $y_t = 1$" are two opposite bets, the regret due to the inability to tell apart $\mathbf{f}_{a,T}$ and $\mathbf{f}_{b,T}$ can be bounded using Theorem 2.2 (iii) of Tsybakov [36]. For distributions $\hat{f}_t$ and $\hat{g}_t$ on $\mathcal{F}^t$ and function ϕ

from $\mathcal{F}^t$ to $\{0,1\}$, the theorem states that

$$\mathbb{P}_{\hat{f}_t}[\phi(\mathbf{D}_{[1t]})=0] \vee \mathbb{P}_{\hat{g}t}[\phi(\mathbf{D}_{[1t]})=1] \geq \frac{1}{4}\cdot \exp\left(-\mathrm{D}_{KL}(\hat{f}_t||\hat{g}_t)\right), \tag{12.12}$$

where $\mathrm{D}_{KL}(\hat{f}_t||\hat{g}_t)$ is the Kullback-Leibler divergence between $\hat{f}_t$ and $\hat{g}_t$, such that

$$\mathrm{D}_{KL}(\hat{f}_t||\hat{g}_t) \equiv \sum_{d_1=0}^{\bar{d}} \cdots \sum_{d_t=0}^{\bar{d}} \hat{f}_t(\mathbf{d}_{[1t]}) \cdot \ln\left(\frac{\hat{f}_t(d_{[1t]})}{\hat{g}_t(d_{[1t]})}\right). \tag{12.13}$$

In our execution, $\hat{f}_t$ will be substituted by $\hat{f}_{b,T,t-1}$, the distribution for $\mathbf{d}_{[1,t-1]}$ that results from applying $(\mathbf{k},\mathbf{y})$ to the demand-distribution vector $\mathbf{f}_{b,T}$; whereas, $\hat{g}_t$ will be substituted by $\hat{f}_{a,T,t-1}$, the distribution for $\mathbf{d}_{[1,t-1]}$ that results from applying $(\mathbf{k},\mathbf{y})$ to the demand-distribution vector $\mathbf{f}_{a,T}$. The divergence $\mathrm{D}_{KL}(\hat{f}_{b,T,t-1}||\hat{f}_{a,T,t-1})$ that appear on the right-hand side of (12.12), on the other hand, is equal to $\mathrm{D}_{KL}(f^1_{b,T}||f^1_{a,T})$ times $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}^1_{t-1}]$, the number of times that price choice 1 is taken up to time t when applying $(\mathbf{k},\mathbf{y})$ to scenario $\mathbf{f}_{b,T}$.

To see this, suppose demand-distribution vectors $\mathbf{f} \equiv (f^k)_{k=1,2,\ldots,\bar{k}}$ and $\mathbf{g} \equiv (g^k)_{k=1,2,\ldots,\bar{k}}$ in $(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ along with decision rule $(\mathbf{k},\mathbf{y})$ result with $\hat{f}_t$ and $\hat{g}_t$ on $\mathcal{F}^t$; i.e.,

$$\begin{cases} \hat{f}_t(\mathbf{d}_{[1t]}) = f^{k_1}(d_1)\cdot f^{k_2(d_1)}(d_2)\cdots f^{k_t(\mathbf{d}_{[1,t-1]})}(d_t), \\ \hat{g}_t(\mathbf{d}_{[1t]}) = g^{k_1}(d_1)\cdot g^{k_2(d_1)}(d_2)\cdots g^{k_t(\mathbf{d}_{[1,t-1]})}(d_t). \end{cases} \tag{12.14}$$

Then from (12.13),

$$\begin{aligned} &\mathrm{D}_{KL}(\hat{f}_t||\hat{g}_t) = \textstyle\sum_{d_1=0}^{\bar{d}} \cdots \sum_{d_t=0}^{\bar{d}} f^{k_1}(d_1)\cdot f^{k_2(d_1)}(d_2)\cdots f^{k_t(\mathbf{d}_{[1,t-1]})}(d_t)\times \\ &\times \ln[f^{k_1}(d_1)\cdot f^{k_2(d_1)}(d_2)\cdots f^{k_t(\mathbf{d}_{[1,t-1]})}(d_t)/(g^{k_1}(d_1)\cdot g^{k_2(d_1)}(d_2)\cdots g^{k_t(\mathbf{d}_{[1,t-1]})}(d_t))], \end{aligned} \tag{12.15}$$

