Submitted to *Management Science*
manuscript MS-0000-0000.00

Bayesian Demand Learning and Revenue Management under Limited Capacity

Mihalis G. Markakis
IESE Business School, University of Navarra MMarkakis@iese.edu

Victor Martínez-de-Albéniz
IESE Business School, University of Navarra VAlbeniz@iese.edu

Marcos Serrano-Mayor
Graduate School of Business, Stanford University MarcosSM@stanford.edu

In Revenue Management (RM) problems with limited capacity, the optimal price or quantity decision is mainly driven by the ratio of supply over total demand during the sales season. Under ex ante uncertainty about certain demand statistics, primarily its rate, the seller not only adjusts pricing or capacity allocation to optimize revenues but also to learn the demand. When capacity is limited, however, conceptually simple strategies based on the principle of "estimate, then optimize" are unlikely to work well because the opportunity cost of every unit of capacity is high. We are interested in characterizing effective experimentation strategies in this context. We thus consider a stylized quantity-based RM problem whereby a fixed amount of capacity needs to be sold within a given horizon, into either a high margin, low volume channel, or a low margin, high volume one. The demand rate in either channel is uncertain, but a prior distribution over it is available. We formulate the dynamic optimization problem with Bayesian demand learning. We provide a clean and intuitive structural characterization for the general case of the problem; a closed-form solution for the special case where there is one unit of capacity to sell (precisely the regime of limited capacity); and an efficient heuristic policy for the multi-unit case, which provides near-optimal performance in various regimes in our numerical experiments. Somewhat surprisingly, we find that higher uncertainty regarding the demand rate may push the seller to opt for a high margin, low volume position for longer, rather than look for higher volumes to accelerate learning, as intuition may suggest. Finally, we show that the monetary value of Bayesian demand learning is comparable to the value of allocating capacity in an optimal way, suggesting that demand learning could be a first-order consideration in RM.

Key words: Quantity-based RM; Bayesian demand learning; exploration-exploitation; margin-volume trade-off.

History: Submitted: April 8, 2024

1. Introduction

Revenue Management (RM), i.e., the tactical optimization of pricing and capacity allocation decisions, supported by statistical methods for estimation and forecasting, economic theory, and advances in Information Technology, has revolutionized the practice of demand management since the 1980s; starting with the airline and hospitality industries, and gradually extending to mobility, retailing, online advertising, and many others. While the scientific rigor and the potential value that RM can add are beyond doubt, its most successful applications have some common characteristics: a company has pricing power to shape the demand for its products or services, and the available capacity (e.g., hotel rooms to rent, units of inventory to sell) is perishable and relatively scarce compared to the demand. Indeed, it is in the regime of limited capacity where every pricing and/or allocation decision has to be made very carefully, because every mistake counts. In contrast, when capacity is plentiful relative to demand and the selling horizon long, there is not much need or benefit from tactical optimization, as there is hardly any opportunity cost and, thus, heuristic decision making rules tend to work reasonably well.

The classical formulation of RM problems is to assume that demand is stochastic but its statistics are known inputs, e.g., Brumelle and McGill (1993), Gallego and Van Ryzin (1994). In many contexts of interest, however, the firm has limited ex ante knowledge of the demand rate, and pricing decisions not only serve to maximize expected revenue, but are an essential piece of experimentation, e.g., Caro and Gallien (2007), Araman and Caldentey (2009), Besbes and Zeevi (2009), Farias and Van Roy (2010), Harrison et al. (2012), V. den Boer and Zwart (2014). The goal of this paper is to understand how to perform RM effectively, when the company does not know with certainty (the statistics of) the demand for its products or services, and thus cannot be sure to what extent its available capacity is scarce. For concreteness, let us provide some examples. Consider, for instance, a small-sized luxury apparel brand, deciding whether to sell its products through its direct channels or through an online marketplace for luxury clothing and accessories such as Farfetch. Typically, production batches of luxury items are small, so each unit of available capacity must be allocated carefully to the appropriate channel. The trade-off that the brand faces is higher margin but lower sales volume in the direct channels, against higher volume and lower margin through the platform (due to the latter's hefty commission on revenues). What makes the capacity allocation decision hard is that luxury items have short life cycles, and that their demand is highly volatile and unpredictable ex ante. An analogous dilemma is posed to a small-sized resort hotel, deciding whether to allocate its available capacity through their direct channels or through an online travel agent such as Booking.com: again, the trade-off is margin vs. volume, and the decision hindered by limited capacity and the high degree of ex ante uncertainty regarding the demand. In both these two cases, channels cater to different audiences and are hence mostly independent from each other. As a result, the choice of channel should be driven by the preference of the firm for margin vs. volume, as well as by the desired speed of learning, which usually favors the marketplace due to its larger scale.

To deal with the above challenge, one would ideally learn the demand as quickly as possible and use the data and insights obtained in order to make better informed decisions. Of course, the value of in-season demand learning is undisputed: a long-standing rule of thumb in retailing is that within the first few weeks of the sales season, a company can safely distinguish hit from flop products, e.g., Fisher and Raman (1996), Caro and Martínez-de Albéniz (2015). Leveraging this insight in the context of RM though, has been a more recent matter: the tactical optimization of pricing and capacity allocation decisions, combined with demand learning, gives rise to challenging dynamic learning problems (sometimes termed "learning and earning"), whose solutions are intellectually rich and practically relevant. This sub-field of RM has attracted significant attention from academics the last 15 years, and typically employs the regret as performance evaluation criterion, primarily through a frequentist approach and with an emphasis on how regret scales with time. In our case, given the limited capacity and selling horizon, this approach does not seem appropriate. In contrast, a Bayesian approach based on a prior distribution of the demand that is updated during the sales season, and an exact (Dynamic Programming) analysis of the dynamic learning problem at hand seems like a more fruitful avenue of investigation. Such formulations are usually analytically intractable and often lead to computationally inefficient algorithms – something that frequentist approaches usually succeed in. Fortunately, by focusing on a canonical model with two possible actions, we are able to provide an exact analysis and deliver near-optimal algorithms to support RM policies that learn and exploit effectively.

Specifically, we consider a fundamental quantity-based RM problem: a seller has a fixed amount of capacity to sell, in a given amount of time, with the price menu also being fixed. For each unit of capacity, the seller has to decide whether to sell through a high margin, low volume channel or through a low margin, high volume channel, and at which point in time each channel makes sense. Demand arrives according to a Poisson process, scaled versions of which are observed in the two channels, but the rate of this process is uncertain. In other words, the seller knows the sizes of the two channels in relative, but not in absolute terms. The seller has a Gamma distributed prior over the Poisson rate parameter, which is updated dynamically as time passes and sales are made. We formulate the decision problem as an Optimal Control one with Bayesian learning. Among the contribution of our work, we consider the following.

1. We provide a clean structural characterization of the solution for the general case, whereby a unit of capacity is to be allocated to the low margin, high volume channel at a given point in time,

if and only if the value of learning minus the value of capacity of that particular unit exceeds a certain threshold that depends on the revenues and the relative sizes of the two channels. This is reminiscent of the structural result in the seminal paper Gallego and Van Ryzin (1994) regarding revenue vs. opportunity cost of capacity, albeit in a price-based RM context and without any learning considerations;

2. For the special case of one unit of capacity, we provide a closed-form solution for the optimal allocation decision. We note that this special case is precisely our regime of interest, where capacity is truly limited and, thus, the decision most impactful. Through this closed-form solution, we derive an important managerial insight: the optimal time to allocate the capacity to the low margin, high volume channel is non-monotonic with respect to the uncertainty regarding the demand statistics. In other words, higher uncertainty regarding the mean demand may imply allocating earlier or later. This phenomenon, which we liken to a "lottery ticket," implies that a markdown strategy with progressive price/margin reductions is inappropriate in our setting, because it may be optimal for the seller to reverse a mark-down with a subsequent mark-up on certain occasions;

3. Based on the closed-form solution that we obtain for the single-unit case, we devise a heuristic policy for the general case, where the parameters of the Gamma distribution are scaled appropriately. The algorithm requires, virtually, no computation and performs well in our numerical experiments;

4. Using the special single-unit case and the corresponding closed-form solution as backdrop, we evaluate the dollar value of Bayesian demand learning, and find it of the same order of magnitude as the value of sharing optimally (rather than sharing in a static way). This finding suggests that demand learning could be a first-order consideration for a firm; at least, as important as the practice of RM itself.

