Stochastic Regret Minimization for Revenue Management Problems with Nonstationary Demands

Huanan Zhang*, Cong Shi*, Chao Qin†, Cheng Hua‡

* Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109
{zhanghn, shicong}@umich.edu

† Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208,
chaoqin2019@u.northwestern.edu

‡ Yale School of Management, Yale University, New Haven, CT 06511, cheng.hua@yale.edu

We study an admission control model in revenue management with non-stationary and correlated demands over a finite discrete time horizon. The arrival probabilities are updated by current available information, i.e., past customer arrivals and some other exogenous information. We develop a regret-based framework, which measures the difference in revenue between a clairvoyant optimal policy that has access to all realizations of randomness a priori and a given feasible policy which does not have access to this future information. This regret minimization framework better spells out the trade-offs of each accept/reject decision. We proceed using the lens of approximation algorithms to devise a conceptually simple regret-parity policy. We show the proposed policy achieves 2-approximation of the optimal policy in terms of total regret for a two-class problem, and then extend our results to a multi-class problem with a fairness constraint. Our goal in this paper is to make progress towards understanding the marriage between stochastic regret minimization and approximation algorithms in the realm of revenue management and dynamic resource allocation.

Key words: algorithms; admission control; revenue management; regret minimization; non-homogeneous Poisson processes; worst-case bounds

History: Received October 2014; revisions received June 2015, May 2016, August 2016; accepted August 2016.

1. Introduction

We propose and analyze a simple admission control policy for a class of revenue management problems under non-stationary customer arrivals. There is a given positive and non-replenishable inventory M of some product to be sold to arriving customers of two different classes in a finite time horizon T. Class-1 customers are willing to pay r_1 per unit of product which is more than what Class-2 customers are willing to pay (i.e., $r_1 \geq r_2 \geq 0$). The demand for each class is modeled as a non-homogeneous Poisson process whose (instantaneous) arrival rates are time-varying and correlated and whose distributions can be updated by current available information, i.e., past customer arrivals and some other exogenous information. This is the main new feature of this paper, which captures realistic phenomena such as demand seasonality and forecast updating mechanisms. Unsatisfied demand units are lost with no penalty cost. The firm can decide whether to accept or reject an arriving customer, so as to maximize the expected revenue over the planning horizon.

The model is motivated by a wide range of applications, such as vacation timeshare management, online retailing, and workforce management. For example, Hilton Grand Vacations Club offers timeshares at different prices and levels of membership. A platinum customer will have a higher priority than a non-platinum (or regular) customer when it comes to selecting a particular home resort (e.g., reserving a room in Elara on the Las Vegas strip). The arrival process of customers clearly depends on the tourism seasons (e.g., more platinum customers will select Elara Las Vegas during the Christmas season) as well as the total number of members of Hilton Grand Vacations Club. Another example is in online retailing. Anthropologie clothing online offers discount coupons to customers. The regular customers (without coupons), who are willing to pay the tagged prices, are always accepted. On the other hand, the discounted customers (with coupons) could be rejected since these coupon code offers are subject to discretion and availability. The demand process is also non-stationary and evolving, according to the season and the product's popularity.

It is worthwhile noting that our model encompasses many important non-stationary demand processes studied in the literature, including Markov modulated demand processes (described in §6), time series models [43], the martingale model of forecast evolution [23, 28] and models with advance demand information [18]. However, finding the optimal admission control policies using brute-force dynamic programming is computationally intractable, since the state space of the corresponding dynamic programs is usually large (which is extensively discussed in §2.1). Hence, our main focus of this paper is to prescribe an effective and provably-good heuristic policy for this class of problems.

1.1. Main Results and Contributions

The main results and contributions of this paper are summarized as follows.

First, we study the aforementioned class of admission control based revenue management problems using a regret minimization framework. The regret of a feasible policy is defined as the difference in revenue between a clairvoyant optimal policy (that has access to all realizations of randomness a priori at the beginning of the time horizon) and the feasible policy (which does not have access to this future information). We propose a conceptually simple admission control policy, called the *regret-parity* policy $\tilde{\pi}$, that perfectly balances the regret of an acceptance decision against that of a rejection decision. We show that the regret ratio of $\tilde{\pi}$ (defined as the ratio of the regret of $\tilde{\pi}$ to the regret of an optimal policy) is always bounded above by 2 in a two-class setting (Theorem 1). We then extend our model and results to a multi-class setting under a fairness constraint (Theorem 2).

This regret-based performance measure is different from the conventional revenue-based performance measure (defined as the ratio of the revenue of a feasible policy to that of an optimal policy). In many applications such as online retailing with low price discrimination (e.g., Anthropologie clothing online), the after-tax profit margin is very thin (around 5%) and even a small improvement is significant. In such cases, the revenue difference between two feasible policies would be small, and the regret ratio could arguably better gauge the effectiveness of a given feasible policy by quantifying its operational mistakes, thereby improving the firm's overall decision making and profitability. We discuss their connections between these two performance measures in §3.3.

Second, our numerical results demonstrate the efficacy of the proposed regret-parity policy $\tilde{\pi}$ under a large set of demand and parameter instances. The empirical performance of $\tilde{\pi}$ is usually much better than 2. More specifically, Tables 1–3 show that $\tilde{\pi}$ performs consistently well in term of expected revenue and regret, compared to an optimal policy. Compared to the robust benchmark algorithms proposed in Ball and Queyranne [1], we gain around 16% more expected revenue, which is quite significant. It is worth noting that the proposed policy can be efficiently implemented in an online manner, i.e., the decision at any time is computed based only on the current observed state of the system and does not depend on future decisions. This is in contrast to solving an optimal policy exactly using a brute-force dynamic programming approach, which suffers from the curse of dimensionality.

Finally, we note that the regret-parity policy belongs to the family of cost-balancing policies that are predominantly used in stochastic inventory control problems (cf. Levi et al. [37, 39, 36]). The main idea underlying this approach is to isolate and quantify the marginal impact of each operational decision (from the moment it is made until the end of the planning horizon). When we consider the problem of interest from the view of revenue maximization, it is straightforward to count the immediate revenue resulted from each acceptance/rejection decision, but it is difficult (or perhaps impossible) to measure how each acceptance/rejection decision impacts the overall (future) revenue. As a result, the conventional methods developed in their papers cannot be directly applied in the revenue management setting. In contrast, under the regret minimization framework, we are able to readily quantify the marginal impact of each acceptance/rejection decision in terms of regret (relative to a full-information benchmark). This enables us to design an efficient and effective cost-balancing algorithm, and compare the costs of two different policies. Our worst-case analysis involves dealing with this regret-based (mistake-based) objective as well as a randomized decision rule, which advances the current methodology in cost-balancing algorithms.

We believe that the ideas and techniques developed in this paper could be applied to other classes of revenue management or resource allocation problems. The notion of approximation ratios (or worst-case performance guarantees) has also been gaining acceptance in the revenue management literature (see, e.g., Chen and Farias [10], Dragos and Farias [15], Chan and Farias [8]). For instance, Chen and Farias [10] gives a class of *re-optimized fixed price* (RFP) policies that yields at least 0.342 of the optimal policy for a classical single-product dynamic pricing problem but allowing the scale of demand intensity to be modulated by an exogenous market size stochastic process.

1.2. Relevant Literature

Our work is closely related to the following streams of literature.

Revenue management. Most revenue management models assume that the demand process is a time-homogeneous, mainly for its mathematical tractability (see, e.g., [19, 27, 6, 54, 14, 5, 51]). Revenue management models with non-stationary demand environments are much less common in the literature. Gallego and van Ryzin [20] studied dynamic pricing problems in a changing demand environment where the temporal evolution of the demand model is known. They established asymptotic optimality for their policies by solving a deterministic counterpart problem. Netessine [44] analyzed the pricing problem with a limited number of price changes in a dynamic environment in which demand depends on the current price and time. Zhao and Zheng [57] considered a continuous time dynamic pricing problem with non-homogeneous Poisson processes, and showed that the optimal price decreases with inventory. They also identified a sufficient condition under which the optimal price decreases over time for a given inventory level. Cao et al. [7] considered a similar problem with non-homogeneous Poisson customer arrival processes, and obtained the structural properties of optimal policies by a Hamilton-Jacobi-Bellman (HJB) equation.

It shall be noted that our worst-case regret ratio has a similar philosophical underpinning to the rapidly growing area of robust optimization. Since the future demand information is often uncertain and evolving, the firm only has limited (present) information to make good decisions. Robust optimization has been widely adopted under limited and sparse information to protect the firm from the worst-case scenarios (see, e.g., Ben-Tal and Nemirovski [2] and Bertsimas and Sim [3]). In the revenue management literature, Perakis and Roels [46] developed robust formulations for capacity allocation in network revenue management. Birbil et al. [4] devised an efficient algorithm based on robust optimization to compute the maximum booking limits in a single-leg airline revenue management problem. Lan et al. [35] focused on the relative regret in overbooking and fare-class allocation for a multi-fare, single-resource problem in revenue management. Rusmevichientong and

Topaloglu [49] studied robust formulations of assortment optimization problems under the multinomial logit choice model. Geng et al. [21] studied a two-customer sequential resource allocation problem with a max-min fill-rate objective, and characterized the structure of optimal solutions with a bounded discrete distribution. Closer to our work, Ball and Queyranne [1] carried out a thorough competitive analysis of nested booking limits in an online adversarial setting. The main point of departure from their work is that we consider a regret-based (mistake-based) objective, and therefore the main results obtained are incomparable.

Stochastic knapsack and dynamic resource allocation. Another relevant line of research to this work is the class of dynamic and stochastic knapsack problems. Papastavrou et al. [45] and Kleywegt and Papastavrou [32, 33] considered variants of dynamic and stochastic knapsack problems where items (with random size and rewards) arrive according to a time-homogeneous Poisson process, and an accept/reject decision needs to be made upon each item's arrival so as to maximize the expected profit (rewards minus costs) accumulated. They showed that a threshold-type policy is optimal and also derived a number of monotonicity and convexity properties. Lueker [41] gave an $O(\log n)$-competitive algorithm for the 0/1 online knapsack problem, where n is the number of arriving items. Dean et al. [13] also considered a stochastic 0/1 knapsack problem with deterministic arrivals and item values but random item sizes. They bounded its adaptivity gap by developing a polynomial-time algorithm that computes a non-adaptive policy whose expected value approximates that of an optimal adaptive policy within a factor of 4. These models are very similar to ours; however, our work considers a non-stationary, correlated and evolving demand process, which requires new analytical methods to be analyzed.

Our work is related to the domain of online reservation or selection problems. Elmachtoub and Levi [17, 16] considered online versions of supply chain management and logistics models where customers arrive sequentially, and one has to decide whether to accept or reject the customer upon her arrival. They developed several algorithms with small constant competitive ratios, i.e., for any sequence of arriving customers, the cost incurred by the online algorithm is within a fixed constant factor of the cost incurred by the respective optimal solution that has full knowledge upfront on the sequence of arriving customers. Van Hentenryck et al. [53] proposed constant approximation algorithms for online reservation or online multi-knapsack problems with or without overbooking. We also refer interested readers to Coffman Jr. et al. [12] for an excellent survey on online bin packing problems. Our regret-parity framework shares some similarities with the competitive performance measures used there, in that the common benchmark involves the full-information (or offline) solution.