which is further equal to

$$\begin{aligned}
&\textstyle\sum_{d_1=0}^{\bar{d}} f^{k_1}(d_1)\cdot[\ln(f^{k_1}(d_1)/g^{k_1}(d_1)) \\
&\qquad+\textstyle\sum_{d_2=0}^{\bar{d}} f^{k_2(d_1)}(d_2)\cdot[\ln(f^{k_2(d_1)}(d_2)/g^{k_2(d_1)}(d_2))+\cdots\cdots \\
&\qquad\qquad+\textstyle\sum_{d_t=0}^{\bar{d}} f^{k_t(\mathbf{d}_{[1,t-1]})}(d_t)\cdot\ln(f^{k_t(\mathbf{d}_{[1,t-1]})}(d_t)/g^{k_t(\mathbf{d}_{[1,t-1]})}(d_t))\cdots]] \\
&=\mathrm{D}_{KL}(f^{k_1}||g^{k_1})+\textstyle\sum_{d_1=0}^{\bar{d}} f^{k_1}(d_1)\cdot[\mathrm{D}_{KL}(f^{k_2(d_1)}||g^{k_2(d_1)})+\cdots\cdots \\
&\qquad+\textstyle\sum_{d_{t-1}=0}^{\bar{d}} f^{k_{t-1}(\mathbf{d}_{[1,t-2]})}(d_{t-1})\cdot\mathrm{D}_{KL}(f^{k_t(\mathbf{d}_{[1,t-1]})}||g^{k_t(\mathbf{d}_{[1,t-1]})})\cdots] \\
&=\textstyle\sum_{k=1}^{\bar{k}}\mathrm{D}_{KL}(f^k||g^k)\cdot\{\mathbf{1}(k_1=k)+\sum_{d_1=0}^{\bar{d}} f^{k_1}(d_1)\cdot[\mathbf{1}(k_2(d_1)=k)+\cdots \\
&\qquad+\textstyle\sum_{d_{t-1}=0}^{\bar{d}} f^{k_{t-1}(\mathbf{d}_{[1,t-2]})}(d_{t-1})\cdot\mathbf{1}(k_t(\mathbf{d}_{[1,t-1]})=k)\cdots]\} \\
&=\textstyle\sum_{k=1}^{\bar{k}}\mathrm{D}_{KL}(f^k||g^k)\cdot\mathbb{E}_{\hat{f}_t}[\mathcal{N}_t^k].
\end{aligned} \tag{12.16}$$

Replacing $\mathbf{f}$ by our current $\mathbf{f}_{b,T}$, and $\mathbf{g}$ by our current $\mathbf{f}_{a,T}$, while noting that $\mathbf{f}_{b,T}$ and $\mathbf{f}_{a,T}$ differ only under the price choice 1, we can obtain from (12.15) and (12.16) that

$$\mathrm{D}_{KL}(\hat{f}_{b,T,t-1}||\hat{f}_{a,T,t-1})=\mathrm{D}_{KL}(f_{b,T}^1||f_{a,T}^1)\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_{t-1}^1]. \tag{12.17}$$

Since $\mathrm{D}_{KL}(f_{b,T}^1||f_{a,T}^1)$ is proportional to ϵ_T^2, the above indicates that one lower bound for $\sup_{\mathbf{f}\in(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}} R_{\mathbf{f}}^T(\mathbf{k},\mathbf{y})$ will come in the form of

$$R_{ba,T}=A''\cdot\epsilon_T\cdot\sum_{t=1}^{T}\exp\left(-B''\cdot\epsilon_T^2\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_{t-1}^1]\right), \tag{12.18}$$

where A'' and B'' are constants. The better an upper bound we have for $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$, the better a lower bound we will have for $\sup_{\mathbf{f}\in(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}} R_{\mathbf{f}}^T(\mathbf{k},\mathbf{y})$. For instance, if $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$ and hence $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_{t-1}^1]$ were bounded by a constant times $T^{2/3}$, then we could choose ϵ_T as "bad" as being proportional to $T^{-1/3}$. This would make $R_{ba,T}$ bounded by a constant times $T^{2/3}$.

Our introduction of $\mathbf{f}_{c,T}$ is for the purpose of bounding $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$. For $R_{\mathbf{f}_{c,T}}^T(\mathbf{k}.\mathbf{y})$ which is above a constant times $\eta_T\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ to be upper-bounded by a constant

times $T^{2/3}$, while η_T is decreasing in T very slowly say at the pace of $T^{-\nu}$ for some tiny $\nu > 0$, we must have $\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ being upper-bounded by a constant times $T^{2/3+\nu}$. Because $\mathrm{D}_{KL}(f_{c,T}^1||f_{b,T}^1)$ is proportional to η_T^2, which though slowly will decrease to 0 as $T \to +\infty$. So decisions under $\mathbf{f}_{b,T}$ and $\mathbf{f}_{c,T}$ should be close. If this were exploited efficiently, this might lead to

$$\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1] \leq \mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1] + \text{a slow-growing term.} \tag{12.19}$$

If we succeeded in (12.19) and arrived to a bound for $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1]$ that is proportional to $T^{2/3+2\nu}$ say, then we would be able to prove a lower bound for the regret that is proportional to $T^{2/3-\nu}$—just consider $\epsilon_T = T^{-1/3-\nu}$ in (12.18).

This now leads to our potentially improvable lower bound, Theorem 4.

Proof of Theorem 4: Throughout, we let items be perishable. The same lower bound will certainly be true when (3) is further required. Let β still stand for the parameter $\bar{b}/(\bar{h}+\bar{b}) \in (0,1)$. We first concentrate on the case with (12.11). Define $\gamma \in (\beta, 1)$ such that

$$\gamma = 1 - \frac{\bar{p}^1 - \bar{c} - \bar{b}}{\bar{p}^2 - \bar{c} - \bar{b}} \cdot (1-\beta). \tag{12.20}$$