To put the contributions of the present work in perspective, we note that the same volume vs. margin trade-off lies at the heart of the most studied – and most applied in practice – model of quantity-based RM, that of capacity allocation in a single-leg flight with two fare classes, e.g., Business and Economy, and uncertain demand. The concept of booking limit captures the fact that a certain amount of seats need to be protected for the high-margin class of customers that arrives later, and Littlewood's rule (equivalent to the newsvendor model in an inventory context) quantifies the trade-off and provides a prescription for the booking limit in this stylized model. We argue that our study is closely related to the aforementioned single-leg, two-class capacity control problem: the booking limit and Littlewood's rule educate the tactical decision of how many seats to protect for the high-margin customers, which is made at the beginning of the sales season, several months before the flight. In our model, this corresponds to deciding the available capacity that the seller starts with. In RM practice, the above tactical decision is operationalized through frequent reoptimization, a heuristic but effective approach. Our work can be viewed as providing the optimal way to execute the booking limit, while dynamically learning the demand.

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature to our work. Sections 3 and 4 provide a detailed description of our quantity-based RM model with Bayesian demand learning; the main analytical results and managerial insights derived from the special single-unit case; and an efficient heuristic for the general case. In Section 5, we compare the value of learning to the value of sharing optimally, and show that they are of the same order. We conclude the paper with a broader discussion in Section 6.

2. Literature Review

We start our literature review with the papers related to quantity-based RM; specifically, the ones that focus on a single resource, e.g., single-leg flights in an airline context. This literature began with the seminal work of Littlewood (1972) on the booking limit/protection level for the static single-leg, two-class capacity allocation problem. The EMSR heuristics in Belobaba (1989) and Belobaba (1992) capitalized on "Littlewood's rule" in order to solve the n-class problem in an approximate yet computationally efficient way, and were adopted widely by the airline industry. In parallel, a series of papers provided the optimal solution to the n-class problem under different assumptions; e.g., see Curry (1990), Wollmer (1992), Brumelle and McGill (1993) and Robinson (1995). Other works focused on relaxing an important assumption of Littlewood's work, that customers of a given fare class arrive all together, and different classes arrive in a segregated, sequential way; e.g., see Lee and Hersh (1993) and Lautenbacher and Stidham Jr. (1999). Within this strand of literature, the work that comes closest to ours is that of Martínez-de Albéniz et al. (2022): their model is quite similar to the one introduced here, but their focus is quite different, as the optimal capacity allocation policy of the seller is used to study the best-response pricing of the partner channel.

A common characteristic of the aforementioned works is that a probabilistic description of demand uncertainty is readily available, hence there is no need to learn the demand. An early paper that has a learning "flavor" is the adaptive algorithm of van Ryzin and McGill (2000) for the two-class capacity problem, which employs a stochastic approximation approach in order to update directly the booking limit based on new data, without resorting to cycles of forecasting and optimization. In a quantity-based Network RM setting with parametric demand uncertainty, Jasin (2015) develops a Linear Programming (LP)-based probabilistic allocation control, based on re-optimization of allocation decisions and re-estimation of parameter values, which outperforms standard LP-based booking limit and bid price control policies.

In the broader class of quantity-based RM, one can also include assortment optimization problems. Of particular relevance to our paper is Caro and Gallien (2007), which formulates the problem of dynamic assortment optimization for seasonal consumer goods as a multi-armed bandit problem with Bayesian demand learning. Specifically, the authors assume a Poisson demand model with a Gamma prior distribution on its rate parameter. The fact that the two distributions are conjugate implies that the posterior distribution is also Gamma, resulting in tractable updating of its shape and rate parameters. We follow a similar approach in this paper.

Related to capacity allocation, although not RM strictly speaking, Chod et al. (2021) study the question of investing in flexible resources in a two-period model, and argue that the monetary value of learning the demand from censored observations is of the same order as the celebrated risk pooling benefit of resource flexibility; hence, potentially a first-order consideration for a firm. Our results confirm and strengthen this insight in the context of a continuous-time model of quantity-based RM with Bayesian learning.

On the other hand, during the last two decades, we have witnessed intense research activity on price-based RM with demand learning. The part of this literature that methodologically comes closer to our work is the one that employs Bayesian demand learning. We can roughly classify these works in two types. The first one assumes a conjugate prior to the demand process, much like in Caro and Gallien (2007); e.g., in Aviv and Pazgal (2002), this is integrated in a certainty-equivalence approach; in Cope (2007), it is the basis of a one-step lookahead with recourse policy; in Farias and Van Roy (2010), the core of their proposed decay-balancing heuristic; and in Papanastasiou and Savva (2017), to study the effect of social learning on dynamic pricing with strategic consumers. The second type assumes that the unknown demand parameter can only take two values (e.g., high and low), resulting in the Bayesian updating of a single parameter, and maintaining tractability that way; e.g., Araman and Caldentey (2009) and Harrison et al. (2012) follow such an approach in their greedy heuristic and myopic Bayesian policy, respectively. It is worth emphasizing that, even though the Bayesian updating of the unknown parameters is tractable in the above works, the corresponding dynamic learning problems are not. The first two papers, chronologically, provide no theoretical guarantees and resort to numerical experiments for performance evaluation, whereas the remaining ones provide interesting analytical results in the form of upper and lower bounds that are more meaningful in asymptotic regimes though. (Papanastasiou and Savva (2017) is an exception, by postulating a two-period model that they solve to optimality.)

In parallel, a significant body of work has been developing on price-based RM of a single product/resource, adopting non-Bayesian learning approaches. For instance, Besbes and Zeevi (2009) propose a nonparametric pricing algorithm, based on the separation of learning and optimization; Broder and Rusmevichientong (2012) introduce the MLE cycle policy, where the selling horizon is split in exploration-exploitation cycles; Chen and Farias (2013) focus on the re-optimized fixed price policy; Wang et al. (2014) provide an algorithm that searches for the optimal price in progressively shrinking intervals; V. den Boer and Zwart (2014) develop the controlled variance pricing approach, which combines the certainty equivalence principle with a progressively shrinking taboo interval; and Besbes and Zeevi (2015) show the sufficiency of simple parametric models in this setting. In contrast to the papers above, V. den Boer and Zwart (2015) consider a fixed amount of inventory that needs to be sold within a fixed amount time – a setting that comes closer to our work – and employ a combination of certainty equivalence and maximum-likelihood estimation. The analyses in the aforementioned papers are based on deriving order-wise matching upper and lower bounds, and all but one employ the regret as performance evaluation criterion. On the other hand, there is also literature on price-based Network RM with non-Bayesian learning, e.g., Besbes and Zeevi (2012) and Chen et al. (2019) follow nonparametric approaches, whereas Chen et al. (2021) employ a parametric demand model.

A work that, perhaps, bridges the two strands of literature on dynamic demand learning in price-based RM, Bayesian and non-Bayesian, is Johnson Ferreira et al. (2018), which employs Thompson (posterior) sampling to solve the online version of the Network RM problem. The approach is Bayesian in nature but the notion of regret is adopted for performance evaluation; in this case, the expected regret with respect to the posterior demand distribution, often termed as Bayesian regret. As already argued though, performance guarantees in terms of the regret make more sense in problems where the inventory/capacity and the horizon are large, which is not our regime of interest.