Another relevant domain is dynamic resource allocation in controlled queueing and communication networks (see Kelly [30] for an overview). Most papers on dynamic resource allocation problems also assume time-homogeneous Poisson arrival processes (see, e.g., [31], [22], [29]). Closer to our work, Levi and Radovanović [38] used a simple knapsack-type linear program (LP) to decide whether to accept or reject incoming customer requests. They showed that their proposed policy is guaranteed to achieve at least half of the optimal long-run revenue. However, the counterpart models with non-stationary arrivals are invariably much harder to study (e.g., [48] and [26]). Yoon and Lewis [55] proposed a pointwise stationary approximation (PSA) to approximate the optimal policies in a multi-class queueing system with non-homogeneous Poisson arrival processes and periodically varying parameters. Green and Kolesar [24] and Massey and Whitt [42] considered peak hour congestion in a multi-server queuing system under non-homogeneous Poisson arrival processes. Kumar et al. [34] devised dynamic control policies for a single-server queue with Markov modulated arrivals. A key difference between this line of research and our work is that the resource units in our model are non-reusable, i.e., once sold, they cannot be used to satisfy other customers.

Other related work. Our work is also closely related to the development of approximation algorithms that admit constant worst-case performance guarantees (see, e.g., Levi et al. [37, 39, 36], Levi and Shi [40], Shi et al. [50], Chao et al. [9]) predominantly in various stochastic inventory control settings). As mentioned earlier, the conventional techniques and methods developed in their papers cannot be directly applied to the revenue management setting; in order to establish a worst-case performance guarantee of 2, one needs to combine them with the regret minimization framework.

1.3. Organization

The rest of this paper is organized as follows. In §2, we first describe the discrete time model formulation for a two-class revenue management problem under non-stationary customer demands. We then present a dynamic programming formulation in §2.1 and a stochastic regret minimization formulation in §2.2. In §3, we propose a different regret accounting scheme based on decisions. Then we devise and analyze the regret-parity policy in §4. We extend our model and results to the multi-class setting under a fairness constraint in §5, and conduct numerical experiments of our proposed policy in §6. We then point out some plausible future research avenues in §7.

2. Two-Class Problem Formulation

We present the mathematical model for a two-class problem under non-stationary and correlated customer demands. As a general convention, we often distinguish between a random variable and its realization using capital letters and lower case letters, respectively.

Consider a firm selling a fixed number of M perishable homogeneous items to two classes of customers, indexed by $i = 1, 2$, over a finite planning horizon T periods numbered $t = 1, \ldots, T$. Inventory is not replenishable. Unsatisfied demand is lost with no penalty cost, and any unsold items at the end of period T have no residual value or disposal penalty. Each class-i customer pays r_i dollars ($r_1 \geq r_2 \geq 0$) for a single item. Class-1 customers are always served whenever inventory units are available; however, the firm needs to decide whether to accept or reject an arriving Class-2 customer, depending on information available, such as current inventory level, the number of periods remaining, and conditional future demand distributions. The objective is to develop a provably-good admission control policy (for accepting Class-2 customers) that maximizes the expected total revenue over the planning horizon.

We describe the demand process of our model. In each time period $t = 1, \ldots, T$, there is at most one arriving customer who wishes to request a single item. The probabilities of having no customer request, a Class-1 customer request, and a Class-2 customer request are denoted by p_t^0, p_t^1, and $p_t^2 = 1 - p_t^0 - p_t^1$, respectively. As part of the model, we assume that at the beginning of each period $t = 1, \ldots, T$, the firm is endowed with an observed *information set* f_t, which contains all the realized demand information that is available at the beginning of time period t. More specifically, the information set f_t consists of the realized customer requests over the set of periods $[1, t]$, and possibly some external information such as the state of the economy and the weather. The information set f_t is a specific realization from the set of all possible realizations, denoted by F_t. The future arrival probabilities over the set of periods $(t, T]$ are updated by the information set f_t, i.e., $p_s^0 = p_s^0(f_t)$, $p_s^1 = p_s^1(f_t)$ and $p_s^2 = p_s^2(f_t)$ for all $s \in (t, T]$. With these updated arrival probabilities, the firm knows the conditional joint distribution of future customer requests, denoted by $I_t = I_t(f_t)$. Our model allows for non-stationarity and correlation among the demands in different periods. We note again that by allowing for correlation we let I_t be dependent on the realization of the customer requests over the set of periods $[1, t]$ and possibly on some external information, i.e., I_t is a function of f_t. However, the information set f_t as well as the conditional joint distribution I_t are assumed to be *independent* of the specific admission control policy being considered. In other words, the admission control policy does not have any effect on the evolution of the future demands.

Next, we describe the system dynamics. At the beginning of each period $t = 1, \ldots, T$, the firm observes the customer request (if any) and its class, and then makes a decision whether to accept or reject the incoming customer request. We let $\alpha_t \in \{0, 1\}$ be a binary decision variable, where 0 denotes a rejection and 1 denotes an acceptance. We always accept Class-1 customers (i.e., $\alpha_t = 1$) as long as the inventory is non-empty, since $r_1 \geq r_2 \geq 0$. The firm needs to decide α_t whenever a

Class-2 customer arrives in period t. Let X_t and Y_t be the inventory levels in period t before and after a decision in period t is made, respectively. We have that the initial inventory $X_1 = M$, and $X_{t+1} = Y_t = X_t - \alpha_t$ for all $t = 1, \ldots, T$. We only restrict our attentions to *state-dependent* policies which are non-anticipatory, i.e., in each period t, the information that a feasible admission control policy π can use consists of the information set f_t, and past decisions and inventory levels up to period t.

2.1. Dynamic Programming Formulation

Our two-class model can be formulated using dynamic programming below. We denote $V_t(x_t, f_t)$ as the optimal expected revenue over the set of periods $[t, T]$, with the starting inventory level x_t and the information set f_t. Since optimal policies will always accept Class-1 customers as long as there is positive inventory, the Bellman's equation is given by

$$\begin{aligned} V_t(x_t, f_t) = \mathbb{E}\Bigg[\Big(& p_t^0 \cdot V_{t+1}(x_t, F_{t+1}) + p_t^1 \cdot (r_1 + V_{t+1}(x_t - 1, F_{t+1})) \\ & + p_t^2 \cdot \max\left\{r_2 + V_{t+1}(x_t - 1, F_{t+1}), V_{t+1}(x_t, F_{t+1})\right\}\Big) \,\Big|\, f_t\Bigg], \end{aligned} \tag{1}$$

with boundary conditions $V_{T+1}(\cdot) = 0$ and $V_t(0, f_t) = 0$, $t = 1, \ldots, T$. It can be seen that the state space grows exponentially fast when the arrival rates are correlated over time. As a result, computing exact optimal policies using dynamic programming is intractable, due to the well-known *curse of dimensionality* [47]. This motivates us to devise a conceptually simple and provably-good approximation algorithm to solve this class of problems.

2.2. Sample Path Regret and its Explicit Expression

We define the random variable $W(\pi; f_t)$ as the total revenue of any feasible policy π given information set f_t. Then the *full-information* revenue $W(\pi; f_T)$ represents the total revenue of π given a fully realized sample path f_T, which is a deterministic value. To properly define our *regret*, we carefully distinguish between two different notions of optimality.

(a) Let the *clairvoyant* optimal policy along a given sample path f_T be

$$\pi^* = \pi^*(f_T) = \arg\max_{\pi} W(\pi; f_T).$$

Note that π^* knows the full-information f_T *a priori* at the beginning of period 1. Given a specific realization f_T, one can write down π^* instantly without any optimization procedures.

(b) Let the optimal policy for our original stochastic control problem be

$$\pi^o = \arg\max_{\pi} \mathbb{E}\left[W\left(\pi; f_1\right)\right],$$

where the expectation is taken over all possible realizations f_T. Note that π^o only knows f_t at the beginning of period $t = 1, \ldots, T$, respectively, and π^o is in fact the optimal control of the dynamic programming (1) that attains an expected optimal revenue $V_1(x_1, f_1)$.

With the above notion, the sample path regret of a given feasible policy π is defined as

$$\mathcal{R}\left(\pi; f_T\right) \triangleq W\left(\pi^*; f_T\right) - W\left(\pi; f_T\right), \tag{2}$$

which is the difference in revenue between the clairvoyant optimal policy π^* (which has access to the entire realization f_T *a priori* at the beginning of period 1) and the feasible policy π (which only knows f_1 at the beginning of period 1). The expected regret via (2) is defined as

$$\mathbb{E}\left[\mathcal{R}(\pi)\right] = \mathbb{E}\left[\mathcal{R}\left(\pi; F_T\right)\right], \tag{3}$$

where the expectation is taken over all possible realizations $f_T \in F_T$.

Now we find a more explicit expression of the sample path regret defined in (2). Given any feasible policy π and any sample path f_T, we let $C^1_\pi = C^1_\pi\left(f_T\right)$ and $C^2_\pi = C^2_\pi\left(f_T\right)$ be the numbers of Class-1 and Class-2 customers accepted by π, respectively.

Proposition 1. *The sample path regret $\mathcal{R}\left(\pi; f_T\right)$ defined in (2) can be re-written as*

$$\mathcal{R}\left(\pi; f_T\right) = (r_1 - r_2)(C^2_\pi - C^2_{\pi^*})^+ + r_2(C^2_{\pi^*} - C^2_\pi)^+. \tag{4}$$

We relegate the detailed proof of Proposition 1 to the Appendix. There is an intuitive explanation of (4). If the number of Class-2 customers accepted by π is greater than that accepted by π^* (i.e., $C^2_\pi \geq C^2_{\pi^*}$), then π "wrongly" accepts $C^2_\pi - C^2_{\pi^*}$ Class-2 customers rather than Class-1 customers, and the regret is the cost difference $r_1 - r_2$ for each such wrong admission. On the other hand, if the number of Class-2 customers accepted by π is less than that accepted by π^* (i.e., $C^2_{\pi^*} \geq C^2_\pi$), then π loses sales of $C^2_{\pi^*} - C^2_\pi$ Class-2 customers and the regret is r_2 for each such lost-sale.

3. Regret-Based Reformulation

There is a clear trade-off between accepting and rejecting an arriving Class-2 customer. That is, if we accept the Class-2 customer, we gain a revenue rate of r_2; however, we may potentially lose a sale of a Class-1 customer when the inventory is used up. On the other hand, if we reject the Class-2 customer, we may eventually lose a sale of r_2 if the inventory remains positive at the end of the planning horizon. Each acceptance or rejection comes with a regret (loss). Our approach attempts to exploit the trade-off of each decision. We introduce additional notation. Let the random variable $A^i_{(t,T]}$ $(i=1,2)$ denote the number of class-i customers that will arrive over the set of periods $(t,T]$. Similarly, let $A^{1,2}_{(t,T]}$ denote the total number of customers that will arrive over the set of periods $(t,T]$.

3.1. Regret of Acceptance

Given any feasible policy π, let $RA^\pi_t(\alpha^\pi_t)$ be the regret of acceptance decision α^π_t made in period t under π when the arriving customer in period t belongs to Class-2, which is given by

$$\begin{aligned} RA^\pi_t(\alpha^\pi_t) = RA_t(\alpha^\pi_t; X^\pi_t) &= (r_1 - r_2)\cdot \mathbb{1}\left(A^1_{(t,T]} \geq X^\pi_t > 0 \text{ and } \alpha^\pi_t = 1\right) \\ &= (r_1 - r_2)\left[\left(A^1_{(t,T]} - X^\pi_t + \alpha^\pi_t\right)^+ - \left(A^1_{[t,T]} - X^\pi_t\right)^+\right] \\ &= (r_1 - r_2)\left[\left(A^1_{(t,T]} - Y^\pi_t\right)^+ - \left(A^1_{[t,T]} - X^\pi_t\right)^+\right]. \end{aligned} \tag{5}$$

This is because that if a Class-2 customer arrives in period t, by accepting her, the firm incurs a regret of $r_1 - r_2$ only when the event $\left\{A^1_{(t,T]} \geq X^\pi_t > 0\right\}$ occurs, since the firm could have sold this item to a Class-1 customer. Note that the last equality of (5) also remains valid for the other two cases. If no customer arrives in period t, we incur zero regret. If a Class-1 customer arrives in period t, by accepting her as long as the inventory is positive, the firm incurs zero regret, i.e., $RA^\pi_t(1) = 0$ since $A^1_{(t,T]} + 1 = A^1_{[t,T]}$ and $Y^\pi_t + 1 = X^\pi_t$.