Consider $\mathbf{f} \equiv (f^k)_{k=1,2,...,\bar{k}} \in (\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ such that

$$f^1(0) = \beta, \qquad f^1(1) = 1-\beta, \qquad f^1(2) = f^1(3) = \cdots = 0, \tag{12.21}$$

$$f^2(0) = \gamma, \qquad f^2(1) = 1-\gamma, \qquad f^2(2) = f^2(3) = \cdots = 0, \tag{12.22}$$

and for $k = 3, 4, ..., \bar{k}$,

$$f^k(0) = 1, \qquad f^k(1) = 0, \qquad f^k(2) = f^k(3) = \cdots = 0. \tag{12.23}$$

From (9), we know $y^*_{f^1} = y^*_{f^2} = y^*_{f^3} \cdots = y^*_{f^{\bar{k}}} = 0$; thus, by (37), (38), (102), and (12.20),

$$V^1_{f^1} = V^2_{f^2} = (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1-\beta) = (\bar{p}^2 - \bar{c} - \bar{b}) \cdot (1-\gamma) > V^3_{f^3} = \cdots = V^{\bar{k}}_{f^{\bar{k}}} = 0. \quad (12.24)$$

At each T, consider alternatives $\mathbf{f}_{a,T}$, $\mathbf{f}_{b,T}$, and $\mathbf{f}_{c,T}$ to $\mathbf{f}$, with $\mathbf{f}_{a,T} = (f^1_{a,T}, f^2, f^3, ..., f^{\bar{k}})$, $\mathbf{f}_{b,T} = (f^1_{b,T}, f^2, f^3, ..., f^{\bar{k}})$, and $\mathbf{f}_{c,T} = (f^1_{c,T}, f^2, f^3, ..., f^{\bar{k}})$. Let ϵ_T and η_T be constants within $(0, [\beta \wedge (1-\gamma)]/2)$ that satisfy $\epsilon_T < \eta_T$; also, let

$$f^1_{a,T}(0) = \beta - \epsilon_T \qquad f^1_{a,T}(1) = 1 - \beta + \epsilon_T, \qquad f^1_{a,T}(2) = f^1_{a,t}(3) = \cdots = 0, \quad (12.25)$$

$$f^1_{b,T}(0) = \beta + \epsilon_T, \qquad f^1_{b,T}(1) = 1 - \beta - \epsilon_T, \qquad f^1_{b,T}(2) = f^1_{b,T}(3) = \cdots = 0, \quad (12.26)$$

$$f^1_{c,T}(0) = \beta + \eta_T, \qquad f^1_{c,T}(1) = 1 - \beta - \eta_T, \qquad f^1_{c,T}(2) = f^1_{c,T}(3) = \cdots = 0, \quad (12.27)$$

Due to (9), we have $y^*_{f^1_{a,T}} = 1$ and $y^*_{f^1_{b,T}} = y^*_{f^1_{c,T}} = 0$. By (37) and (38), it also follows that

$$\begin{aligned} V^1_{f^1_{a,T}} &= V_{f^1_{a,T}}(\bar{p}^1, 1) = (\bar{p}^1 - \bar{c}) \cdot (1 - \beta + \epsilon_T) - \bar{h} \cdot (\beta - \epsilon_T) \\ &= (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1-\beta) + (\bar{p}^1 - \bar{c} + \bar{h}) \cdot \epsilon_T = V^2_{f^2} + (\bar{p}^1 - \bar{c} + \bar{h}) \cdot \epsilon_T, \end{aligned} \quad (12.28)$$

$$V_{f^1_{a,T}}(\bar{p}^1, 0) = (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1 - \beta + \epsilon_T) = V^2_{f^2} + (\bar{p}^1 - \bar{c} - \bar{b}) \cdot \epsilon_T, \quad (12.29)$$

$$\begin{aligned} V^1_{f^1_{b,T}} &= V_{f^1_{b,T}}(\bar{p}^1, 0) = (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1 - \beta - \epsilon_T) \\ &= (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1-\beta) - (\bar{p}^1 - \bar{c} - \bar{b}) \cdot \epsilon_T = V^2_{f^2} - (\bar{p}^1 - \bar{c} - \bar{b}) \cdot \epsilon_T, \end{aligned} \quad (12.30)$$

$$V_{f^1_{b,T}}(\bar{p}^1, 1) = (\bar{p}^1 - \bar{c}) \cdot (1 - \beta - \epsilon_T) - \bar{h} \cdot (\beta + \epsilon_T) = V^2_{f^2} - (\bar{p}^1 - \bar{c} + \bar{h}) \cdot \epsilon_T, \quad (12.31)$$

$$\begin{aligned} V^1_{f^1_{c,T}} &= V_{f^1_{c,T}}(\bar{p}^1, 0) = (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1 - \beta - \eta_T) \\ &= (\bar{p}^1 - \bar{c} - \bar{b}) \cdot (1-\beta) - (\bar{p}^1 - \bar{c} - \bar{b}) \cdot \eta_T = V^2_{f^2} - (\bar{p}^1 - \bar{c} - \bar{b}) \cdot \eta_T, \end{aligned} \quad (12.32)$$

$$V_{f^1_{c,T}}(\bar{p}^1, 1) = (\bar{p}^1 - \bar{c}) \cdot (1 - \beta - \eta_T) - \bar{h} \cdot (\beta + \eta_T) = V^2_{f^2} - (\bar{p}^1 - \bar{c} + \bar{h}) \cdot \eta_T. \quad (12.33)$$