Bayesian approaches to dynamic learning problems in RM can also be found in a strand of literature that may be smaller compared to quantity-based and price-based RM, but is of increasing importance, which we could call information-based RM, e.g., see Bertsimas and Mersereau (2007) and Drakopoulos et al. (2021). It is clear though that (limited) capacity is not an issue there, leading to different dynamics and insights compared to our work.

Finally, it is worthwhile acknowledging the sizeable literature on inventory management with Bayesian demand learning. The most followed setting of investigation is the inventory management of a perishable product over repeated selling seasons (effectively, a repeated newsvendor model), where the demand over different seasons is i.i.d. and belongs to a particular family of probability distributions, with one or more parameters unknown. A prior on these parameters is assumed to be known though, and the belief about the true parameter values is updated with observed demand samples through Bayes rule. Early works in that direction include Scarf (1959), Karlin (1960), and Iglehart (1964), which focus on exponential families of demand distributions. Murray and Silver (1966), Azoury (1985), and Eppen and Iyer (1997) study the inventory management of non-perishable products and incorporate Bayesian learning into a Dynamic Programming framework, while Chang and Fyffe (1971) uses a Kalman filtering approach to achieve efficient learning/forecasting. Lovejoy (1990) shows the near-optimality of simple myopic inventory management policies, when combined with adaptive tuning of the parameters via Bayes rule or exponential smoothing.

More recently, the focus of this strand of literature has shifted to learning under censored data, e.g., by observing sales rather than the actual demand. Lariviere and Porteus (1999) provides the Bayesian optimal inventory level if the demand belongs to the class of "newsvendor distributions," and confirms that it is optimal to enhance learning through stocking higher; a phenomenon that has been termed information stalking. Besbes et al. (2022), building on the framework of Lariviere and Porteus (1999), provides both analytical and numerical evidence to the fact that the cost in being myopic (rather than far-sighted, in the Dynamic Programming sense) is actually small. The picture becomes more complicated in the case of nonperishable products: the inventory carried over from previous periods may force the inventory manager to stock higher or lower compared to the Bayesian myopic benchmark; see Chen and Plambeck (2008). We note that these settings are distinctly different than ours due to the ability to re-stock inventory before the next period or selling season. In contrast, we have a fixed inventory to sell, in a fixed amount of time, with the selling channel being the main decision. Thus, the dynamics of our model are much closer to those of quantity-based RM.

3. Model

In this section, we introduce a canonical quantity-based RM problem, which we construct with analytical tractability and insights in mind rather than exact prescriptions. Specifically, we consider the problem of a seller that has Q units of capacity to sell within T units of time. We denote by t, a continuous variable, the time remaining until the end of the selling season which, following the quantity-based RM convention, we define as $t = 0$. In other words, $t = T$ corresponds to the beginning of the selling season, and time goes backward to zero.

Let us assume that, at time $t \in [0, T]$, the seller has q_t units of capacity remaining. If $q_t > 0$, then the seller can allocate capacity to one of two distinct selling channels: either to a lower-volume channel with margin $r_H = 1$, or to a higher-volume partner channel with margin $r_L < 1$. We denote these actions by $a_t = H$ (high margin) or $a_t = L$ (low margin). Note that, in practice, $a_t = L$ may involve selling in both channels, e.g., see Martínez-de Albéniz et al. (2022), but in which case our formulation is still valid by setting r_L to be the weighted average of margins across channels. In contrast, if $q_t = 0$, then there is no capacity for sale anymore, which we note as action $a_t = 0$.

There is a demand intensity parameter λ that is uncertain (below, we become more precise regarding the information that the seller possesses about λ), so that the demand in each selling channel follows a Poisson process of intensity $\theta_{a_t}\lambda$, when the chosen action is a_t. Without loss of generality, we set $\theta_H = 1$, so the demand intensity is λ in the lower-volume channel; and $\theta_L > 1$ in the higher-volume channel, so that the demand intensity is $\theta_L\lambda$. (If θ_L was less than or equal to one, then the seller would have no incentive to ever opt for channel L). Summarizing, the Poisson multipliers, which represent the relative sizes of the different channels and are known to the seller a priori, are:

$$\theta_{a_t} = \begin{cases} 1, & \text{if } a_t = H \\ \theta_L > 1, & \text{if } a_t = L \\ 0, & \text{otherwise.} \end{cases}$$

Note that we assume that these multipliers are independent of the capacity allocation decision. This assumption is reasonable in our chosen context, e.g., if the high-margin channel corresponds to a brick-and-mortar retail store and the low-margin channel to an online store, dedicated to serving locations where the company does not have physical presence: the number of those locations could be large, resulting in high volume of traffic, while covering the fulfilment cost reduces the company's margin. There is no reason to suspect that allocating capacity to one channel should have an effect on the demand in the other channel, as the two serve distinct markets.

Let N_t be the cumulative number of arrivals from T to t. Since we are counting time backward, $N_T = 0$ and N_t decreases in t. Given the properties of the Poisson process, we have that the variation of N_t in the infinitesimal interval $[t, t-dt]$ is equal to $dN_t = -1$ with probability $\theta_{a_t}\lambda dt$, and $dN_t = 0$ otherwise.

As mentioned already, our focus in this paper is to examine the role of demand uncertainty, and learning, on RM with limited capacity. For this purpose, we adopt a Bayesian framework and assume that λ is the realization of a random variable Λ, as well as that the seller starts with a prior belief on Λ given by a Gamma distribution with a shape parameter α_T, and a rate parameter β_T. Specifically,

$$f_{\Lambda_T}(\lambda|\alpha_T, \beta_T) = \frac{\beta_T(\beta_T\lambda)^{\alpha_T-1}e^{-\beta_T\lambda}}{\Gamma(\alpha_T)}, \quad \lambda \geqslant 0.$$

We choose a Gamma-distributed prior for analytical tractability, because it is the conjugate to the Poisson distribution. As a result, Λ_t remains Gamma-distributed as t decreases from T to zero. To see why this is true, consider the distribution of Λ_{t-dt}, given that at t, Λ_t is Gamma-distributed with parameters (α_t, β_t). If the arrival rate is $\theta\lambda$, then the probability of one arrival in the interval $[t, t-dt]$ is equal to

$$\begin{aligned}\mathbb{P}[-dN_t = 1] = \int_0^\infty f_{\Lambda_t}(\lambda|\alpha_t, \beta_t)\mathbb{P}[-dN_t = 1|\Lambda_t = \lambda]d\lambda &= \int_0^\infty \frac{\beta_t(\beta_t\lambda)^{\alpha_t-1}e^{-\beta_t\lambda}}{\Gamma(\alpha_t)}\theta\lambda dt e^{-\theta\lambda dt}d\lambda \\ &= \frac{\Gamma(\alpha_t+1)}{\Gamma(\alpha_t)}\frac{\beta_t^{\alpha_t}}{(\beta_t+\theta dt)^{\alpha_t+1}}\theta dt.\end{aligned}$$

Within an interval of infinitesimal length dt, there can be 0 or 1 arrivals, hence $\mathbb{P}[dN_t = 0] = 1 - \mathbb{P}[-dN_t = 1]$. We can apply Bayes' rule, conditional on observing that dN_t units were sold: the posterior distribution of the market parameter Λ_{t-dt} becomes a Gamma distribution with a shape parameter equal to α_t plus the number of units sold during $[t, t-dt]$, and a rate parameter that equals to β_t plus the constant θdt. In other words, throughout the sales season, the seller's belief about the market condition will remain Gamma, but with parameters that are updated over time as follows:

$$(\alpha_t, \beta_t) \to \begin{cases} (\alpha_t, \beta_t) & \text{if } a_t = 0 \\ (\alpha_t - dN_t, \beta_t + dt) & \text{if } a_t = H \\ (\alpha_t - dN_t, \beta_t + \theta_L dt) & \text{if } a_t = L. \end{cases}$$

In other words, the shape parameter mimics the sales process (i.e., has jumps), and the rate parameter grows in a continuous way, but faster in the high-volume channel.