3.2. Regret of Rejection

Given any feasible policy π, let $RR^\pi_t(\alpha^\pi_t)$ be the regret of rejection decision α^π_t made in period t under π when the arriving customer in period t belongs to Class-2, which is given by

$$\begin{aligned} RR^\pi_t(\alpha^\pi_t) = RR_t(\alpha^\pi_t; X^\pi_t) &= r_2 \cdot \mathbb{1}\left(A^{1,2}_{(t,T]} < X^\pi_t \text{ and } X^\pi_t > 0 \text{ and } \alpha^\pi_t = 0\right) \\ &= r_2\left[\left(X^\pi_t - A^{1,2}_{(t,T]} - \alpha^\pi_t\right)^+ - \left(X^\pi_t - A^{1,2}_{[t,T]}\right)^+\right] \\ &= r_2\left[\left(Y^\pi_t - A^{1,2}_{(t,T]}\right)^+ - \left(X^\pi_t - A^{1,2}_{[t,T]}\right)^+\right]. \end{aligned} \tag{6}$$

This is because that if a Class-2 customer arrives in period t, by rejecting her, the firm incurs a regret of r_2 only when the event $\left\{A^{1,2}_{(t,T]} < X^\pi_t \text{ and } X^\pi_t > 0\right\}$ occurs, since the firm has positive inventory at the end of period T and could have gained r_2 from this Class-2 customer. Note that the last equality of (6) also remains valid for the other two cases. If no customer arrives in period t, we incur zero regret. If a Class-1 customer arrives in period t, by rejecting her only when the inventory is zero, the firm incurs zero regret, i.e., $RR_t^\pi(0)=0$ since $Y_t^\pi = X_t^\pi = 0$.

3.3. Regret-Based Performance Measure

The next result asserts that the regrets (associated with each individual decision) defined in §3.1–3.2 add up to the total regret defined in §2.2. We delegate its proof to the Appendix.

Proposition 2. *The total regret of decisions can be re-written (along every sample path) as*

$$\mathcal{R}(\pi; f_T) = \sum_{t=1}^{T} \left[RA_t^\pi(\alpha_t^\pi) + RR_t^\pi(\alpha_t^\pi) \right] \Big| f_T.$$

Proposition 2 allows us to reformulate the dynamic programming (1) from a viewpoint of regret minimization. The original dynamic programming (1) views this revenue management problem as gradually gaining revenue from zero to the final total revenue as time progresses. The regret minimization problem takes a dual view of the original revenue maximization problem. More specifically, we start with the *clairvoyant* optimal revenue $W(\pi^*; f_T)$ at the beginning, and in each period we make an admission decision, incurring either the regret of acceptance or the regret of rejection. After each decision is made, the revenue is penalized by the computed regret. From this dual view, we start with the highest possible revenue and gradually decrease it as time progresses.

This regret minimization reformulation can also be cast as a dynamic program. Denote $G_t(x_t, f_t)$ as the minimum expected regret over the periods $[t, T]$, with the starting inventory x_t and the information set f_t. To minimize the regret, the Bellman's equation is given by

$$\begin{aligned} G_t(x_t, f_t) &= \mathbb{E}\left[p_t^0 \cdot G_{t+1}(x_t, F_{t+1}) + p_t^1 \cdot G_{t+1}(x_t - 1, F_{t+1}) \,\Big|\, f_t \right] \\ &+ p_t^2 \cdot \min\left\{ \mathbb{E}\left[RA_t(1; x_t) + G_{t+1}(x_t - 1, F_{t+1}) \,\Big|\, f_t \right], \mathbb{E}\left[RR_t(0; x_t) + G_{t+1}(x_t, F_{t+1}) \,\Big|\, f_t \right] \right\}, \end{aligned} \tag{7}$$

with boundary conditions $G_{T+1}(\cdot) = 0$ and $G_t(0, f_t) = 0$, $t = 1, \ldots, T$. By Proposition 2 and (2), the optimal decisions of (1) and (7) are identical. Moreover, the expected total regret and the expected total revenue sum up to the expected clairvoyant optimal revenue. It is important to note that the regret minimization formulation gives the same optimal stochastic control as the revenue maximization formulation. However, the performance measure is different under this regret-based reformulation. We define the *regret ratio* of π to be

$$\mathscr{R}(\pi) \triangleq \frac{\mathbb{E}\left[\mathcal{R}(\pi)\right]}{\mathbb{E}\left[\mathcal{R}(\pi^o)\right]}, \tag{8}$$

where the expected regret $\mathbb{E}[\mathcal{R}(\pi^o)]$ defined in (3) is in fact the optimal expected regret $G_1(x_1, f_1)$ solved using the dynamic program defined in (7). It is clear that $1 \leq \mathscr{R}(\pi) \leq \infty$.

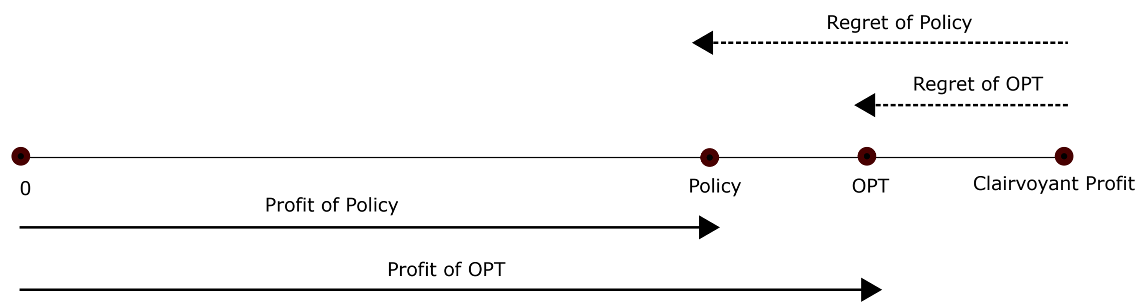

Figure 1 **Regret-Based Reformulation and Performance Measure**

We shall draw the connection between the regret ratio defined in (8) under this equivalent regret minimization reformulation and the conventional revenue ratio defined by the ratio of the total revenue of π to that of the optimal policy π^o under the original revenue maximization formulation. As illustrated in Figure 1, for each instance of the problem, a lower regret ratio of π always leads to a higher revenue ratio of π, and vice versa. However, these two performance measures are *incomparable*. In many retail industries with low price discrimination, the regret ratio can better capture how good a feasible policy π is (see Example 1 below). The regret ratio "zooms in" on the operational mistakes that a firm makes, thereby improving the (already thin) profit margin.

Example 1. Let us consider a practical example in which the price discrimination is rather low, i.e., r_1 and r_2 are quite close, say $r_1 = 100$ and $r_2 = 95$. This example is abundant in practical settings, e.g., coupon discounts, omnichannel retailing (in-store vs. online), etc. Let us consider a simple (and clearly sub-optimal) feasible policy that accepts all customers as long as the inventory does not run out, which we call it "all-accept" policy. It is clear that the revenue difference between an optimal policy and the all-accept policy is always bounded by 5% (regardless of the input stochastic processes), if the firm chooses to use the traditional revenue-based performance measure. However, this all-accept policy is undoubtedly a very poor heuristic policy, since it ignores all the inventory and demand information.

In contrast, in such cases, the regret ratios of sub-optimal policies similar to all-accept policies are usually very high, which better captures the "real" performance of such policies. For instance, one can fix $M = 3$ and construct a sample path with Class-$2,2,2,1,1,1$ arrival sequence, the all-accept policy would accept $2,2,2$ while the optimal policy would accept $1,1,1$. In this case, the revenue error is only 5% while the regret error is infinity (since none of the decisions from the all-accept policy are correct)! □

The above example shows that the regret measure quantifies the regret or the cost of a poor decision in a much lucid way, which allows the firm to better trade-off between acceptance and rejection. We remark that this notion of regret ratio defined in (8) has also been proposed and used in other fields such as the theory of online and statistical learning (see, e.g., Guha and Munagala [25] that gives a conceptually similar regret ratio).

4. Approximation Algorithm: Regret-Parity Policy

In this section, we propose and analyze an efficient and effective admission-control policy called the *regret-parity* policy, denoted by $\tilde{\pi}$, which aims to exactly balance between the regret of acceptance and the regret of rejection. The proposed policy $\tilde{\pi}$ is a randomized policy, which makes a randomized admission decision in each period based on computed probabilities.

4.1. Policy Description

To fully describe and analyze $\tilde{\pi}$ which invovles *randomzied* decision rules, we introduce the *expanded information set* f_t^+ that not only includes the original information set f_t but also all the randomized decisions of $\tilde{\pi}$ up to period $t-1$. Thus, given $\tilde{\pi}$ and f_t^+, the inventory level x_t at the beginning of period t is *known* but the decision in period t remains *unknown*. In addition, we define f_T^{++} as f_T^+ plus the decision made in period T, which constitutes a full sample path.

Now we describe $\tilde{\pi}$ as follows. In each period $t = 1, \ldots, T$, if a Class-1 customer arrives, we accept her as long as the inventory $x_t > 0$. On the other hand, if a Class-2 customer arrives and $x_t > 0$, $\tilde{\pi}$ accepts her with probability θ_t and reject her with probability $1 - \theta_t$. That is, we set

$$\alpha_t^{\tilde{\pi}} = \begin{cases} 1, \text{ with probability } \theta_t, \\ 0, \text{ with probability } 1 - \theta_t, \end{cases}$$

where probability θ_t is computed by solving

$$\theta_t \cdot \mathbb{E}\left[RA_t^{\tilde{\pi}}(1) \mid f_t^+\right] = (1 - \theta_t) \cdot \mathbb{E}\left[RR_t^{\tilde{\pi}}(0) \mid f_t^+\right], \tag{9}$$

where $RA_t^{\tilde{\pi}}(\cdot)$ and $RR_t^{\tilde{\pi}}(\cdot)$ are defined in (5) and (6), respectively. It is clear that the proposed regret-parity policy $\tilde{\pi}$ strikes an exact balance between the two types of regrets via (9).

We also discuss here how to efficiently evaluate the expectation $\mathbb{E}\left[RA_t^{\tilde{\pi}}(1) \mid f_t^+\right]$ (and similarly $\mathbb{E}\left[RR_t^{\tilde{\pi}}(0) \mid f_t^+\right]$) in practical implementations. First observe via (5) that because $A_{[t,T]}^1$ takes integer values from 0 to $T - t + 1$, evaluating $\mathbb{E}\left[RA_t^{\tilde{\pi}}(1) \mid f_t^+\right]$ has the same complexity of computing $\mathbb{P}(A_{[t,T]}^1 = i)$ for $i = 0, \ldots, T - t + 1$. When the demands are i.i.d., the computation is easy since

$A^1_{[t,T]}$ follows a binomial distribution. When the demands are generally correlated, e.g., the Markov modulated demand processes tested in §6.1, computing the exact values of $\mathbb{P}(A^1_{[t,T]} = i)$ for $i = 0,\ldots,T-t+1$ is not straightforward. For practical purposes, we use Monte Carlo simulation method to obtain a very close estimation. In our numerical experiments, the coefficient of variation of using 5000 sample paths to estimate the expectation is generally less than 1%. We note that [36, 56] face the same computational challenges when evaluating similar expectations in other (inventory) settings. The main computational advantage of $\tilde{\pi}$ lies in that $\tilde{\pi}$ can be efficiently implemented in an online manner, i.e., the decision at any time is computed based only on the current observed state of the system and does not depend on future decisions. This is a desired property if one wishes to avoid the prohibitive (recursive) computational burden of solving large dynamic programs.