Now from (12.13) and (12.25) to (12.27), we can obtain

$$\mathrm{D}_{KL}(f^1_{b,T}||f^1_{a,T}) = (\beta+\epsilon_T)\cdot\ln\left(\frac{\beta+\epsilon_T}{\beta-\epsilon_T}\right)+(1-\beta-\epsilon_T)\cdot\ln\left(\frac{1-\beta-\epsilon_T}{1-\beta+\epsilon_T}\right), \quad (12.34)$$

$$\mathrm{D}_{KL}(f^1_{c,T}||f^1_{b,T}) = (\beta+\eta_T)\cdot\ln\left(\frac{\beta+\eta_T}{\beta+\epsilon_T}\right)+(1-\beta-\eta_T)\cdot\ln\left(\frac{1-\beta-\eta_T}{1-\beta-\epsilon_T}\right). \quad (12.35)$$

From (12.34),

$$\begin{aligned}\mathrm{D}_{KL}(f^1_{b,T}||f^1_{a,T}) &= (\beta+\epsilon_T)\cdot\ln(1/(1-x))+(1-\beta-\epsilon_T)\cdot\ln(1/(1+y))\\ &\le (\beta+\epsilon_T)\cdot(x+x^2)+(1-\beta-\epsilon_T)\cdot(-y+y^2)\\ &= 4\epsilon_T^2\cdot[1/(\beta+\epsilon_T)+1/(1-\beta-\epsilon_T)] \le 8\epsilon_T^2/(\beta\cdot(1-\beta)),\end{aligned} \quad (12.36)$$

where $x = 2\epsilon_T/(\beta+\epsilon_T)$ and $y = 2\epsilon_T/(1-\beta-\epsilon_T)$. From (12.35),

$$\begin{aligned}\mathrm{D}_{KL}(f^1_{c,T}||f^1_{b,T}) &= (\beta+\eta_T)\cdot\ln(1/(1-x))+(1-\beta-\eta_T)\cdot\ln(1/(1+y))\\ &\le (\beta+\eta_T)\cdot(x+x^2)+(1-\beta-\eta_T)\cdot(-y+y^2)\\ &= (\eta_T-\epsilon_T)^2/((\beta+\eta_T)\cdot(1-\beta-\eta_T)) < 2\eta_T^2/(\beta\cdot(1-\beta)),\end{aligned} \quad (12.37)$$

where $x = (\eta_T-\epsilon_T)/(\beta+\eta_T)$ and $y = (\eta-\epsilon_T)/(1-\gamma-\eta_T)$.

For $i = a, b, c$, let $\hat{f}_{i,T,t-1}$ be the distribution on $\mathcal{F}^{t-1}$ resulting from applying the current policy to demand-distribution vector $\mathbf{f}_{i,T}$ for $t-1$ periods. Thus, with (12.15), the definitions of $\mathbf{f}_{a,T}$, $\mathbf{f}_{b,T}$, and $\mathbf{f}_{c,T}$, as well as (12.36) and (12.37),

$$\mathrm{D}_{KL}(\hat{f}_{b,T,t-1}||\hat{f}_{a,T,t-1}) = \mathrm{D}_{KL}(f^1_{b,T}||f^1_{a,T})\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}^1_{t-1}] \le \frac{8\epsilon_T^2}{\beta\cdot(1-\beta)}\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}^1_{t-1}], \quad (12.38)$$

$$\mathrm{D}_{KL}(\hat{f}_{c,T,t-1}||\hat{f}_{b,T,t-1}) = \mathrm{D}_{KL}(f^1_{c,T}||f^1_{b,T})\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}^1_{t-1}] \le \frac{2\eta_T^2}{\beta\cdot(1-\beta)}\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}^1_{t-1}]. \quad (12.39)$$

Under a given adaptive policy $(\mathbf{k}, \mathbf{y})$, let $\mathcal{N}^{k,y}_t$ be the number of periods from 1 to t

under which the price choice is k and ordering decision is y:

$$\mathcal{N}_t^{k,y} = \sum_{s=1}^{t} \mathbf{1}(k_s = k \text{ and } y_s = y). \tag{12.40}$$

Compared to $\mathcal{N}_t^k$ in (42), it is certainly true that $\mathcal{N}_t^k = \sum_{y=0}^{\bar{d}} \mathcal{N}_t^{k,y}$.

Meanwhile, (41), (12.24), and (12.28) to (12.33) reveal that

$$\begin{aligned} R^T_{\mathbf{f}_{a,T}}(\mathbf{k},\mathbf{y}) \;\; \geq\; & (\bar{h}+\bar{b})\cdot\epsilon_T\cdot\mathbb{E}_{\hat{f}_{a,T}}[\mathcal{N}_T^{1,0}] + (\bar{p}^1-\bar{c}+\bar{h})\cdot\epsilon_T\cdot\mathbb{E}_{\hat{f}_{a,T}}[\mathcal{N}_T^{2}] \\ & +(\bar{p}^1-\bar{c}-\bar{b})\cdot(1-\beta)\cdot\mathbb{E}_{\hat{f}_{a,T}}[T-\mathcal{N}_T^{1,0}-\mathcal{N}_T^{1,1}-\mathcal{N}_T^{2}], \end{aligned} \tag{12.41}$$