Integrating the equation above, which is contingent on the policy employed and hence path-dependent, leads to the following (implicit) description of the current state, while $q_t > 0$ (so that action $a_t = 0$ is never employed):

$$\alpha_t = \alpha_T + N_t \text{ and } \beta_t = \beta_T + (T - t) + (\theta_L - 1)\Delta, \tag{1}$$

where $\Delta = \int_t^T 1_{a_t = L} dt$ is the duration in which the firm has opted for the low margin-high volume channel. The expected arrival intensity is $\mathbb{E}[\Lambda_t] = \alpha_t/\beta_t$, the standard deviation is $\sigma[\Lambda_t] = \sqrt{\alpha_t}/\beta_t$ and, consequently, the coefficient of variation of Λ_t is equal to $1/\sqrt{\alpha_t}$. This implies that opting for the high-volume channel has an advantage in terms of learning: it reduces the coefficient of variation of Λ_t by stochastically increasing N_t and, thus, accelerates learning.

The objective of the seller is to maximize their expected revenue:

$$\sup_{\boldsymbol{a_t}} \mathbb{E}\left[\int_0^T -r_{a_t} dN_t\right], \tag{2}$$

with $a_t = \pi_q(t, \alpha, \beta)$, where $\pi_q(\cdot)$ is a mapping from the state to the action space that satisfies $\pi_0(t, \alpha, \beta) = 0$ and $\pi_q(0, \alpha, \beta) = 0$, for all α and β. This is an Optimal Control problem, with initial conditions α_T, β_T, and q_T.

We denote by $V_q(t, \alpha, \beta)$ the revenue-to-go in $[t, 0]$, at state $\alpha_t = \alpha$, $\beta_t = \beta$, $q_t = q$, with boundary conditions $V_q(0, \alpha, \beta) = 0$ and $V_0(t, \alpha, \beta) = 0$. To derive the Hamilton-Jacobi-Bellman (HJB) equation, we look at the infinitesimal decision in $[t, t-dt]$, in state $\alpha_t, \beta_t, q_t > 0$:

$$V_q(t, \alpha, \beta) = \sup_a \mathbb{E}_{\Lambda_t} \left\{ \begin{array}{l} \theta_a \Lambda_t dt \big(r_a + V_{q-1}(t - dt, \alpha + 1, \beta + \theta_a dt)\big) \\ +(1 - \theta_a \Lambda_t dt) V_q(t - dt, \alpha, \beta + \theta_a dt) \end{array} \right\}.$$

Note that, through a Taylor expansion, $V_q(t - dt, \alpha, \beta + \theta_a dt)$ can be written as

$$V_q(t, \alpha, \beta) + \left(-\frac{\partial V_q}{\partial t}(t, \alpha, \beta) + \theta_a \frac{\partial V_q}{\partial \beta}(t, \alpha, \beta)\right) dt + o(dt).$$

In the limit, and through the standard, formal derivation, we obtain the HJB equation:

$$\frac{\partial V_q}{\partial t}(t,\alpha,\beta) = \frac{\alpha}{\beta}\sup_a \left\{\theta_a\left(r_a + V_{q-1}(t,\alpha+1,\beta) - V_q(t,\alpha,\beta) + \frac{\beta}{\alpha}\frac{\partial V_q}{\partial \beta}(t,\alpha,\beta)\right)\right\}. \tag{3}$$

We define the *value of learning* as

$$\Delta L_q(t,\alpha,\beta) \equiv \frac{\beta}{\alpha}\frac{\partial V_q}{\partial \beta}(t,\alpha,\beta) + V_{q-1}(t,\alpha+1,\beta) - V_{q-1}(t,\alpha,\beta), \tag{4}$$

and the *value of capacity* as

$$\Delta I_q(t,\alpha,\beta) \equiv V_q(t,\alpha,\beta) - V_{q-1}(t,\alpha,\beta). \tag{5}$$

Clearly, the quantities above are defined for $q > 0$; by convention, we assume their values to be equal to zero for $q \leq 0$. Note that the value of learning includes information both from selling and from not selling the q^{th} unit of capacity (time elapsing without sales is informative too); while the value of capacity can be interpreted as the marginal expected value of the q^{th} unit of capacity, similarly to Gallego and Van Ryzin (1994).

We have assumed, without loss of generality, that $r_H = \theta_H = 1$. Given that $\theta_L > 1$, Equation (3) implies that it is optimal to choose the low margin, high volume channel if and only if the value of learning minus the value of capacity exceeds a threshold that depends on the revenues and the relative sizes of the different channels:

$$a_t^* = L \iff \Delta L_q(t,\alpha,\beta) - \Delta I_q(t,\alpha,\beta) \geq \frac{1-\theta_L r_L}{\theta_L - 1}.$$

Let us define

$$\phi_q(t,\alpha,\beta) \equiv \Delta L_q(t,\alpha,\beta) - \Delta I_q(t,\alpha,\beta) - \frac{1-\theta_L r_L}{\theta_L - 1}, \tag{6}$$

which implies that

$$a_t^* = L \iff \phi_q(t,\alpha,\beta) \geq 0. \tag{7}$$

Using this notation, we can re-write Equation (3) as follows:

$$\frac{\partial V_q}{\partial t}(t,\alpha,\beta) = \frac{\alpha}{\beta}\sup_a \left\{\theta_a\left(r_a + \phi_q(t,\alpha,\beta) + \frac{1-\theta_L r_L}{\theta_L - 1}\right)\right\}. \tag{8}$$

By taking the partial derivative with respect to t in Equation (6), and substituting for Eqs. (4)-(5), we have that

$$\begin{aligned}\frac{\partial \phi_q}{\partial t}(t,\alpha,\beta) &= \frac{\partial}{\partial t}\left\{\frac{\beta}{\alpha}\frac{\partial V_q}{\partial \beta}(t,\alpha,\beta) + V_{q-1}(t,\alpha+1,\beta) - V_q(t,\alpha,\beta) - \frac{1-\theta_L r_L}{\theta_L - 1}\right\}\\ &= \left(\frac{\partial}{\partial \beta}\left\{\frac{\beta}{\alpha}\frac{\partial V_q}{\partial t}(t,\alpha,\beta)\right\} - \frac{1}{\alpha}\frac{\partial V_q}{\partial t}(t,\alpha,\beta)\right) + \frac{\partial}{\partial t}\left\{V_{q-1}(t,\alpha+1,\beta) - V_q(t,\alpha,\beta)\right\}\\ &= \frac{\partial}{\partial \beta}\left\{\frac{\beta}{\alpha}\frac{\partial V_q}{\partial t}(t,\alpha,\beta)\right\} + \frac{\partial V_{q-1}}{\partial t}(t,\alpha+1,\beta) - \frac{\alpha+1}{\alpha}\frac{\partial V_q}{\partial t}(t,\alpha,\beta). \end{aligned} \tag{9}$$

Let θ_q and r_q be the relative size of the channel and the revenue, respectively, corresponding to the optimal allocation decision at state (t,α,β) with remaining capacity q; and similarly, θ_{q-1} and r_{q-1} at state $(t,\alpha+1,\beta)$ with remaining capacity $q-1$. Substituting Equation (8) to Equation (9), results in

$$\begin{aligned}\frac{\partial\phi_q}{\partial t}(t,\alpha,\beta) = \theta_q\frac{\partial\phi_q}{\partial\beta}(t,\alpha,\beta)&\\ +\frac{\alpha+1}{\beta}\left\{\theta_{q-1}\left(r_{q-1}+\phi_{q-1}(t,\alpha+1,\beta)+\frac{1-\theta_L r_L}{\theta_L-1}\right)\right\}&\\ -\frac{\alpha+1}{\beta}\left\{\theta_q\left(r_q+\phi_q(t,\alpha,\beta)+\frac{1-\theta_L r_L}{\theta_L-1}\right)\right\}.&\end{aligned} \tag{10}$$

As we can see, Equation (10) is a partial differential equation (p.d.e.) on four dimensions: q,t,α and β. In reality, however, the difficulty stems from the partial derivative with respect to t and β. For this reason, we can take q and α as parameters and solve the p.d.e. for $q=1$ and any α, then for $q=2$, etc.