4.2. Performance Analysis

To establish a worst-case performance guarantee of 2, we wish to show that, on expectation, the total regret of the optimal policy π^o "pays" for at least half of that of the regret-parity policy $\tilde{\pi}$. In the subsequent analysis, we use superscript π^o to refer to the optimal policy that solves the dynamic programming (1), and superscript $\tilde{\pi}$ to refer to our regret-parity policy.

We first define a stopping time τ which records the first period time when the inventory of $\tilde{\pi}$ runs out. More specifically, we define

$$\tau = \inf\left\{t \in \{1,\ldots,T+1\} : X_t^{\tilde{\pi}} = 0\right\}. \tag{10}$$

Note that τ is well-defined since it is measurable w.r.t. the expanded information set f_t^+.

We then partition the set of periods $\{1,\ldots,T\}$ into three disjoint random subsets

$$\mathscr{T}_a = \left\{t \in \{1,\ldots,T\} : t < \tau \text{ and } Y_t^{\pi^o} \le X_t^{\tilde{\pi}} - 1\right\}, \tag{11}$$

$$\mathscr{T}_b = \left\{t \in \{1,\ldots,T\} : t < \tau \text{ and } Y_t^{\pi^o} \ge X_t^{\tilde{\pi}}\right\}, \tag{12}$$

$$\mathscr{T}_c = \{t \in \{1,\ldots,T\} : \tau \le t \le T\}. \tag{13}$$

Note that the above subsets are disjoint and exhaustive, and the indicators $\mathbb{1}(t \in \mathscr{T}_a)$, $\mathbb{1}(t \in \mathscr{T}_b)$ and $\mathbb{1}(t \in \mathscr{T}_c)$ become known with the expanded information set f_t^+.

Then, we prove two important lemmas below. We want to show that the total regret of acceptance incurred by π^o is higher than that incurred by $\tilde{\pi}$ in the set $\mathscr{T}_a$ in Lemma 1, and the total regret of rejection incurred by π^o is higher than that incurred by $\tilde{\pi}$ in the set $\mathscr{T}_b$ in Lemma 2.

LEMMA 1. *The total regret of acceptance by π^o is no smaller than that by $\tilde{\pi}$ in the set $\mathscr{T}_a$, i.e., along every sample path,*

$$\sum_{t=1}^{T} RA_t^{\pi^o}(\alpha_t^{\pi^o}) \geq \sum_{t\in\mathscr{T}_a} RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}).$$

Proof. We fix an arbitrary sample path f_T^{++}. Suppose there are l customer arrivals, and let $1 \leq t_1 \leq \ldots \leq t_l \leq T$ denote all these l customer arriving epochs. We then denote t_s to be the last customer arriving epoch that belongs to the set $\mathscr{T}_a$.

For each $k = 1, \ldots s-1$, since there is no customer arrival over (t_k, t_{k+1}), it is clear that $a^1_{(t_k,T]} = a^1_{[t_{k+1},T]}$ and $y^{\pi^o}_{t_k} = x^{\pi^o}_{t_{k+1}}$. This implies that

$$\left(a^1_{(t_k,T]} - y^{\pi^o}_{t_k}\right)^+ = \left(a^1_{[t_{k+1},T]} - x^{\pi^o}_{t_{k+1}}\right)^+. \tag{14}$$

Now using (5), we sum up the regret of acceptance over these arriving epochs along f_T^{++},

$$\begin{aligned}\sum_{k=1}^{s} RA^{\pi^o}_{t_k}(\alpha^{\pi^o}_{t_k}) &= \sum_{k=1}^{s}(r_1 - r_2)\left[\left(a^1_{(t_k,T]} - y^{\pi^o}_{t_k}\right)^+ - \left(a^1_{[t_k,T]} - x^{\pi^o}_{t_k}\right)^+\right] \\ &= (r_1 - r_2)\left[\left(a^1_{(t_s,T]} - y^{\pi^o}_{t_s}\right)^+ - \left(a^1_{[t_1,T]} - x^{\pi^o}_{t_1}\right)^+\right] \\ &= (r_1 - r_2)\left[\left(a^1_{(t_s,T]} - y^{\pi^o}_{t_s}\right)^+ - \left(a^1_{[1,T]} - M\right)^+\right],\end{aligned} \tag{15}$$

where the second equality follows from expanding the telescoping sum and (14); the third equality holds because there is no customer arrival before t_1, and thus $a^1_{[t_1,T]} = a^1_{[1,T]}$ and $x^{\pi^o}_{t_1} = M$.

Using the identical argument above, we also have the same expression for $\tilde{\pi}$,

$$\sum_{k=1}^{s} RA^{\tilde{\pi}}_{t_k}(\alpha^{\tilde{\pi}}_{t_k}) = (r_1 - r_2)\left[\left(a^1_{(t_s,T]} - y^{\tilde{\pi}}_{t_s}\right)^+ - \left(a^1_{[1,T]} - M\right)^+\right]. \tag{16}$$

Because $t_s \in \mathscr{T}_a$ implies that $y^{\pi^o}_{t_s} \leq x^{\tilde{\pi}}_{t_s} - 1 \leq y^{\tilde{\pi}}_{t_s}$, then we have

$$\sum_{t=1}^{T} RA_t^{\pi^o}(\alpha_t^{\pi^o}) \geq \sum_{k=1}^{s} RA^{\pi^o}_{t_k}(\alpha^{\pi^o}_{t_k}) \geq \sum_{k=1}^{s} RA^{\tilde{\pi}}_{t_k}(\alpha^{\tilde{\pi}}_{t_k}) \geq \sum_{t\in\mathscr{T}_a} RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}),$$

where the second inequality follows from comparing (15) and (16) with $y^{\pi^o}_{t_s} \leq y^{\tilde{\pi}}_{t_s}$. **Q.E.D.**

LEMMA 2. *The total regret of rejection by π^o is no smaller than that by $\tilde{\pi}$ in the set $\mathscr{T}_b$, i.e., along every sample path,*

$$\sum_{t=1}^{T} RR_t^{\pi^o}(\alpha_t^{\pi^o}) \geq \sum_{t\in\mathscr{T}_b} RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}).$$

Proof. We fix an arbitrary sample path f_T^{++}. Suppose there are l customer arrivals, and let $1 \le t_1 \le \ldots \le t_l \le T$ denote all these l customer arriving epochs. We then denote t_s to be the last customer arriving epoch that belongs to the set $\mathscr{T}_b$.

For each $k = 1, \ldots s-1$, since there is no customer arrival over (t_k, t_{k+1}), it is clear that $a^{1,2}_{(t_k,T]} = a^{1,2}_{[t_{k+1},T]}$ and $y^{\pi^o}_{t_k} = x^{\pi^o}_{t_{k+1}}$. This implies that

$$\left(y^{\pi^o}_{t_k} - a^{1,2}_{(t_k,T]}\right)^+ = \left(x^{\pi^o}_{t_{k+1}} - a^{1,2}_{[t_{k+1},T]}\right)^+. \tag{17}$$

Now using (6), we sum up the regret of rejection over these arriving epochs along f_T^{++},

$$\begin{aligned} \sum_{k=1}^{s} RR^{\pi^o}_{t_k}(\alpha^{\pi^o}_{t_k}) &= \sum_{k=1}^{s} r_2 \left[\left(y^{\pi^o}_{t_k} - a^{1,2}_{(t_k,T]}\right)^+ - \left(x^{\pi^o}_{t_k} - a^{1,2}_{[t_k,T]}\right)^+ \right] \\ &= r_2 \left[\left(y^{\pi^o}_{t_s} - a^{1,2}_{(t_s,T]}\right)^+ - \left(x^{\pi^o}_{t_1} - a^{1,2}_{[t_1,T]}\right)^+ \right] \\ &= r_2 \left[\left(y^{\pi^o}_{t_s} - a^{1,2}_{(t_s,T]}\right)^+ - \left(M - a^{1,2}_{[1,T]}\right)^+ \right], \end{aligned} \tag{18}$$

where the second equality follows from expanding the telescoping sum and (17); the third equality holds because there is no customer arrival before t_1, and thus $a^{1,2}_{[t_1,T]} = a^{1,2}_{[1,T]}$, and $x^{\pi^o}_{t_1} = M$.

Using the identical argument above, we also have the same expression for $\tilde{\pi}$,

$$\sum_{k=1}^{s} RR^{\tilde{\pi}}_{t_k}(\alpha^{\tilde{\pi}}_{t_k}) = r_2 \left[\left(y^{\tilde{\pi}}_{t_s} - a^{1,2}_{(t_s,T]}\right)^+ - \left(M - a^{1,2}_{[1,T]}\right)^+ \right]. \tag{19}$$

Because $t_s \in \mathscr{T}_b$ implies that $y^{\pi^o}_{t_s} \ge x^{\tilde{\pi}}_{t_s} \ge y^{\tilde{\pi}}_{t_s}$, then we have

$$\sum_{t=1}^{T} RR^{\pi^o}_{t}(\alpha^{\pi^o}_{t_k}) \ge \sum_{k=1}^{s} RR^{\pi^o}_{t_k}(\alpha^{\pi^o}_{t_k}) \ge \sum_{k=1}^{s} RR^{\tilde{\pi}}_{t_k}(\alpha^{\tilde{\pi}}_{t_k}) \ge \sum_{t \in \mathscr{T}_b} RR^{\tilde{\pi}}_{t}(\alpha^{\tilde{\pi}}_{t}),$$

where the second inequality follows from comparing (18) and (19) with $y^{\pi^o}_{t_s} \ge y^{\tilde{\pi}}_{t_s}$. **Q.E.D.**

Lemmas 1 and 2 establish a connection between π^o and $\tilde{\pi}$. To complete the worst-case analysis, we need a new variable $Z^{\tilde{\pi}}_t$ defined as

$$Z^{\tilde{\pi}}_t \triangleq \mathbb{E}\left[RA^{\tilde{\pi}}_t(\alpha^{\tilde{\pi}}_t) \mid F^+_t\right] = \mathbb{E}\left[RR^{\tilde{\pi}}_t(\alpha^{\tilde{\pi}}_t) \mid F^+_t\right]. \tag{20}$$

Note that $Z^{\tilde{\pi}}_t$ is a random variable that is realized with the information set f^+_t at the beginning of period t. Observe that by the construction of $\tilde{\pi}$, the random variable $Z^{\tilde{\pi}}_t$ is well-defined since the expected regret of acceptance and the expected regret of rejection are always balanced.

Lemma 3 below shows that the expected total regret of $\tilde{\pi}$ can be expressed using the $Z_t^{\tilde{\pi}}$ variables defined in (20).