$$\begin{aligned} R^T_{\mathbf{f}_{b,T}}(\mathbf{k},\mathbf{y}) \;\; \geq\; & (\bar{p}^1-\bar{c}-\bar{b})\cdot\epsilon_T\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^{1,0}] + (\bar{p}^1-\bar{c}+\bar{h})\cdot\epsilon_T\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^{1,1}] \\ & +(\bar{p}^1-\bar{c}-\bar{b})\cdot(1-\beta)\cdot\mathbb{E}_{\hat{f}_{b,T}}[T-\mathcal{N}_T^{1,0}-\mathcal{N}_T^{1,1}-\mathcal{N}_T^{2}], \end{aligned} \tag{12.42}$$

$$\begin{aligned} R^T_{\mathbf{f}_{c,T}}(\mathbf{k},\mathbf{y}) \;\; \geq\; & (\bar{p}^1-\bar{c}-\bar{b})\cdot\eta_T\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^{1,0}] + (\bar{p}^1-\bar{c}+\bar{h})\cdot\eta_T\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^{1,1}] \\ & +(\bar{p}^1-\bar{c}-\bar{b})\cdot(1-\beta)\cdot\mathbb{E}_{\hat{f}_{c,T}}[T-\mathcal{N}_T^{1,0}-\mathcal{N}_T^{1,1}-\mathcal{N}_T^{2}]. \end{aligned} \tag{12.43}$$

Hence,

$$\begin{cases} R^T_{\mathbf{f}_{a,T}}(\mathbf{k},\mathbf{y}) \geq (\bar{h}+\bar{b})\cdot\epsilon_T\cdot\sum_{t=1}^{T}\mathbb{P}_{\hat{f}_{a,T}}[k_t\neq 1 \text{ or } y_t\neq 1], \\ R^T_{\mathbf{f}_{b,T}}(\mathbf{k},\mathbf{y}) \geq (\bar{h}+\bar{b})\cdot\epsilon_T\cdot\sum_{t=1}^{T}\mathbb{P}_{\hat{f}_{b,T}}[k_t= 1 \text{ and } y_t= 1], \end{cases} \tag{12.44}$$

which, due to (12.12), will lead to

$$R^T_{\mathbf{f}_{a,T}}(\mathbf{k},\mathbf{y}) \vee R^T_{\mathbf{f}_{b,T}}(\mathbf{k},\mathbf{y}) \geq \frac{1}{8}\cdot(\bar{h}+\bar{b})\cdot\epsilon_T\cdot\sum_{t=1}^{T}\exp\left(-\mathrm{D}_{KL}(\hat{f}_{b,T,t-1}||\hat{f}_{a,T,t-1})\right), \tag{12.45}$$

where we have used the fact that, for positive numbers $x_1, y_1, ..., x_n, y_n$,

$$(x_1+\cdots+x_n)\vee(y_1+\cdots+y_n) \geq \frac{x_1+y_1}{2}+\cdots+\frac{x_n+y_n}{2} \geq \frac{x_1\vee y_1}{2}+\cdots+\frac{x_n\vee y_n}{2}. \tag{12.46}$$

Now combine (12.38) and (12.43) with (12.45), and we can obtain

$$\sup_{\mathbf{f}\in(\mathcal{F}_\infty(\bar{d}))^{\bar{k}}} R_{\mathbf{f}}^T(\mathbf{k},\mathbf{y}) \geq R_{ba,T} \vee R_{c,T}, \tag{12.47}$$

where

$$R_{ba,T} = \frac{1}{8}\cdot(\bar{h}+\bar{b})\cdot\epsilon_T\cdot\sum_{t=1}^{T}\exp\left(-\frac{8\epsilon_T^2}{\beta\cdot(1-\beta)}\cdot\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_{t-1}^1]\right), \tag{12.48}$$

$$R_{c,T} = (\bar{p}^1-\bar{c}-\bar{b})\cdot\eta_T\cdot\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]. \tag{12.49}$$

Note that $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_{t-1}^1] \leq \mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1] \leq T$. When $\epsilon_T = T^{-1/2}$, every exponential term in (12.48) will be bounded from below by a constant. This will lead to a $T^{1/2}$-sized lower bound for $R_{ba,T}$. We will thus obtain the theorem's bound using (12.47).

Our bound would improve if (12.19) could be achieved by exploiting the yet untouched (12.39) and (12.49). The best of our effort is summarized. Using a trick from Auer et al. [3], we see that $\mathbb{E}_{\hat{f}_b}[\mathcal{N}_{t-1}^1]$ is linked with $\mathbb{E}_{\hat{f}_c}[\mathcal{N}_{t-1}^1]$. By (12.39), $\mathbb{E}_{\hat{f}_{b,T}}[\mathcal{N}_T^1] - \mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ is equal to