4. Analytical Results and Insights

4.1. Tractable Special Case: a Single Unit of Capacity

In this section, we focus on a special case of the problem with a single unit of capacity. Somewhat surprisingly relative to existing literature, we show that this case is analytically tractable, and we provide a closed-form solution to it. Importantly, this special case coincides with the regime of interest in our study, which is the regime of limited capacity. Later, we use this closed-form solution to conduct a sensitivity analysis, and from that to derive managerial insights regarding the role of uncertainty regarding the demand statistics in quantity-based RM with dynamic learning.

Let us assume that $q=1$. Recall that $\theta_0=0$, and there is neither value of learning nor value of capacity that does not exist. Hence, the second term of the right-hand side of Equation (10) disappears, and we have that

$$\frac{\partial\phi_1}{\partial t}(t,\alpha,\beta) = \theta_1\frac{\partial\phi_1}{\partial\beta}(t,\alpha,\beta) - \frac{\alpha+1}{\beta}\left\{\theta_1\left(r_1+\phi_1(t,\alpha,\beta)+\frac{1-\theta_L r_L}{\theta_L-1}\right)\right\}. \tag{11}$$

We note that, in the equation above, α is fixed to the value dictated by the prior Gamma distribution: as the α parameter is updated with sales and in this case there is only one unit of capacity, as long as that unit is not sold, α is not updated. On the other hand, as soon as the unit is sold, there are no more decisions to make.

We distinguish between two cases in our analysis. First, let us consider the case where $\theta_L r_L > 1$. In that case,

$$\phi_1(0,\alpha,\beta) = -\frac{1-\theta_L r_L}{\theta_L-1} > 0,$$

which, by Equation (7), implies that $a_0^* = L$. Consider some $t > 0$ in the vicinity of 0, for which $a_t^* = L$. (Such t exists from the continuity of ϕ_1 at $t = 0$.) Equation (11) implies that this point in time must satisfy the p.d.e. in t and β (recall that α is constant in this special case of the problem):

$$\frac{\partial \phi_1}{\partial t}(t,\alpha,\beta) = \theta_L \frac{\partial \phi_1}{\partial \beta}(t,\alpha,\beta) - \frac{\alpha+1}{\beta}\left\{\theta_L\left(r_L + \phi_1(t,\alpha,\beta) + \frac{1-\theta_L r_L}{\theta_L - 1}\right)\right\}. \tag{12}$$

Using as a boundary condition $\phi_1(0,\alpha,\beta) = -(1-\theta_L r_L)/(\theta_L - 1)$, the solution to Equation (12) in this regime can be obtained in closed form:

$$\phi_1(t,\alpha,\beta) = r_L\left(\frac{\beta}{\beta + t\theta_L}\right)^{\alpha+1} + \frac{r_L - 1}{\theta_L - 1}. \tag{13}$$

This solution is valid as long as $\phi_1 < 0$.

On the other hand, let us consider the case where $\theta_L r_L \le 1$. In that case,

$$\phi_1(0,\alpha,\beta) = -\frac{1-\theta_L r_L}{\theta_L - 1} \le 0,$$

which, by Equation (7), implies that $a_0^* = H$. Again, let us consider some $t > 0$ in the vicinity of 0, for which $a_t^* = H$, i.e., $\phi_1 > 0$. Equation (11) implies that this point in time must satisfy the p.d.e. in t and β (recall that $\theta_H = r_H = 1$, without loss of generality):

$$\frac{\partial \phi_1}{\partial t}(t,\alpha,\beta) = \frac{\partial \phi_1}{\partial \beta}(t,\alpha,\beta) - \frac{\alpha+1}{\beta}\left\{1 + \phi_1(t,\alpha,\beta) + \frac{1-\theta_L r_L}{\theta_L - 1}\right\}. \tag{14}$$

Using as a boundary condition $\phi_1(0,\alpha,\beta) = -(1-\theta_L r_L)/(\theta_L - 1)$, the solution to Equation (14) can be obtained in closed form:

$$\phi_1(t,\alpha,\beta) = \left(\frac{\beta}{\beta + t\theta_L}\right)^{\alpha+1} + \frac{\theta_L(r_L - 1)}{\theta_L - 1}. \tag{15}$$

Again, this solution remains valid while $\phi_1 > 0$.

PROPOSITION 1. *Consider the capacity allocation problem in Section 3 with $Q = 1$ unit, and fix α and β. There exists a unique $t_1(\alpha,\beta) \in [0,T]$ such that it is optimal for the firm to allocate the unit of capacity to the low margin, high volume channel for all $t \le t_1(\alpha,\beta)$. In other words, $\phi_1(t,\alpha,\beta) \ge 0$ if and only if $t \le t_1(\alpha,\beta)$.*

Proof. Following the line of analysis above, we distinguish between two cases in the proof. If $\theta_L r_L > 1$, then

$$\phi_1(t,\alpha,\beta) = r_L\left(\frac{\beta}{\beta + t\theta_L}\right)^{\alpha+1} + \frac{r_L - 1}{\theta_L - 1}.$$

Note that $\phi_1(t,\alpha,\beta)$ is continuous and monotonically decreasing in t. Given that $\phi_1(0,\alpha,\beta)>0$, $\phi_1(\cdot,\alpha,\beta)$ either crosses zero at a single point, or it never crosses it. In both cases there is a unique $t_1(\alpha,\beta)$ such that $a_t^*=L$, for all $t\le t_1(\alpha,\beta)$; in the former case, that point is the (unique) root of $\phi_1(t,\alpha,\beta)=0$; in the latter case, $t_1(\alpha,\beta)=T$.

Similarly, if $\theta_L r_L\le 1$, then

$$\phi_1(t,\alpha,\beta)=\left(\frac{\beta}{\beta+t\theta_L}\right)^{\alpha+1}+\frac{\theta_L(r_L-1)}{\theta_L-1}.$$

Given that $\phi_1(t,\alpha,\beta)$ is continuous and monotonically decreasing in t and $\phi_1(0,\alpha,\beta)\le 0$, this implies that $\phi_1(t,\alpha,\beta)\le 0$, and hence $a_t^*=H$, for all t. In that case, the unique $t_1(\alpha,\beta)$ is 0. □

A consequence of this result is that, in the extreme case where $t_1(\alpha,\beta)=0$, it is optimal for the firm to keep the unit of capacity in the high margin, low volume channel throughout the selling season. In contrast if $t_1(\alpha,\beta)=T$, selling through the low margin, high volume channel from the beginning of the season is optimal for the firm. Our main analytical result, presented below, provides a closed-form expression for $t_1(\alpha,\beta)$ for different regions of the parameter space and, thus, a complete solution for the special case of the problem with a single unit of capacity.