LEMMA 3. *The expected total regret incurred by $\tilde{\pi}$ is*

$$\mathbb{E}\left[\mathcal{R}(\tilde{\pi})\right] = 2 \cdot \sum_{t=1}^{T} \mathbb{E}\left[Z_t^{\tilde{\pi}}\right].$$

Proof. By Proposition 2 and standard arguments of conditional expectations, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}(\tilde{\pi})\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\left[RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) + RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}})\right]\right] = \sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left[RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) + RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \mid F_t^+\right]\right] \\
&= \sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left[RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \mid F_t^+\right] + \mathbb{E}\left[RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \mid F_t^+\right]\right] = \sum_{t=1}^{T}\mathbb{E}\left[2Z_t^{\tilde{\pi}}\right] = 2 \cdot \sum_{t=1}^{T}\mathbb{E}\left[Z_t^{\tilde{\pi}}\right],
\end{aligned}$$

where the fourth equality follows directly from the definition in (20). **Q.E.D.**

Lemma 4 below shows that the expected total regret of π^o can be upper bounded using the $Z_t^{\tilde{\pi}}$ variables defined in (20).

LEMMA 4. *The expected total regret incurred by π^o is lower bounded by*

$$\mathbb{E}\left[\mathcal{R}(\pi^o)\right] \geq \sum_{t=1}^{T}\mathbb{E}\left[Z_t^{\tilde{\pi}}\right].$$

Proof. Combining Lemmas 1 and 2 and the fact the $Z_t^{\tilde{\pi}} = 0$ when $t \in \mathcal{T}_c$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}(\pi^o)\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\left[RA_t^{\pi^o}(\alpha_t^{\pi^o}) + RR_t^{\pi^o}(\alpha_t^{\pi^o})\right]\right] \\
&\geq \mathbb{E}\left[\sum_{t\in\mathscr{T}_a} RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) + \sum_{t\in\mathscr{T}_b} RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) + \sum_{t\in\mathscr{T}_c} 0\right] \\
&= \sum_{t=1}^{T}\mathbb{E}\left[RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \cdot \mathbb{1}(t\in\mathscr{T}_a) + RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \cdot \mathbb{1}(t\in\mathscr{T}_b) + 0 \cdot \mathbb{1}(t\in\mathscr{T}_c)\right] \\
&= \sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left[RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \cdot \mathbb{1}(t\in\mathscr{T}_a) + RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}) \cdot \mathbb{1}(t\in\mathscr{T}_b) + 0 \cdot \mathbb{1}(t\in\mathscr{T}_c)\right] \mid F_t^+\right] \\
&= \sum_{t=1}^{T}\mathbb{E}\left[(\mathbb{1}(t\in\mathscr{T}_a) + \mathbb{1}(t\in\mathscr{T}_b) + \mathbb{1}(t\in\mathscr{T}_c))Z_t^{\tilde{\pi}}\right] = \sum_{t=1}^{T}\mathbb{E}\left[Z_t^{\tilde{\pi}}\right],
\end{aligned}$$

where the fourth equality holds since $\mathbb{1}(t\in\mathscr{T}_a)$, $\mathbb{1}(t\in\mathscr{T}_b)$ and $\mathbb{1}(t\in\mathscr{T}_c)$ are measurable with the expanded information set F_t^+. **Q.E.D.**

Combining Lemmas 3 and 4, we have established that the $\tilde{\pi}$ policy has a constant worst-case performance guarantee of 2, which is stated formally below.

THEOREM 1. *For each instance of the two-class revenue management problem under non-stationary customer demands, the expected total regret incurred by the regret-parity policy $\tilde{\pi}$ is at most two times the expected total regret incurred by the optimal policy π^o, i.e.,*

$$\mathbb{E}\left[\mathcal{R}(\tilde{\pi})\right] \leq 2 \cdot \mathbb{E}\left[\mathcal{R}(\pi^o)\right].$$

Theorem 1 asserts that $\tilde{\pi}$ achieves a 2-approximation of the optimal policy in terms of stochastic relative regret, which establishes an interesting link between the relative regret and approximation algorithms in the revenue management setting.

5. Multi-Class Extension with Fairness

We consider an extension of our model to the multi-class setting which incorporates a *fairness* constraint. The model is almost identical to that defined in §2 but with I classes of customers. Without loss of generality, we let $r_1 \geq r_2 \geq \ldots \geq r_I \geq 0$, and we say class-$i$ has a higher priority than class-j whenever $i < j$. Similar to the two-class model, the total arrival rate and the probability of being class-i customer are both evolving over time. In this model, Class-1 customers are always accepted if there are inventory units available, but we need to make a decision whether to accept a customer if she is not a Class-1 customer. We define our notion of fairness.

DEFINITION 1. We say a feasible policy π is said to be fair if the following condition holds. If for each period $t = 1, \ldots, T$ with an arriving customer (of class $1 \leq j \leq T$), then for all $1 \leq i < j < k \leq I$, i.e., class-i (class-k) has a higher (lower) revenue or priority than class-j, the policy π is allowed to accept this class-j customer in period t only if when there is no class-i customer rejected by π before period t, and π is allowed to reject this class-j customer in period t only if when there is no class-k customer accepted by π before period t.

This notion of fairness asserts that when π accepts a customer, π needs to accept all the customers with higher priorities; when π rejects a customer, π needs to reject all the customers with lower priorities. In many practical settings, not enforcing strict fairness may adversely affect customer loyalty to the firm. In the example of vacation timeshare management mentioned in §1, Hilton Grand Vacations Club offers timeshares at four different levels, namely, platinum, gold, silver, bronze [11]. During a particular selling season, the management will not sell a home resort (e.g., a room in Elara on the Las Vegas strip) to a gold customer if it has previously rejected a platinum customer. Likewise, the management will not reject a gold customer if it has previously accepted a silver customer. The management has the incentives to enforce such fairness, because of the extensive interactions between timeshare users in online forums [52].

5.1. Extended Definitions of Regret

We extend the definitions of $RA_t^P(\alpha_t^P)$ and $RR_t^P(\alpha_t^P)$ defined in §4. For any feasible policy π, there can be two scenarios when a random customer arrives in period t.

(a) Active decision making: π needs to decide whether to accept or reject this customer;

(b) Passive decision making: π does not need to make an active decision, if this arriving customer is "automatically" accepted or rejected due to either fairness or stock-out.

In the former scenario (a), the active decision made by π incurs a regret of acceptance or rejection, whereas in the latter scenario (b), there is no regret incurred in period t since π does not make any decisions. For any feasible policy π, to define our regrets $RA_t^\pi(\alpha_t^\pi)$ and $RR_t^\pi(\alpha_t^\pi)$, we use

$$W(\pi^*; f_T, \alpha_0^\pi, \ldots, \alpha_t^\pi)$$

to denote the *modified* clairvoyant optimal revenue given a fixed sample path f_T *and* fixed decisions $\alpha_0^\pi, \ldots, \alpha_t^\pi$ in the first t periods. Intuitively, the *modified* clairvoyant optimal policy π^* takes the first t-period (potentially sub-optimal) decisions as given, and generates the highest revenue over the remaining periods $[t, T]$ along the sample path f_T. With this modified definition, we define the two regrets below. Fix a sample path f_T and examine any period $t = 1, \ldots, T$.

(a) If π accepts the incoming customer in period t, then the regret of acceptance is defined by

$$RA_t^\pi(\alpha_t^\pi = 1) \mid f_T = W(\pi^*; f_T, \alpha_0^\pi, \ldots, \alpha_{t-1}^\pi) - W(\pi^*; f_T, \alpha_0^\pi, \ldots, \alpha_{t-1}^\pi, 1). \tag{21}$$

The underlying idea is simple. The regret of acceptance in period t is exactly the difference between two revenues, one resulted from taking the "optimal" actions in hindsight from period t onwards, and the other one resulted from taking an acceptance decision in period t and then taking the "optimal" actions in hindsight from period $t+1$ onwards.

(b) If π rejects the incoming customer in period t, then the regret of acceptance is defined by

$$RR_t^\pi(\alpha_t^\pi = 0) \mid f_T = W(\pi^*; f_T, \alpha_0^\pi, \ldots, \alpha_{t-1}^\pi) - W(\pi^*; f_T, \alpha_0^\pi, \ldots, \alpha_{t-1}^\pi, 0). \tag{22}$$

The idea is similar by merely taking a dual view of (a).

5.2. Extended Regret-Parity Policy

With these extended definitions (21–22), we re-define the regret-parity policy $\tilde{\pi}$ below. Since $\tilde{\pi}$ is randomized, we use the expanded information set f_t^+ that not only includes the original information set f_t but also all the randomized decisions of $\tilde{\pi}$ up to period $t-1$.

Suppose a customer arrives in period t and $\tilde{\pi}$ needs to make an *active* decision. Then $\tilde{\pi}$ will accept her with probability θ_t and reject with probability $1-\theta_t$, i.e., $\alpha_t^{\tilde{\pi}}=1$ with probability θ_t and $\alpha_t^{\tilde{\pi}}=0$ with probability $1-\theta_t$, where θ_t is computed by solving

$$\theta_t \cdot \mathbb{E}[RA_t^{\tilde{\pi}}(1) \mid f_t^+] = (1-\theta_t) \cdot \mathbb{E}[RR_t^{\tilde{\pi}}(0) \mid f_t^+].$$

5.3. Performance Analysis

At any customer arrival time t (post-decision), we keep track of three critical numbers associated with any feasible policy π, namely, γ_t^π being the lowest revenue class that π has accepted, β_t^π being the highest revenue class that π has rejected (not due to stock-out), and the ending inventory level Y_t^π. As a convention, we initialize $\gamma_0^\pi = 0$ and $\beta_0^\pi = I+1$.

Lemma 5. *At any customer arrival time t where $Y_t^{\tilde{\pi}} > 0$, we have*
(a) If $Y_t^{\pi^o} < Y_t^{\tilde{\pi}}$, then $\gamma_t^{\tilde{\pi}} \le \beta_t^{\tilde{\pi}} \le \gamma_t^{\pi^o} \le \beta_t^{\pi^o}$.
(b) If $Y_t^{\pi^o} > Y_t^{\tilde{\pi}}$, then $\gamma_t^{\pi^o} \le \beta_t^{\pi^o} \le \gamma_t^{\tilde{\pi}} \le \beta_t^{\tilde{\pi}}$.
(c) If $Y_t^{\pi^o} = Y_t^{\tilde{\pi}}$, then $\gamma_t^{\tilde{\pi}} = \gamma_t^{\pi^o} \le \beta_t^{\tilde{\pi}} = \beta_t^{\pi^o}$.

At a high-level, with the fairness constraint, we can clearly keep track of the accept/reject status of each class of two different policies by merely comparing their aggregate ending inventory levels. In the absence of fairness, these relationships in Lemma 5 will not hold, thereby making the cost comparison between two different policies very challenging. With aid of Lemma 5, we can prove the following result, similar to Lemmas 1 and 2 in the two-class case.

Lemma 6. *Along every sample path f_T^{++}, we have*

$$\sum_{t=1}^{T} RA_t^{\pi^o}(\alpha_t^{\pi^o}) \ge \sum_{t\in\mathscr{T}_a} RA_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}), \qquad \sum_{t=1}^{T} RR_t^{\pi^o}(\alpha_t^{\pi^o}) \ge \sum_{t\in\mathscr{T}_b} RR_t^{\tilde{\pi}}(\alpha_t^{\tilde{\pi}}).$$

With identical arguments, Lemmas 3 and 4 hold for the multi-class setting with fairness as well. Combining these results, we have the following theorem.

Theorem 2. *For each instance of the multi-class admission-control policy based revenue management problem under fairness, the expected total regret incurred by the regret-parity policy $\tilde{\pi}$ is at most two times the expected total regret incurred by the optimal policy π^o, i.e.,*

$$\mathbb{E}\left[\mathcal{R}(\tilde{\pi})\right] \le 2 \cdot \mathbb{E}\left[\mathcal{R}(\pi^o)\right].$$

6. Numerical Experiments

To test the empirical performances of our proposed policy $\tilde{\pi}$, we conduct an extensive numerical study and report our numerical results.