$$\begin{aligned}
&\textstyle\sum_{d_1=0}^{\bar{d}}\cdots\sum_{d_{T-1}=0}^{\bar{d}}\mathbf{1}(\mathcal{N}_T^1(\mathbf{d}_{[1,T-1]}))\cdot[\hat{f}_{b,T}(\mathbf{d}_{[1,T-1]})-\hat{f}_{c,T}(\mathbf{d}_{[1,T-1]})]\\
&\quad\leq T\cdot\textstyle\sum_{d_1=0}^{\bar{d}}\cdots\sum_{d_{T-1}=0}^{\bar{d}}\mathbf{1}(\hat{f}_{b,T}(\mathbf{d}_{[1,T-1]})\geq\hat{f}_{c,T}(\mathbf{d}_{[1,T-1]})\times\\
&\qquad\qquad\times[\hat{f}_{b,T}(\mathbf{d}_{[1,T-1]})-\hat{f}_{c,T}(\mathbf{d}_{[1,T-1]})]\\
&\quad= T\cdot||\hat{f}_{b,T,t-1}-\hat{f}_{c,T,t-1}||_1/2 \leq T\cdot(2\cdot\mathrm{D}_{KL}(\hat{f}_{c,T,t-1}||\hat{f}_{b,T,t-1}))^{1/2}/2\\
&\quad\leq T\cdot\eta_T\cdot(\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_{t-1}^1])^{1/2}/(\beta\cdot(1-\beta))^{1/2},
\end{aligned} \tag{12.50}$$

where the first equality realizes that $\mathcal{N}_T^1$ achieves the same value under the same demand path under both distributions, and the first inequality is attributable to $\mathcal{N}_T^1 \leq T$. Note that $||\hat{f}_{b,T,t-1}-\hat{f}_{c,T,t-1}||_1$ stands for the sum of terms $|\hat{f}_{b,T,t-1}(\mathbf{d}_{[1,t-1]})-\hat{f}_{c,T,t-1}(\mathbf{d}_{[1,t-1]})|$ and its relation with $\mathrm{D}_{KL}(\hat{f}_{c,T,t-1}||\hat{f}_{b,T,t-1})$ is known as Pinsker's inequality. From (12.49), we can obtain a $T^{2/3+2\nu}$-sized upper

bound for $\mathbb{E}_{\hat{f}_{c,T}}[\mathcal{N}_T^1]$ if $R_{c,T}$ is to be kept below some constant times $T^{2/3+\nu}$ while $\eta_T = T^{-\nu}$. Unfortunately, this fails to make (12.50) much closer to (12.19).

For the general case where (102) but not necessarily (12.11) is maintained, let $f_{00} \in \mathcal{F}_0$ be the all-zero demand distribution such that

$$f_{00}(0) = 1, \qquad f_{00}(1) = f_{00}(2) = \cdots = 0. \tag{12.51}$$

For $V_f(p, y)$ defined at (37), note that $V_{f_{00}}(p, y) = -\bar{b} \cdot y$ for any price p and any order-up-to level $y = 0, 1, ..., \bar{d}$; hence, for V_f^k defined at (38), it follows that $V_{f_{00}}^k = V_{f_{00}}(\bar{p}^k, 0) = 0$. Let $f_\beta \in \mathcal{F}_0$ be the demand distribution with a β portion on $\bar{d} - 1$ and a $1 - \beta$ portion on $\bar{d}$:

$$f_\beta(0) = \cdots = f_\beta(\bar{d} - 2) = f_\beta(\bar{d} + 1) = \cdots = 0, \qquad f_\beta(\bar{d} - 1) = \beta, \qquad f_\beta(\bar{d}) = 1 - \beta. \tag{12.52}$$

From (9), note that $y_{f_\beta}^* = \bar{d} - 1$. The earlier (102) amounts to

$$V_{f_\beta}^{\bar{k}} = V_{f_\beta}(\bar{p}^{\bar{k}}, \bar{d} - 1) = V_{f_\beta}(\bar{p}^{\bar{k}}, \bar{d}) = (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - 1) - (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot (1 - \beta) > 0. \tag{12.53}$$

Consider $\mathbf{f} \equiv (f^1, f^2, ..., f^{\bar{k}}) \in (\mathcal{F}_\infty(\bar{d}))^{\bar{k}}$ such that

$$f^1 = f^2 = \cdots = f^{\bar{k}-1} = f_{00}, \qquad f^{\bar{k}} = f_{00}. \tag{12.54}$$

Our construction has ensured that

$$V_{f^{\bar{k}}}^{\bar{k}} = (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - 1) - (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot (1 - \beta) > 0 = V_{f^1}^1 = \cdots = V_{f^{\bar{k}-1}}^{\bar{k}-1}. \tag{12.55}$$

Consider perturbations $\mathbf{f}_{a,T} \equiv (f^1, ..., f^{\bar{k}-1}, f_{a,T}^{\bar{k}})$ and $\mathbf{f}_{b,T} \equiv (f^1, ..., f^{\bar{k}-1}, f_{b,T}^{\bar{k}})$. Let ϵ_T

be a constant within $(0, [\beta \wedge (1-\beta)]/2)$ and

$$\begin{cases} f^{\bar{k}}_{a,T}(0) = \cdots = f^{\bar{k}}_{a,T}(\bar{d}-2) = f^{\bar{k}}_{a,T}(\bar{d}+1) = f^{\bar{k}}_{a,T}(\bar{d}+2) = \cdots = 0, \\ f^{\bar{k}}_{a,T}(\bar{d}-1) = \beta - \epsilon_T, \qquad\qquad f^{\bar{k}}_{a,T}(\bar{d}) = 1 - \beta + \epsilon_T, \end{cases} \tag{12.56}$$

$$\begin{cases} f^{\bar{k}}_{b,T}(0) = \cdots = f^{\bar{k}}_{b,T}(\bar{d}-2) = f^{\bar{k}}_{b,T}(\bar{d}+1) = f^{\bar{k}}_{b,T}(\bar{d}+2) = \cdots = 0, \\ f^{\bar{k}}_{b,T}(\bar{d}-1) = \beta + \epsilon_T, \qquad\qquad f^{\bar{k}}_{a,T}(\bar{d}) = 1 - \beta - \epsilon_T. \end{cases} \tag{12.57}$$