Theorem 1. *Consider the capacity allocation problem in Section 3 with $Q=1$ unit, and fix α and β. If $\theta_L r_L\le 1$, then $t_1(\alpha,\beta)=0$, for all α and β; else if $\left(\frac{T\theta_L}{\beta}+1\right)^{\alpha+1}\le m$, then $t_1(\alpha,\beta)=T$; otherwise,*

$$t_1(\alpha,\beta)=\frac{\beta}{\theta_L}\left(m^{\frac{1}{\alpha+1}}-1\right); \tag{16}$$

where

$$m\equiv\frac{r_L(\theta_L-1)}{(1-r_L)}. \tag{17}$$

Proof. The claim that, if $\theta_L r_L\le 1$, then $t_1(\alpha,\beta)=0$, for all α and β, is established in the proof of Proposition 1. Consequently, henceforth, we assume that $\theta_L r_L>1$. Equation (13) implies that the (unique) root of $\phi_1(\cdot,\alpha,\beta)$ is equal to

$$\frac{\beta}{\theta_L}\left(m^{\frac{1}{\alpha+1}}-1\right).$$

If the above root is greater than or equal to T, which is the case if

$$\left(\frac{T\theta_L}{\beta}+1\right)^{\alpha+1}\le m,$$

then $t_1(\alpha,\beta)=T$; otherwise,

$$t_1(\alpha,\beta)=\frac{\beta}{\theta_L}\left(m^{\frac{1}{\alpha+1}}-1\right),$$

as the statement of the theorem suggests. □

4.2. Comparative Statics and Managerial Insights

As discussed in the previous section, if $\theta_L r_L \le 1$, then $t_1(\alpha,\beta)=0$, i.e., it is never optimal for the firm to employ the low margin, high volume channel. For that reason, in our analysis henceforth, we focus on the more interesting case where $\theta_L r_L > 1$, particularly on scenarios where $t_1(\alpha,\beta)\in(0,T)$. We perform comparative statics on the optimal time to switch sales channels with respect to the uncertainty regarding the demand statistics, which in our model translates in the parameters α and β of the Gamma prior distribution.

Recall that the mean of the Gamma distribution is equal to $\mathbb{E}[\Lambda]=\alpha/\beta$. To make the results and insights that follow more intuitive, we re-parameterize the optimal switching time in terms of the expected value of the distribution, $\mathbb{E}[\Lambda]$, and the β parameter which then captures the uncertainty around it: the variance of the Gamma distribution is equal to $\sigma^2[\Lambda]=\alpha/\beta^2=\mathbb{E}[\Lambda]/\beta$. Hence, for a given $\mathbb{E}[\Lambda]$, the higher the β, the lower the uncertainty around $\mathbb{E}[\Lambda]$. With this parameterization, Theorem 1 implies that

$$t_1\big(\mathbb{E}[\Lambda],\beta\big)=\frac{\beta}{\theta_L}\left(m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}}-1\right), \tag{18}$$

We note that $\theta_L r_L > 1$ implies that the constant m is strictly greater than 1, so that $\ln(m)>0$.

Figures 1 and 2 illustrate the optimal switching time in Equation (18) as a function of β, for different combinations of θ_L and r_L values, for a relatively low and a relatively high $\mathbb{E}[\Lambda]$, respectively. A note may be worthwhile making regarding the right-most plot in Figure 1, which is the only case where the optimal switching time is capped at T, according to the condition provided in Theorem 1. Pertaining to the dependence of $t_1(\cdot)$ on the two main quantities of interest, $\mathbb{E}[\Lambda]$ and β, in these figures we observe that: (i) for given β, higher $\mathbb{E}[\Lambda]$ implies switching later to the low-margin, high-volume channel; (ii) for given $\mathbb{E}[\Lambda]$, higher β, i.e., lower uncertainty around $\mathbb{E}[\Lambda]$, may imply switching earlier or later to the low-margin, high-volume channel. The corollaries below formalize and quantify these observations.

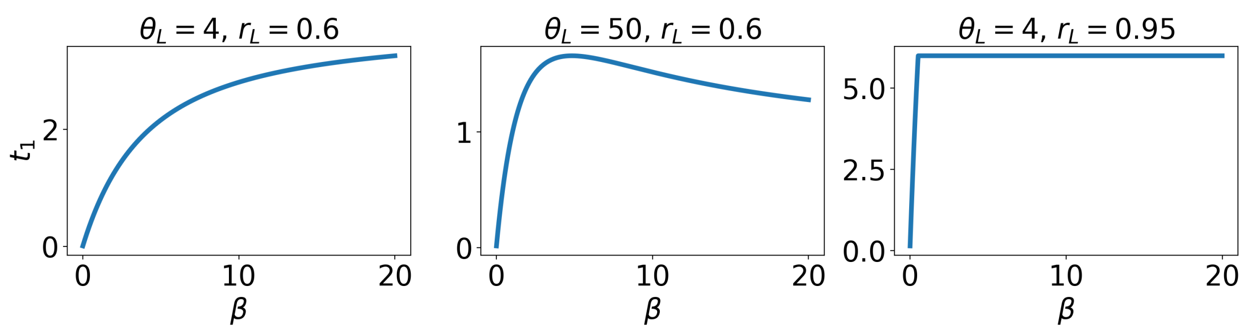

Figure 1 t_1 **as a function of** β **for low demand,** $\mathbb{E}[\Lambda]=0.1$, $T=6$.

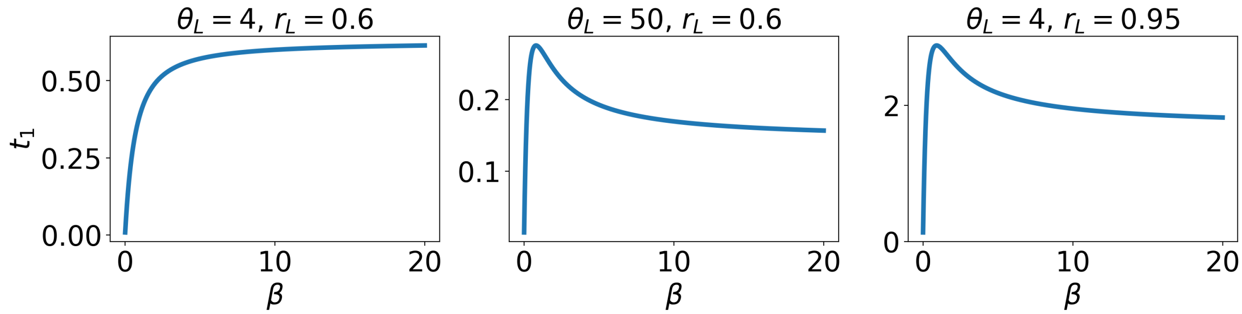

Figure 2 t_1 **as a function of** β **for high demand,** $\mathbb{E}[\Lambda] = 0.6$**,** $T = 6$**.**

COROLLARY 1. *Consider the capacity allocation problem in Section 3 with $Q = 1$ unit, $\theta_L r_L > 1$, and fix β. The optimal switching time $t_1(\cdot)$ is monotonically decreasing in $\mathbb{E}(\Lambda)$, i.e., higher expected value of the Poisson rate implies switching to the low margin, high volume channel later.*

Proof. By taking the partial derivative in Equation (18) with respect to $\mathbb{E}[\Lambda]$, we have that

$$\frac{\partial t_1\big(\mathbb{E}[\Lambda], \beta\big)}{\partial \mathbb{E}[\Lambda]} = -\ln(m) m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}} \frac{\beta^2}{\theta_L} \frac{1}{(1+\mathbb{E}[\Lambda]\beta)^2} < 0.$$

□

COROLLARY 2. *Consider the capacity allocation problem in Section 3 with $Q = 1$ unit, $\theta_L r_L > 1$, and fix $\mathbb{E}[\Lambda]$. The optimal switching time $t_1(\cdot)$ is monotonically decreasing in β, i.e., less uncertainty around the expected value of the Poisson rate implies switching to the low margin, high volume channel later, if and only if*

$$m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}} \left(1 - \frac{\mathbb{E}[\Lambda]\beta}{(1+\mathbb{E}[\Lambda]\beta)^2} \ln(m)\right) < 1. \tag{19}$$

Proof. By taking the partial derivative in Equation (18) with respect to β, we have that

$$\frac{\partial t_1(\mathbb{E}[\Lambda], \beta)}{\partial \beta} = \frac{m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}} - 1}{\theta_L} - \frac{\beta}{\theta_L} \ln(m) m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}} \mathbb{E}[\Lambda]\big(1+\mathbb{E}[\Lambda]\beta\big)^{-2}.$$

Using the equation above, it can be verified that

$$\frac{\partial t_1(\mathbb{E}[\Lambda], \beta)}{\partial \beta} < 0 \iff m^{\frac{1}{1+\mathbb{E}[\Lambda]\beta}} \left(1 - \frac{\mathbb{E}[\Lambda]\beta}{(1+\mathbb{E}[\Lambda]\beta)^2} \ln(c)\right) < 1.$$

□

The results above lead to crisp managerial insights regarding the practice of quantity-based RM with limited capacity, in the presence of (Bayesian) demand learning. The first insight suggests that, for a given level of uncertainty regarding the demand statistics, the higher the anticipated demand rate, the later the seller should switch to the low margin-high volume channel. This finding is quite intuitive because, if the seller believes that demand will be strong enough in the high-margin channel, it can sustain high prices longer and just introduce discounts through the low-margin channel later in the season, all else being equal.