6.1. Design of Experiments

We set the discrete time horizon $T = 50$ periods. We normalize the revenue rate of Class-1 customers $r_1 = 100$ and vary the revenue rate of Class-2 customers $r_2 \in \{20, 30, 40, \ldots, 80\}$. We consider two types of demand processes described as follows.

(a) I.I.D. demands: In the i.i.d. demand setting, in each period $t = 1, \ldots, T$, we test a range of arrival probabilities $p^1, p^2 \in \{0.2, 0.25, 0.3, 0.35, 0.4\}$. The probability of having no arrivals is then $1 - p^1 - p^2$. In addition, we set the initial inventory $M = T(p^1 + \kappa \cdot p^2)$, where the initial inventory M is controlled by $\kappa \in \{-0.2, 0, 0.2\}$.

(b) Correlated demands: We also consider a correlated demand setting where the instantaneous rates are time-varying and correlated, which are modulated by an exogenous Markov chain. In this Markov modulated demand setting, we keep the choices of parameters T, r_1, r_2, p_1 p_2, and M the same as in the i.i.d. demand setting. In addition, we introduce three states of economy, namely, good (denoted by state 1), fair (denoted by state 2), poor (denoted by state 3), and the arrival rate is affected by the state of economy. We set the initial state to be state 2. The state transition is modulated by an exogenous Markov chain. More specifically, the arrival probability depends on p_1, p_2 and the state of economy. Let $p^i(j)$ denote the arrival probability for class-i customer when the state of economy is j, where $i = 1, 2$ and $j = 1, 2, 3$. We set

$$\begin{array}{ll} p^1(1) = 1.5p^1, & p^2(1) = 0.5p^2, \\ p^1(2) = p^1, & p^2(2) = p^2, \\ p^1(3) = 0.5p^1, & p^2(3) = 1.5p^2. \end{array} .$$

The above construction captures the fact that customers will buy higher-class (lower-class) products with a higher probability when the state of economy is better (poorer). We consider two transition probability matrices for the exogenous Markov chain whose states are the three states of economy:

$$P_1 = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.6 & 0.3 & 0.1 \end{bmatrix}.$$

The above two transition probability matrices P_1 and P_2 represent positively-correlated and negatively-correlated demands, respectively.

6.2. Performance Measure and Benchmark Policies

The two standard performance measures, the regret and revenue errors of $\tilde{\pi}$, are defined by

$$\epsilon^{\text{regret}} = \left[\frac{\mathbb{E}\left[\mathcal{R}(\tilde{\pi})\right]}{\mathbb{E}\left[\mathcal{R}(\pi^o)\right]} - 1 \right] \times 100\%, \tag{23}$$

$$\epsilon^{\text{revenue}} = \left[1 - \frac{\mathbb{E}\left[W(\tilde{\pi})\right]}{\mathbb{E}\left[W(\pi^{o})\right]}\right] \times 100\%. \tag{24}$$

Besides the above standard performance measures, we also compare our policy with the policies proposed in Ball and Queyranne [1]. In their paper, the authors presented two robust optimization algorithms that can be applied in our setting. Both policies having the form of setting a threshold level for the maximum amount of Class-2 customer to be admitted.

Benchmark Algorithm 1: The first algorithm is a static policy that have a fixed threshold level as in their Equation (4). The threshold θ is set as either $M - \left\lfloor \frac{M}{2-\frac{r_2}{r_1}} \right\rfloor$ or $M - \left\lceil \frac{M}{2-\frac{r_2}{r_1}} \right\rceil$ depending on which number gives the best competitive ratio in their robust optimization problem.

Benchmark Algorithm 2: The second algorithm is a dynamic policy that keeps updating the threshold level throughout the planning horizon. In each period t, let h' denote the total number of Class-1 customers accepted. Denote $\gamma = \frac{h'}{M}$ and $\alpha = \frac{\gamma r_1 + (1-\gamma r_2)}{r_2}$. Then the threshold level in each period t is defined as $\theta_t = \frac{1-\frac{\gamma r_1}{\alpha r_2}}{1+\frac{1}{\alpha}-\frac{r_2}{r_1}}$.

For every instance of the problem, we denote the better expected revenue of the two robust benchmark algorithms by $W(\pi^{\text{robust}})$. Then the relative gain in expected revenue by using our policy $\tilde{\pi}$ is defined as

$$\eta^{\text{gain}} = \left[\frac{\mathbb{E}\left[W(\tilde{\pi})\right]}{\mathbb{E}\left[W(\pi^{\text{robust}})\right]} - 1\right] \times 100\%. \tag{25}$$

We note that their robust algorithms are established in the online adversarial setting (that does not require any future demand information as an input). Our algorithm, on the other hand, does require evolving conditional future demand information, as time progresses. Hence the numerical comparison between our policy and theirs is not entirely fair. Nevertheless, in the absence of better alternatives, we adapt their algorithms to our setting and compare the numerical performances.

As seen from Tables 1–3, our proposed regret-parity policy $\tilde{\pi}$ performs consistently well in term of expected revenue and regret, compared to the optimal policy. Compared to the robust benchmark algorithms proposed in Ball and Queyranne [1], we gain around 16% more expected revenue, which is quite significant. Also, it is interesting to observe that this ratio η^{gain} is higher when the starting inventory is smaller (i.e., κ is smaller), and when r_2 is smaller. This is because when there is less starting inventory, the optimal policy shall reject almost all the Class-2 customers, but the robust benchmark algorithm always accepts some Class-2 customer as long as the threshold level has not been reached. And when r_2 is small, this loss becomes more significant.

7. Conclusion

In this paper, we have studied a class of revenue management problems with non-homogeneous Poisson customer arrival processes. We have proposed a new regret minimization framework and

Table 1 Performance of $\tilde{\pi}$ under i.i.d. demands

	κ	-0.2			0			0.2		
	r_2	min	mean	max	min	mean	max	min	mean	max
ϵ^{regret}		39.0%	42.2%	45.6%	31.8%	35.1%	40.2%	24.4%	28.8%	34.9%
$\epsilon^{\text{revenue}}$	20	0.03%	0.11%	0.26%	0.11%	0.25%	0.45%	0.17%	0.36%	0.62%
η^{gain}		17.5%	30.7%	54.5%	13.5%	22.0%	34.2%	8.5%	11.6%	14.1%
ϵ^{regret}		28.5%	34.9%	38.9%	23.8%	29.2%	34.3%	19.1%	24.2%	29.5%
$\epsilon^{\text{revenue}}$	30	0.03%	0.12%	0.29%	0.12%	0.27%	0.50%	0.18%	0.37%	0.66%
η^{gain}		17.4%	30.1%	48.4%	13.4%	21.9%	32.2%	8.6%	12.1%	14.6%
ϵ^{regret}		23.9%	30.5%	34.7%	20.0%	25.7%	30.4%	17.0%	21.5%	26.0%
$\epsilon^{\text{revenue}}$	40	0.04%	0.13%	0.30%	0.12%	0.28%	0.51%	0.18%	0.37%	0.66%
η^{gain}		15.7%	27.1%	44.1%	12.4%	20.4%	31.0%	8.0%	11.6%	14.4%
ϵ^{regret}		19.2%	27.4%	30.5%	16.2%	23.4%	26.5%	13.9%	19.9%	22.6%
$\epsilon^{\text{revenue}}$	50	0.03%	0.13%	0.31%	0.12%	0.27%	0.50%	0.18%	0.36%	0.63%
η^{gain}		13.5%	23.5%	38.7%	10.7%	17.9%	26.9%	7.0%	10.4%	13.1%
ϵ^{regret}		20.2%	25.9%	28.8%	18.0%	22.5%	25.3%	16.4%	19.7%	21.8%
$\epsilon^{\text{revenue}}$	60	0.03%	0.12%	0.29%	0.11%	0.26%	0.47%	0.16%	0.34%	0.58%
η^{gain}		11.0%	18.9%	28.6%	8.6%	14.4%	20.9%	5.6%	8.7%	11.3%
ϵ^{regret}		21.1%	25.5%	27.4%	19.3%	22.5%	23.9%	17.9%	20.4%	21.7%
$\epsilon^{\text{revenue}}$	70	0.03%	0.11%	0.27%	0.09%	0.24%	0.43%	0.14%	0.31%	0.53%
η^{gain}		8.3%	14.2%	21.5%	6.4%	11.0%	16.4%	4.1%	6.6%	8.6%
ϵ^{regret}		23.8%	25.6%	27.6%	21.7%	23.4%	25.0%	20.8%	22.0%	23.6%
$\epsilon^{\text{revenue}}$	80	0.03%	0.10%	0.23%	0.09%	0.21%	0.36%	0.14%	0.27%	0.45%
η^{gain}		5.4%	9.5%	14.9%	4.1%	7.2%	10.6%	2.5%	4.4%	6.0%

Table 2 Performance of $\tilde{\pi}$ under (positively-correlated) Markov modulated demands

	κ	-0.2			0			0.2		
	r_2	min	mean	max	min	mean	max	min	mean	max
ϵ^{regret}		24.3%	49.2%	83.4%	24.7%	36.3%	83.2%	23.5%	31.5%	46.5%
$\epsilon^{\text{revenue}}$	20	0.04%	0.13%	0.27%	0.12%	0.27%	0.46%	0.19%	0.38%	0.63%
η^{gain}		17.0%	30.3%	53.8%	13.0%	21.6%	33.6%	8.4%	11.6%	14.2%
ϵ^{regret}		20.6%	36.9%	69.7%	20.1%	30.4%	50.4%	18.4%	24.8%	33.5%
$\epsilon^{\text{revenue}}$	30	0.04%	0.14%	0.31%	0.13%	0.28%	0.50%	0.20%	0.38%	0.67%
η^{gain}		16.8%	29.7%	48.1%	12.9%	21.5%	31.8%	8.4%	12.0%	14.6%
ϵ^{regret}		23.4%	30.0%	38.9%	19.2%	25.2%	33.5%	16.0%	20.7%	24.9%
$\epsilon^{\text{revenue}}$	40	0.05%	0.14%	0.31%	0.13%	0.28%	0.51%	0.19%	0.37%	0.65%
η^{gain}		15.2%	26.8%	43.8%	11.9%	20.0%	30.6%	7.8%	11.5%	14.4%
ϵ^{regret}		17.3%	26.3%	34.8%	17.4%	23.0%	27.3%	13.5%	19.5%	22.4%
$\epsilon^{\text{revenue}}$	50	0.04%	0.14%	0.31%	0.10%	0.25%	0.46%	0.18%	0.35%	0.62%
η^{gain}		13.1%	23.2%	38.3%	10.3%	17.5%	26.5%	6.8%	10.3%	13.1%
ϵ^{regret}		20.3%	24.3%	28.2%	17.2%	21.3%	25.4%	16.5%	19.2%	21.5%
$\epsilon^{\text{revenue}}$	60	0.03%	0.13%	0.30%	0.10%	0.25%	0.46%	0.15%	0.33%	0.56%
η^{gain}		10.6%	18.7%	28.4%	8.3%	14.1%	20.6%	5.4%	8.6%	11.2%
ϵ^{regret}		20.6%	24.1%	27.6%	19.5%	21.3%	22.9%	16.8%	19.3%	21.2%
$\epsilon^{\text{revenue}}$	70	0.03%	0.12%	0.26%	0.09%	0.23%	0.41%	0.13%	0.30%	0.51%
η^{gain}		8.0%	14.0%	21.4%	6.2%	10.8%	16.2%	3.9%	6.6%	8.5%
ϵ^{regret}		21.9%	24.1%	29.4%	19.7%	22.0%	24.9%	19.0%	20.9%	22.3%
$\epsilon^{\text{revenue}}$	80	0.03%	0.10%	0.23%	0.09%	0.20%	0.35%	0.13%	0.26%	0.44%
η^{gain}		5.2%	9.4%	14.8%	3.9%	7.1%	10.5%	2.4%	4.3%	5.9%

proceeded using the lens of approximation algorithms to devise a conceptually simple and provably-good regret-parity policy. We have made some important progress towards better understanding the intricate link between stochastic regret minimization and approximation algorithms in the realm