Due to (9), $y^*_{f^{\bar{k}}_{a,T}} = \bar{d}$ and $y^*_{f^{\bar{k}}_{b,T}} = y^*_{f^{\bar{k}}_{c,T}} = \bar{d}-1$. By (37) and (38), it also follows that

$$\begin{aligned} V^{\bar{k}}_{f^{\bar{k}}_{a,T}} &= V_{f^{\bar{k}}_{a,T}}(\bar{p}^{\bar{k}}, \bar{d}) = (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta + \epsilon_T) - \bar{h} \cdot (\beta - \epsilon_T) \\ &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta) - \bar{h} \cdot \beta + (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T = V^2_{f^2} + (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T, \end{aligned} \tag{12.58}$$

$$\begin{aligned} V_{f^{\bar{k}}_{a,T}}(\bar{p}^{\bar{k}}, \bar{d}-1) &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta + \epsilon_T) - \bar{b} \cdot (1 - \beta + \epsilon_T) \\ &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta) - \bar{b} \cdot (1 - \beta) - (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot \epsilon_T = V^2_{f^2} - (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot \epsilon_T, \end{aligned} \tag{12.59}$$

$$\begin{aligned} V^{\bar{k}}_{f^{\bar{k}}_{b,T}} = V_{f^{\bar{k}}_{b,T}}(\bar{p}^{\bar{k}}, \bar{d}-1) &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta - \epsilon_T) - \bar{b} \cdot (1 - \beta - \epsilon_T) \\ &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta) - \bar{b} \cdot (1 - \beta) + (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot \epsilon_T = V^2_{f^2} + (\bar{b} + \bar{c} - \bar{p}^{\bar{k}}) \cdot \epsilon_T, \end{aligned} \tag{12.60}$$

$$\begin{aligned} V_{f^{\bar{k}}_{b,T}}(\bar{p}^{\bar{k}}, \bar{d}) &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta - \epsilon_T) - \bar{h} \cdot (\beta + \epsilon_T) \\ &= (\bar{p}^{\bar{k}} - \bar{c}) \cdot (\bar{d} - \beta) - \bar{h} \cdot \beta - (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T = V^2_{f^2} - (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T, \end{aligned} \tag{12.61}$$

From these, we obtain

$$\begin{cases} R^T_{\mathbf{f}_{a,T}}(\mathbf{k}, \mathbf{y}) \geq (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T \cdot \sum_{t=1}^T \mathbb{P}_{\hat{f}_{a,T}}[k_t \neq \bar{k} \text{ or } y_t \neq \bar{d}], \\ R^T_{\mathbf{f}_{b,T}}(\mathbf{k}, \mathbf{y}) \geq (\bar{p}^{\bar{k}} - \bar{c} + \bar{h}) \cdot \epsilon_T \cdot \sum_{t=1}^T \mathbb{P}_{\hat{f}_{b,T}}[k_t = \bar{k} \text{ and } y_t = \bar{d}], \end{cases} \tag{12.62}$$

much like the earlier (12.44). Meanwhile, due to (12.13), (12.56), and (12.57), $\mathrm{D}_{KL}(f^{\bar{k}}_{b,T} || f^{\bar{k}}_{a,T})$ is again upper-bounded by a constant times ϵ^2_T, much like the earlier (12.36). A $T^{1/2}$-sized lower bound for the regret will then follow from using

the same logic employed earlier. ∎

13 References

[1] Araman, V.F. and R. Caldentey. 2009. Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Research*, **57**, pp. 1169-1188.

[2] Auer, P., N. Cesa-Bianchi, and P. Fischer. 2012. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, **47**, pp. 235-256.

[3] Auer, P., N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. 2002. The Nonstochastic Multiarmed Bandit Problem. *SIAM Journal on Computing*, **32**, pp. 48-77.

[4] Auer, P. and R. Ortner. 2007. Logarithmic Online Regret Bounds for Reinforcement Learning. *Advances in Neural Information Processing Systems*, **19**, pp. 49-56.

[5] Aviv, Y. and A. Pazgal. 2005. A Partially Observed Markov Decision Process for Dynamic Pricing. *Management Science*, **51**, pp. 1400-1416.

[6] Besbes, O. and A. Muharremoglu. 2013. On Implications of Demand Censoring in the Newsvendor Problem. *Management Science*, **59**, pp. 1407-1424.

[7] Besbes, O. and A. Zeevi. 2009. Dynamic Pricing without Knowing the Demand Function: Risk Bounds and Near-optimal Algorithms. *Operations Research*, **57**, pp. 1407-1420.

[8] Besbes, O. and A. Zeevi. 2015. On the (Surprising) Sufficiency of Linear Models for Dynamic Pricing with Demand Learning. *Management Science*, **61**, pp. 723-739.