Our second insight is somewhat less intuitive at first sight: for a given level of anticipated demand rate, one would expect that higher uncertainty around that rate would imply allocating the capacity to the low margin, high volume channel earlier. That way, true demand rate could be learned faster, leading, in turn, to a "better" solution to the optimal control problem for the remainder of the selling period. This intuition should certainly hold if the seller had enough capacity, so that they could spare some of it just to learn the demand; along the lines of the "estimate, then optimize" principle. In the regime of limited capacity though, the opportunity cost of every unit of capacity is high, to the extent that once learning materializes (one sale), then there is no more inventory to sell, and as a result there is no possibility of extracting value from such learning. This causes the aforementioned intuition to be reversed on occasions. Specifically, if the anticipated demand rate is relatively low, then, as conventional wisdom suggests, higher uncertainty regarding that rate leads indeed to switching to the low margin, high volume channel earlier. However, if the anticipated demand rate is relatively high, then higher uncertainty regarding that rate leads to switching later.

We liken the mechanism that generates the latter insight to a lottery ticket: unlikely as winning the lottery may be, given that the reward is very high, even a small increase in the chances of winning may make a lottery ticket worth buying in expectation. In our problem, the very high reward stems from the very high opportunity cost of the single unit of capacity. Now, higher ex ante demand rate means higher α, all else being equal. As a result, $\mathbb{E}[\Lambda]$ decreases more slowly as time passes, which implies that time is slowed down, effectively increasing the chances to "win the lottery."

The implications of this result on RM practice become clear if one envisions the dynamical system of Equation (12) evolving in time, with β updated according to Equation (1); recall that the parameter α remains constant in the special case $Q=1$. The non-monotonicity of the optimal switching time with respect to β can lead to the following sequence of events: early in the sales season, the optimal capacity allocation is to the high margin, low volume channel, as Proposition 1 dictates. Let us assume that the initial value of β is small, so that we are in the regime where the optimal time to switch is monotonically increasing in β; see Figures 1 and 2. This implies that there will be a point in time $\bar{t}\in(0,T)$, with $\beta_{\bar{t}}$, where $t_1(\alpha,\beta_{\bar{t}})=\bar{t}$, so that the seller should switch to the low margin, high volume channel. From that point, β starts growing very quickly, meaning that there is now a lot more certainty regarding the value of the Poisson rate. This may imply, depending on the values of θ_L and r_L, that the optimal time to switch is monotonically decreasing in β in this regime. Hence, there may exist a $\hat{t}<\bar{t}$, with $\beta_{\hat{t}}$, where $t_1(\alpha,\beta_{\hat{t}})=\hat{t}$, so the seller switches back to the high margin, low volume channel. Figure 3 illustrates the fact that the optimal solution may involve multiple switches between the two channels, depending on the parameter values of the problem.

The fact that this sample path is feasible in RM problems with limited capacity and significant uncertainty regarding the demand statistics, suggests that a markdown strategy with progressive price/margin reductions, quite popular in retailing, is inappropriate in these problems because it may be optimal for the seller to reverse a mark-down with a subsequent mark-up on certain occasions.

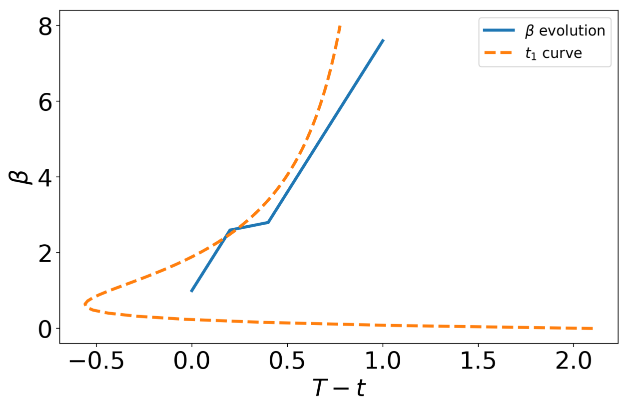

Figure 3 This figure shows the evolution in time of the parameter β of the Gamma distribution, which determines the uncertainty around the expected value, for $q=1$, $T=2.1$, $r_L=0.95$, $\theta_L=4$, $\mathbb{E}[\Lambda_T]=0.6$, and $\beta_T=1$. One can distinguish two periods where the rate of increase is larger, corresponding to the low margin, high volume channel, as well as a period where the rate of increase is smaller, corresponding to the high margin, low volume channel. This illustrates that the optimal policy can involve multiple switches between the two channels.

4.3. A Computational Approach for the Exact Solution in the General Case

Up to this point, our analysis has been focused to a tractable special case of the problem with $Q=1$. Here, we discuss the solution in the general case. Equation (10) provides a recursive way for solving the Optimal Control problem: having obtained $\phi_{q-1}(\cdot)$, one can compute $\phi_q(\cdot)$ by solving the aforementioned p.d.e. in t and β, for any given α. The recursion is initiated with $\phi_1(\cdot)$, which we have obtained in closed form; see Equation (13) for the case where $\theta_L r_L > 1$. In turn, for any fixed α and β, one can obtain the optimal time to switch to the low margin, high volume channel, $t_q(\alpha,\beta)$, as the root of $\phi_q(\cdot,\alpha,\beta)=0$.

Let us illustrate this recursive approach by attempting to derive the optimal policy for $Q=2$. Starting again from Equation (10) and assuming that $\theta_L r_L > 1$, we have that $a_t^*=L$ for $t>0$ in the vicinity of zero. Then, $\phi_q(t,\alpha,\beta)$ can be written as a solution to the p.d.e., provided in the

Appendix. Unfortunately, the ten-line expression includes the Beta function and is intractable. To obtain the optimal time to switch from the high-margin to the low-margin channel, $t_2(\alpha,\beta)$, one has to find the root $\phi_2(\cdot,\alpha,\beta)=0$, which this time does not admit a closed-form expression.

A glance at the expression in the Appendix, which concerns the case of only two units of capacity, makes the point that the recursive approach becomes analytically intractable very quickly, so it has to be pursued via numerical methods. While the latter is possible for small instances of the problem, which can be solved to optimality in a reasonable amount of time, the recursive approach is impractical for larger instances. For this reason, in the following section, we develop a heuristic policy that is very easy to implement and, while suboptimal, it performs very well in our numerical experiments.

4.4. A Heuristic Policy for the General Case

Given the analytical and computational challenges in deriving the optimal capacity allocation policy exactly, we propose an efficient heuristic for the general case of the problem that leverages the fact that the solution for the special case $Q=1$ is available in closed form. The main premise on which our heuristic is based is that the optimal switching time of the q^{th} unit of capacity can be approximated reasonably well by employing the closed-form expression for the optimal switching time of the first unit, with appropriately re-scaled shape and rate parameters:

$$t_q(\alpha,\beta)\approx t_1\big(\alpha q^x,\beta q^y\big),$$

for some x and y.

We expect the heuristic policy to maintain a constant ratio of (expected) demand to supply throughout the selling season. For that to be the case at the optimal switching time for the q^{th} and first unit, we need to have that

$$(\alpha/\beta)/q=\alpha q^x/\beta q^y,$$

which implies that $x=y-1$.