Table 3 Performance of $\tilde{\pi}$ under (negatively-correlated) Markov modulated demands

	κ	-0.2			0			0.2		
	r_2	min	mean	max	min	mean	max	min	mean	max
ϵ^{regret}	20	28.5%	46.3%	71.0%	24.3%	39.1%	54.0%	25.2%	31.5%	39.6%
$\epsilon^{\text{revenue}}$		0.03%	0.12%	0.27%	0.13%	0.27%	0.47%	0.20%	0.39%	0.64%
η^{gain}		17.7%	30.9%	54.7%	13.7%	22.1%	34.4%	8.5%	11.6%	14.1%
ϵ^{regret}	30	27.1%	42.8%	66.8%	21.9%	32.9%	42.6%	17.9%	27.7%	34.7%
$\epsilon^{\text{revenue}}$		0.04%	0.13%	0.30%	0.14%	0.29%	0.49%	0.21%	0.40%	0.65%
η^{gain}		17.5%	30.2%	48.6%	13.5%	22.1%	32.4%	8.6%	12.1%	14.6%
ϵ^{regret}	40	23.3%	34.8%	61.4%	16.7%	27.6%	38.8%	16.7%	22.8%	28.5%
$\epsilon^{\text{revenue}}$		0.04%	0.13%	0.31%	0.14%	0.29%	0.52%	0.20%	0.38%	0.65%
η^{gain}		15.9%	27.2%	44.1%	12.6%	20.5%	31.1%	8.1%	11.6%	14.4%
ϵ^{regret}	50	19.1%	28.9%	34.2%	16.9%	24.8%	30.4%	15.8%	20.7%	27.6%
$\epsilon^{\text{revenue}}$		0.03%	0.13%	0.30%	0.13%	0.27%	0.48%	0.20%	0.36%	0.62%
η^{gain}		13.6%	23.6%	38.8%	10.8%	18.0%	27.1%	7.1%	10.4%	13.2%
ϵ^{regret}	60	19.7%	26.3%	32.3%	20.1%	23.5%	27.7%	16.9%	20.0%	22.1%
$\epsilon^{\text{revenue}}$		0.03%	0.12%	0.29%	0.12%	0.26%	0.47%	0.17%	0.33%	0.58%
η^{gain}		11.1%	19.0%	28.7%	8.8%	14.5%	21.0%	5.7%	8.7%	11.3%
ϵ^{regret}	70	22.3%	26.2%	29.5%	17.4%	22.1%	24.5%	18.6%	20.4%	22.6%
$\epsilon^{\text{revenue}}$		0.03%	0.11%	0.26%	0.09%	0.23%	0.42%	0.14%	0.30%	0.52%
η^{gain}		8.4%	14.3%	21.5%	6.5%	11.1%	16.5%	4.2%	6.7%	8.6%
ϵ^{regret}	80	23.4%	25.3%	32.8%	21.7%	23.0%	25.0%	20.1%	21.5%	25.3%
$\epsilon^{\text{revenue}}$		0.02%	0.09%	0.21%	0.09%	0.20%	0.36%	0.13%	0.26%	0.44%
η^{gain}		5.5%	9.6%	15.0%	4.2%	7.3%	10.7%	2.6%	4.4%	6.0%

of revenue management and dynamic resource allocation. We believe combining the ideas from approximation algorithms with this new regret minimization framework can yield many fruitful results and discussions in many other core resource allocation or revenue management problems.

To close this paper, we would like to point out two immediate and plausible future research directions as follows. (a) One may wish to waive the fairness requirement in the multi-class setting. (b) One can also consider a pricing version of the same problem, in which the firm can dynamically update their prices. However, developing worst-case performance guarantees for the aforementioned directions remains challenging and would require new ideas and methods to be developed.

Acknowledgment

The authors thank the Editor-in-Chief Professor Awi Federgruen, the anonymous associate editor, and the three anonymous referees for their very constructive comments and suggestions, which helped significantly improve both the content and the exposition of this paper. An earlier version has been the finalist of the 2014 INFORMS Undergraduate Operations Research Prize while Chao Qin and Cheng Hua were in the IOE undergraduate program at the University of Michigan. This research is partially supported by NSF grants CMMI-1362619, CMMI-1451078 and CMMI-1634505.

References

[1] Ball, M., M. Queyranne. 2009. Towards robust revenue management: Competitive analysis of online booking. *Operations Research* **57**(4) 950–963.

[2] Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations Research Letters* **25**(1) 1–13.

[3] Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

[4] Birbil, S. I., J. B. G. Frenk, J. A. S. Gromicho, S. Zhang. 2009. The role of robust optimization in single-leg airline revenue management. *Management Science* **55**(1) 148–163.

[5] Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–229.

[6] Broder, J., P. Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* **60**(4) 965–980.

[7] Cao, P., J. Li, H. Yan. 2012. Optimal dynamic pricing of inventories with stochastic demand and discounted criterion. *European Journal of Operational Research* **217**(3) 580 – 588.

[8] Chan, C. W., V. F. Farias. 2009. Stochastic depletion problems: Effective myopic policies for a class of dynamic optimization problems. *Mathematics of Operations Research* **34**(2) 333–350.

[9] Chao, X., X. Gong, C. Shi, H. Zhang. 2015. Approximation algorithms for perishable inventory systems. *Operations Research* **63**(3) 585–601.

[10] Chen, Y., V. F. Farias. 2013. Simple policies for dynamic pricing with imperfect forecasts. *Operations Research* **61**(3) 612–624.

[11] Club, Hilton Grand Vacations. 2016. Hilton grand vacations club reservation options. http://www.hgvclubprogram.com/club-features/reservations/.

[12] Coffman Jr., E. G., J. Csirik, G. Galambos, S. Martello, D. Vigo. 2013. Bin packing approximation algorithms: Survey and classification. Panos M. Pardalos, Ding-Zhu Du, Ronald L. Graham, eds., *Handbook of Combinatorial Optimization*. Springer New York, 455–531.

[13] Dean, B. C., M. X. Goemans, J. Vondrák. 2008. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research* **33**(4) 945–964.

[14] den Boer, A. V., B. Zwart. 2014. Simultaneously learning and optimizing using controlled variance pricing. *Management Science* **60**(3) 770–783.

[15] Dragos, F. C., V. F. Farias. 2012. Model predictive control for dynamic resource allocation. *Mathematics of Operations Research* **37**(3) 501–525.

[16] Elmachtoub, A. N., R. Levi. 2014. From cost sharing mechanisms to online selection problems. *Mathematics of Operations Research* **40**(3) 542–557.

[17] Elmachtoub, A. N., R. Levi. 2016. Supply chain management with online customer selection. *Operations Research* **64**(2) 458–473.

[18] Gallego, G., Ö. Özer. 2001. Integrating replenishment decisions with advance demand information. *Management Science* **47**(10) 1344–1360.

[19] Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* **40**(8) 999–1020.

[20] Gallego, G., G. van Ryzin. 1997. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research* **45**(1) 24–41.

[21] Geng, X., W. T. Huh, M. Nagarajan. 2014. Sequential resource allocation with constraints: Two-customer case. *Operations Research Letters* **42**(1) 70 – 75.

[22] George, J. M., J. M. Harrison. 2001. Dynamic control of a queue with adjustable service rate. *Operations Research* **49**(5) pp. 720–731.

[23] Graves, S., H. Meal, S. Dasu, Y. Qin. 1986. Two-stage production planning in a dynamic environment. *S. Axsater, C. Schneeweiss, E. Silver, eds. Multi-Stage Production Planning and Control. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, Germany* 9–43.

[24] Green, L. V., P. J. Kolesar. 1995. On the accuracy of the simple peak hour approximation for Markovian queues. *Management Science* **41**(8) pp. 1353–1370.

[25] Guha, S., K. Munagala. 2014. Stochastic regret minimization via thompson sampling. *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*. 317–338.

[26] Gupta, V., M. Harchol-Balter, A. S. Wolf, U. Yechiali. 2006. Fundamental characteristics of queues with fluctuating load. *SIGMETRICS Performance Evaluation Review* **34**(1) 203–215.

[27] Harrison, J. M., N. B. Keskin, A. Zeevi. 2012. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science* **58**(3) 570–586.

[28] Heath, D. C., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution system. *IIE Transactions* **26**(3) 17–30.

[29] Iyengar, G., K. Sigman. 2004. Exponential penalty function control of loss networks. *Annals of Applied Probability* **14**(4) 1698–1740.

[30] Kelly, F. P. 1991. Effective bandwidths at multi-class queues. *Queueing Systems* **9**(1-2) 5–16.

[31] Key, P. 1990. Optimal control and trunk reservation in loss networks. *Probability in the Engineering and Informational Sciences* **4** 203–242.

[32] Kleywegt, A. J., J. D. Papastavrou. 1998. The dynamic and stochastic knapsack problem. *Operations Research* **46**(1) 17–35.

[33] Kleywegt, A. J., J. D. Papastavrou. 2001. The dynamic and stochastic knapsack problem with random sized items. *Operations Research* **49**(1) 26–41.

[34] Kumar, R., M. E. Lewis, H. Topaloglu. 2013. Dynamic service rate control for a single-server queue with Markov-modulated arrivals. *Naval Research Logistics* **60**(8) 661–677.

[35] Lan, Y., M. O. Ball, I. Z. Karaesmen. 2011. Regret in overbooking and fare-class allocation for single leg. *Manufacturing & Service Operations Management* **13**(2) 194–208.

[36] Levi, R., G. Janakiraman, M. Nagarajan. 2008. A 2-approximation algorithm for stochastic inventory control models with lost-sales. *Mathematics of Operations Research* **33**(2) 351–374.

[37] Levi, R., M. Pál, R. O. Roundy, D. B. Shmoys. 2007. Approximation algorithms for stochastic inventory control models. *Mathematics of Operations Research* **32**(2) 284–302.

[38] Levi, R., A. Radovanović. 2010. Technical note: Provably near-optimal LP-based policies for revenue management in systems with reusable resources. *Operations Research* **58**(2) 503–507.

[39] Levi, R., R. O. Roundy, D. B. Shmoys, V. A. Truong. 2008. Approximation algorithms for capacitated stochastic inventory models. *Operations Research* **56**(5) 1184–1199.

[40] Levi, R., C. Shi. 2013. Approximation algorithms for the stochastic lot-sizing problem with order lead times. *Operations Research* **61**(3) 593–602.

[41] Lueker, G. S. 1995. Average-case analysis of off-line and on-line knapsack problems. *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '95, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 179–188.

[42] Massey, W., W. Whitt. 1997. Peak congestion in multi-server service systems with slowly varying arrival rates. *Queueing Systems* **25**(1-4) 157–172.

[43] Mills, T. C. 1991. *Time Series Techniques for Economists*. Cambridge University Press, Cambridge, UK.

[44] Netessine, S. 2006. Dynamic pricing of inventory/capacity with infrequent price changes. *European Journal of Operational Research* **174**(1) 553 – 580.