[9] den Boer, A. and B. Zwart. 2014. Simultaneously Learning and Optimizing using Controlled Variance Pricing. *Management Science*, **60**, pp. 770-783.

[10] Broder, J. and P. Rusmevichientong. 2012. Dynamic Pricing under a General Parametric Choice Model. *Operations Research*, **60**, pp. 965-980.

[11] Burnetas, A.N., O. Kanavetas, and M.N. Katehakis. 2016. Asymptotically Optimal Multi-armed Bandit Policies under a Cost Constraint. *Probability in the Engineering and Informational Sciences*, **30**, pp. 1-27.

[12] Burnetas, A.N. and M.N. Katehakis. 1997. Optimal Adaptive Policies for Markov Decision Processes. *Mathematics of Operations Research*, **22**, pp. 222-255.

[13] Burnetas, A.N. and C.E. Smith. 2000. Adaptive Ordering and Pricing for Perishable Products. *Operations Research*, **48**, pp. 436-443.

[14] Chen, B., X. Chao, and H.-S. Ahn. 2015. Coordinating Pricing and Inventory Replenishment with Nonparametric Demand Learning. Working Paper, University of Michigan.

[15] Chen, B., X. Chao, and C. Shi. 2016. Nonparametric Algorithms for Joint Pricing and Inventory Control with Lost-sales and Censored Demand. Working Paper, University of Michigan.

[16] Cheung, W.-C., D. Simchi-Levi, and H. Wang. 2017. Dynamic Pricing and Demand Learning with Limited Price Experimentation. *Operations Research*, **65**, pp. 1722-1731.

[17] Cover, T.M. and J.A. Thomas. 2006. *Elements of Information Theory, 2nd Edition*. Wiley-Interscience, New York.

[18] Dembo, A. and O. Zeitouni. 1998. *Large Deviations Techniques and Applications, 2nd Edition*. Springer, Heidelberg.

[19] Derman, C. 1957. Non-parametric Up-and-down Experimentation. *Annals of Mathematical Statistics*, **28**, pp. 795-798.

[20] Farias, V. and B. van Roy. 2010. Dynamic Pricing with a Prior on Market Response. *Operations Research*, **58**, pp. 16-29.

[21] Federgruen, A. and A. Heching. 1999. Combined Pricing and Inventory Control Under Uncertainty. *Operations Research*, **47**, pp. 454-475.

[22] Ferreira, K.J., D. Simchi-Levi, and H. Wang. 2015. Online Network Revenue Management using Thompson Sampling. Working Paper, Massachusetts Institute of Technology.

[23] Hoeffding, W. 1963. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, **58**, pp. 13-30.

[24] Huh, W.T. and P. Rusmevichientong. 2009. A Non-parametric Asymptotic Analysis of Inventory Planning with Censored Demand. *Mathematics of Operations Research*, **34**, pp. 103-123.

[25] Huh, W.T., R. Levi, P. Rusmevichientong, and J.B. Orlin. 2011. Adaptive Data=driven Inventory Control with Censored Demand Based on Kaplan-Meier Estimator. *Operations Research*, **59**, pp. 929-941.

[26] Jaksch, T, R. Ortner, and P. Auer. 2010. Near-optimal Regret Bounds for Reinforcement Learning. *Journal of Machine Learning Research*, **11**, pp. 1563-1600.

[27] Katehakis, M.N. and H. Robbins. 1995. Sequential Choice from Several Populations. *Proceedings of the National Academy of Sciences*, **92**, pp. 8584-8585.

[28] Kiefer, J. and J. Wolfowitz. 1952. Stochastic Estimation of the Maximum of a Regression Function. *Annals of Mathematical Statistics*, **23**, pp. 462-466.

[29] Lai, T.L. and H. Robbins. 1985. Asymptotically Efficient Adaptive Allocation Rules. *Advances in Applied Mathematics*, **6**, pp. 4-22.

[30] Lariviere, M.A. and E.L. Porteus. 1999. Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales. *Management Science*, **43**, pp. 346-363.

[31] Levi, R., R.O. Roundy, and D.B. Shmoys. 2007. Provably Near-optimal Sampling-based Policies for Stochastic Inventory Control Models. *Mathematics of Operations Research*, **32**, pp. 821-839.

[32] Robbins, H. 1952. Some Aspects of the Sequential Design of Experiments. *Bulletins of American Mathematical Society*, **58**, pp. 527-535.

[33] Robbins, H and S. Monro, S. 1951. A Stochastic Approximation Method. *Annals of Mathematical Statistics*, **22**, pp. 400-407.

[34] Scarf, H. 1959. Bayes Solutions of the Statistical Inventory Problem. *Annals of Mathematical Statistics*, **30**, pp. 490-508.

[35] Tewari, A and P.L. Bartlett. 2007. Optimistic Linear Programming Gives Logarithmic Regret for Irreducible MDPs. *Advances in Neural Information Processing Systems*, **20**, pp. 1-8.

[36] Tsybakov, A. 2008. *Introduction to Nonparametric Estimation*. Springer, Berlin.

[37] Wang, Z., S. Deng, and Y Ye. 2014. Close the Gaps: A Learning-while-Doing Algorithm for Single-Product Revenue Management Problems. *Operations Research*, **62**, pp. 318-331.