[45] Papastavrou, J. D., S. Rajagopalan, A. J. Kleywegt. 1996. The dynamic and stochastic knapsack problem with deadlines. *Management Science* **42**(12) pp. 1706–1718.

[46] Perakis, G., G. Roels. 2010. Robust controls for network revenue management. *Manufacturing & Service Operations Management* **12**(1) 56–76.

[47] Powell, W. B. 2011. *Approximate dynamic programming solving the curse of dimensionality, 2nd ed.*. Wiley, Hoboken, NJ.

[48] Prabhu, N. U., Y. Zhu. 1989. Markov-modulated queueing systems. *Queueing Systems* **5**(1-3) 215–245.

[49] Rusmevichientong, P., H. Topaloglu. 2012. Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations Research* **60**(4) 865–882.

[50] Shi, C., H. Zhang, X. Chao, R. Levi. 2014. Approximation algorithms for capacitated stochastic inventory systems with setup costs. *Naval Research Logistics* **61**(4) 304–319.

[51] Talluri, K. T., G. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Springer, New York.

[52] TugBBS. 2016. Timeshare users group online discussion forums. http://tugbbs.com/forums/.

[53] Van Hentenryck, P., R. Bent, L. Mercier, Y. Vergados. 2009. Online stochastic reservation systems. *Annals of Operations Research* **171**(1) 101–126.

[54] Wang, Z., S. Deng, Y. Ye. 2014. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research* **62**(2) 318–331.

[55] Yoon, S., M. Lewis. 2004. Optimal pricing and admission control in a queueing system with periodically varying parameters. *Queueing Systems* **47**(3) 177–199.

[56] Zhang, H., C. Shi, X. Chao. 2016. Approximation algorithms for perishable inventory systems with setup costs. *Operations Research* **64**(2) 432–440.

[57] Zhao, W., Y. Zheng. 2000. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Science* **46**(3) 375–388.

Appendix: Omitted Technical Proofs

Proof of Proposition 1. We can write

$$W(\pi^*; f_T) = r_1 C^1_{\pi^*} + r_2 C^2_{\pi^*}, \qquad W(\pi; f_T) = r_1 C^1_{\pi} + r_2 C^2_{\pi}.$$

Since the firm accepts Class-1 customers as long as the inventory is positive, we have $C^1_{\pi^*} = \min\left(A^1_{[1,T]}, M\right)$. Moreover, if the number of Class-2 customers accepted by π is greater than that accepted by π^*, then the number of Class-1 customers accepted by π will decrease by $(C^2_\pi - C^2_{\pi^*})$. Otherwise, C^1_π will be equal to $C^1_{\pi^*}$. Thus, combining the two cases above, we have $C^1_\pi = \min\{C^1_{\pi^*}, C^1_{\pi^*} - (C^2_\pi - C^2_{\pi^*})\}$. Hence, by (2) and some simple algebra, we have

$$\begin{aligned}\mathcal{R}(\pi; f_T) &= W(\pi^*; f_T) - W(\pi; f_T) \\ &= r_1 C^1_{\pi^*} + r_2 C^2_{\pi^*} - r_1 C^1_{\pi} - r_2 C^2_{\pi}\end{aligned}$$

$$
\begin{aligned}
&= r_1(C^1_{\pi^*} - C^1_{\pi}) + r_2(C^2_{\pi^*} - C^2_{\pi}) \\
&= r_1\left(C^1_{\pi^*} - \min\{C^1_{\pi^*}, C^1_{\pi^*} - (C^2_{\pi} - C^2_{\pi^*})\}\right) + r_2(C^2_{\pi^*} - C^2_{\pi}) \\
&= r_1\left(C^1_{\pi^*} + \max\{-C^1_{\pi^*}, -C^1_{\pi^*} + (C^2_{\pi} - C^2_{\pi^*})\}\right) + r_2(C^2_{\pi^*} - C^2_{\pi}) \\
&= r_1(C^2_{\pi} - C^2_{\pi^*})^+ + r_2(C^2_{\pi^*} - C^2_{\pi}).
\end{aligned}
$$

This completes the proof. **Q.E.D.**

Proof of Proposition 2. We fix an arbitrary sample path f_T. Suppose there are l customers arrived at the system, and let $1 \le t_1 \le \ldots \le t_l \le T$ denote all these l customer arriving epochs. We then denote t_s to be the last customer arriving epoch in which the firm is not out of stock. Using (5) and (6), we sum up the two regrets (by decisions) along the sample path f_T, and obtain

$$
\begin{aligned}
&\sum_{t=1}^{T} [RA^{\pi}_t(\alpha^{\pi}_t) + RR^{\pi}_t(\alpha^{\pi}_t)] \,\Big|\, f_T \qquad (26)\\
&= \sum_{k=1}^{s} (r_1 - r_2)\left[\left(a^1_{(t_k,T]} - y^{\pi}_{t_k}\right)^+ - \left(a^1_{[t_k,T]} - x^{\pi}_{t_k}\right)^+\right] + \sum_{k=1}^{s} r_2 \left[\left(y^{\pi}_{t_k} - a^{1,2}_{(t_k,T]}\right)^+ - \left(x^{\pi}_{t_k} - a^{1,2}_{[t_k,T]}\right)^+\right] \\
&= (r_1 - r_2)\left[\left(a^1_{(t_s,T]} - y^{\pi}_{t_s}\right)^+ - \left(a^1_{[t_1,T]} - x^{\pi}_{t_1}\right)^+\right] + r_2\left[\left(y^{\pi}_{t_s} - a^{1,2}_{(t_s,T]}\right)^+ - \left(x^{\pi}_{t_1} - a^{1,2}_{[t_1,T]}\right)^+\right] \\
&= (r_1 - r_2)\left[\left(a^1_{(t_s,T]} - y^{\pi}_{t_s}\right)^+ - \left(a^1_{[1,T]} - M\right)^+\right] + r_2\left[\left(y^{\pi}_{t_s} - a^{1,2}_{(t_s,T]}\right)^+ - (M - l)^+\right],
\end{aligned}
$$

where the first equality holds because when the inventory runs out, i.e., $k > s$, both regrets $RA^{\pi}_{t_k}(\alpha^{\pi}_{t_k} = 0) = RR^{\pi}_{t_k}(\alpha^{\pi}_{t_k} = 0) = 0$, and the second equality holds since for each $k = 1, \ldots, s-1$,

$$
x^{\pi}_{t_{k+1}} = y^{\pi}_{t_k}, \quad a^1_{[t_{k+1},T]} = a^1_{(t_k,T]}, \quad \text{and} \quad a^{1,2}_{[t_{k+1},T]} = a^{1,2}_{(t_k,T]}.
$$

There are two cases as follows.

Case 1. There is some inventory left in the end of the horizon, i.e., $y^{\pi}_T > 0$. This implies that $t_s = t_l$ and $y^{\pi}_{t_s} = y^{\pi}_T > 0$. Moreover, the total number of Class-1 customers is less than the initial inventory, i.e., $a^1_{[1,T]} < M$. Hence, (26) becomes

$$
\sum_{t=1}^{T} [RA^{\pi}_t(\alpha^{\pi}_t) + RR^{\pi}_t(\alpha^{\pi}_t)] \,\Big|\, f_T = r_2\left(y^{\pi}_T - (M - l)^+\right).
$$

Moreover, because $C^2_{\pi} \le C^2_{\pi^*}$, (4) becomes $\mathcal{R}(\pi) = r_2(C^2_{\pi^*} - C^2_{\pi})$. Therefore, it suffices to show that

$$
y^{\pi}_T - (M - l)^+ = C^2_{\pi^*} - C^2_{\pi}. \qquad (27)
$$

We know that $C^1_\pi = C^1_{\pi^*} = A^1_{[1,T]}$ in this case, and hence $y^\pi_T = M - C^1_\pi - C^2_\pi = M - C^1_{\pi^*} - C^2_\pi$. Then (27) becomes $M - (M-l)^+ = C^1_{\pi^*} + C^2_{\pi^*}$, which is valid since $C^1_{\pi^*} + C^2_{\pi^*} = \min(M, l)$.

Case 2. All the inventory units are used up at the end of the horizon. This implies that $y^\pi_{t_s} = 0$, and $M \le l$. Hence, (26) becomes

$$\sum_{t=1}^{T} \left[RA^\pi_t(\alpha^\pi_t) + RR^\pi_t(\alpha^\pi_t) \right] \Big| f_T = (r_1 - r_2)\left(a^1_{(t_s,T]} - (a^1_{[1,T]} - M)^+ \right).$$

Moreover, because $C^2_\pi \ge C^2_{\pi^*}$, (4) becomes $\mathcal{R}(\pi) = (r_1 - r_2)(C^2_\pi - C^2_{\pi^*})$. Therefore, it suffices to argue

$$a^1_{(t_s,T]} - (a^1_{[1,T]} - M)^+ = C^2_\pi - C^2_{\pi^*}. \tag{28}$$

We know that $a^1_{[1,t_s]} + C^2_\pi = C^1_\pi + C^2_\pi = M$ in this case, and hence $C^2_\pi = M - a^1_{[1,t_s]}$. Then (28) becomes $a^1_{(t_s,T]} - (a^1_{[1,T]} - M)^+ = M - a^1_{[1,t_s]} - C^2_{\pi^*}$, which is valid due to the fact that if $a^1_{[1,T]} \ge M$, then $C^2_{\pi^*} = 0$, and otherwise, $C^2_{\pi^*} = M - a^1_{[1,T]}$.

Combining the above cases, the two regret accounting schemes are indeed equivalent. **Q.E.D.**

Proof of Lemma 5. We fix a sample path f^{++}_T. It is clear that $\gamma^{\tilde\pi}_t \le \beta^{\tilde\pi}_t$ and $\gamma^{\pi^o}_t \le \beta^{\pi^o}_t$ by Definition 1 of our fairness constraint. In Case (a), we can always find a customer accepted by π^o but rejected by $\tilde\pi$. Denote the class of this particular customer by κ and we can see that $\gamma^{\tilde\pi}_t \le \beta^{\tilde\pi}_t \le \kappa \le \gamma^{\pi^o}_t \le \beta^{\pi^o}_t$. Similarly we can prove Case (b). In Case (c) where $Y^{\pi^o}_t = Y^{\tilde\pi}_t$, if all the decisions for two policies are the same, the claim holds trivially. Otherwise the only possible case under the fairness constraint is that both policies accept and reject the same class customer in a different order, which implies that $\gamma^{\tilde\pi}_t = \gamma^{\pi^o}_t = \beta^{\tilde\pi}_t = \beta^{\pi^o}_t$. **Q.E.D.**

Proof of Lemma 6. Identical to (11–13) in the two-class case, we partition the set of periods $\{1, \ldots, T\}$ into three disjoint subsets $\mathscr{T}_a$, $\mathscr{T}_b$ and $\mathscr{T}_c$. Since $\tilde\pi$ generates no regret in $\mathscr{T}_c$, we focus on $\mathscr{T}_a$ and $\mathscr{T}_b$ only. Now we fix a sample path f^{++}_T, and examine any period $t = 1, \ldots, T$. If $t \in \mathscr{T}_a$, as long as π^o holds positive inventory, by Lemma 5, π^o must accept all the customers that are accepted by $\tilde\pi$. Hence, by the same argument in Lemma 1, the cumulative regret of acceptance incurred by π^o must be higher than or equal to $\tilde\pi$. Similarly, if $t \in \mathscr{T}_b$, we can see by Lemma 5 that π^o must reject all the customers that are rejected by $\tilde\pi$. Hence, by the same argument in Lemma 2, the cumulative regret of rejection incurred by π^o must be higher than or equal to $\tilde\pi$. **Q.E.D.**