We determine the values of x and y numerically: for different values of α, β, θ_L, and r_L, we solve exactly for $t_q(\alpha,\beta)$, for various q's. Then, we estimate the values of x and y by minimizing the Sum of Squared Errors between the exact values, $t_q(\alpha,\beta)$, and our approximations, $t_1\big(\alpha q^x,\beta q^y\big)$. The estimates of x that we get range between .40 and .45. The estimates of y range between 1.62 and 1.89. The estimation with the least minimized error satisfies $y-x=1.17$, which suggests that the condition of $y-x=1$ mentioned before seems reasonable. As representative values we then choose $x=0.5$ and $y=1.5$.

Summarizing, our heuristic policy boils down to the approximation:

$$t_q(\alpha,\beta) \approx t_1\left(\alpha q^{1/2}, \beta q^{3/2}\right).$$

As a sanity check for the proposed heuristic policy, we note that, by solving the p.d.e. numerically as explained in the previous section (for small instances of the problem), we observe that $t_q(\alpha,\beta)$ is monotonically increasing in q, for fixed α and β. The fact that the expression for $t_1\left(\alpha q^{1/2}, \beta q^{3/2}\right)$ is available in closed form, implies that we can verify analytically that the heuristic policy satisfies:

$$\frac{\partial t_1\left(\alpha q^{1/2}, \beta q^{3/2}\right)}{\partial q} > 0.$$

In Figures 4 and 5 we contrast the exact optimal switching time to its approximation through our heuristic, for different values of q, θ_L, and r_L. While not perfect, the approximation seems to be varying between good and very good, depending on the parameter values.

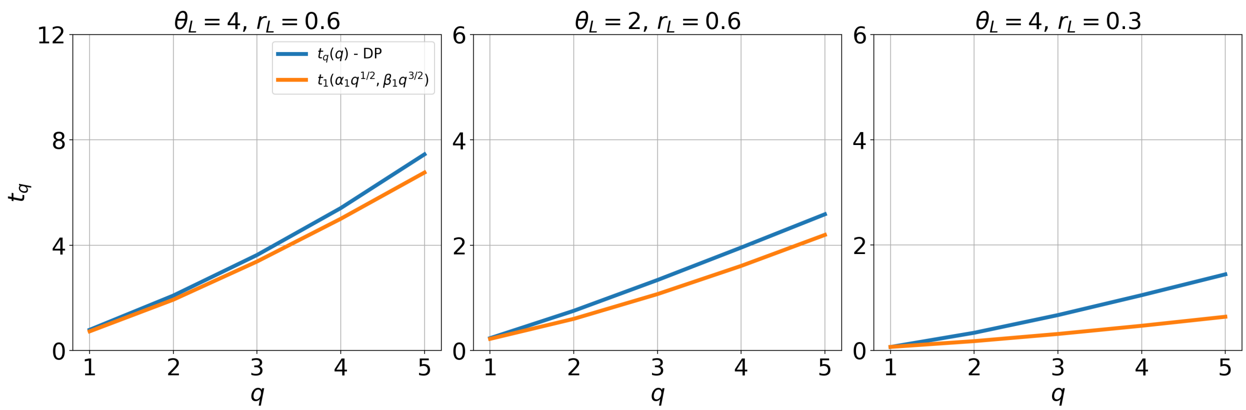

Figure 4 t_q **as a function of** q **for low demand,** $\mathbb{E}[\Lambda] = 0.1$. **Comparison between exact solution and heuristic policy in terms of the optimal switching time.**

5. On the Value of (Bayesian) Demand Learning

Our final set of results pertains to quantifying the value of (Bayesian) demand learning, and the ability of the heuristic policy, introduced in the previous section, to capture the bigger part of that value. To that end, we compare the total expected revenue resulting from the following policies:

A. "Optimal Learning": The exact solution to the Optimal Control problem;

B. "No Learning": The exact solution to a special case of the Optimal Control problem, where there is no uncertainty around the ex ante mean demand, and hence no updating of the shape and rate parameters of the Gamma distribution. Mathematically, this corresponds to the solution to Equation (10) with $\alpha_T = \lambda k$ and $\beta_T = k$, for fixed λ and $k \to \infty$; effectively, in the absence of the partial derivative with respect to β. Equivalently, this is the solution to the Optimal Control problem in the benchmark model of Martínez-de Albéniz et al. (2022);

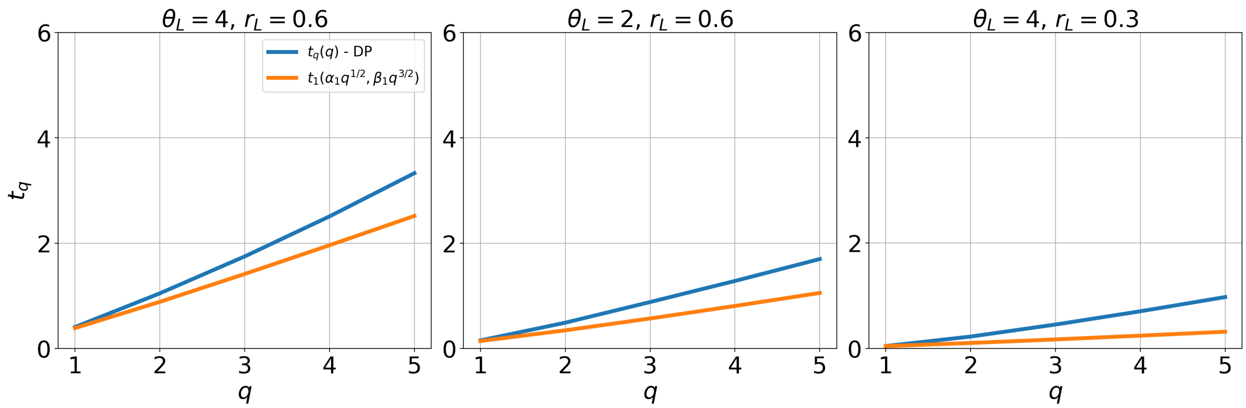

Figure 5 t_q **as a function of** q **for high demand,** $\mathbb{E}[\Lambda] = 0.6$**. Comparison between exact solution and heuristic policy in terms of the optimal switching time.**

C. "No Sharing": $t_q(\alpha, \beta) = 0$, for all q, α, and β, i.e., the seller never employs the low margin, high volume channel;

D. "Heuristic Policy": the heuristic policy presented in Section 4.4.

Figures 6 and 7 present the performance of those policies for $T = 6$; $Q = 1$ and $Q = 5$; $\mathbb{E}[\Lambda_T] = 0.5Q/T$ and $\mathbb{E}[\Lambda_T] = 1.5Q/T$; and β ranging from 0.01 to 10 on a logarithmic scale. The performance of these policies is normalized by the total expected revenue of the optimal policy with demand learning (curve A) which, by definition, is the highest.

The difference between curves A and B quantifies the value of (Bayesian) demand learning, because we compare the optimal allocation policy with dynamic learning to the optimal allocation policy without it. On the other hand, the difference between curves B and C quantifies the value of (optimal) sharing – essentially, the value of the practice of RM – because we have the optimal allocation policy without demand learning against the static allocation policy. We observe that, across all the scenarios considered, the value of learning is of the same order of magnitude as the value of the practice of RM; and actually, much larger than the latter in the regime of high ex ante uncertainty regarding the demand. Hence, for the firms and industries where RM is a first-order consideration, demand learning could very well be too.

Moreover, by comparing curves A and D, we observe that the heuristic policy captures a significant amount of the total expected revenue of the optimal policy for moderate and low levels of demand uncertainty, despite the fact that it is conceptually very simple and computationally extremely fast. Note also that it outperforms the practice of RM without demand learning (curve B) in all scenarios, which reinforces the main message of this section: the benefit from demand learning may outweigh the cost of suboptimal capacity allocation.

Finally, in the regime of high ex ante uncertainty regarding the demand, i.e., small values of β, curves B, C, and D exhibit big gaps from optimality. It is in those situations that we recommend practitioners to lean towards the exact Optimal Control formulation, or to develop more sophisticated heuristic policies.

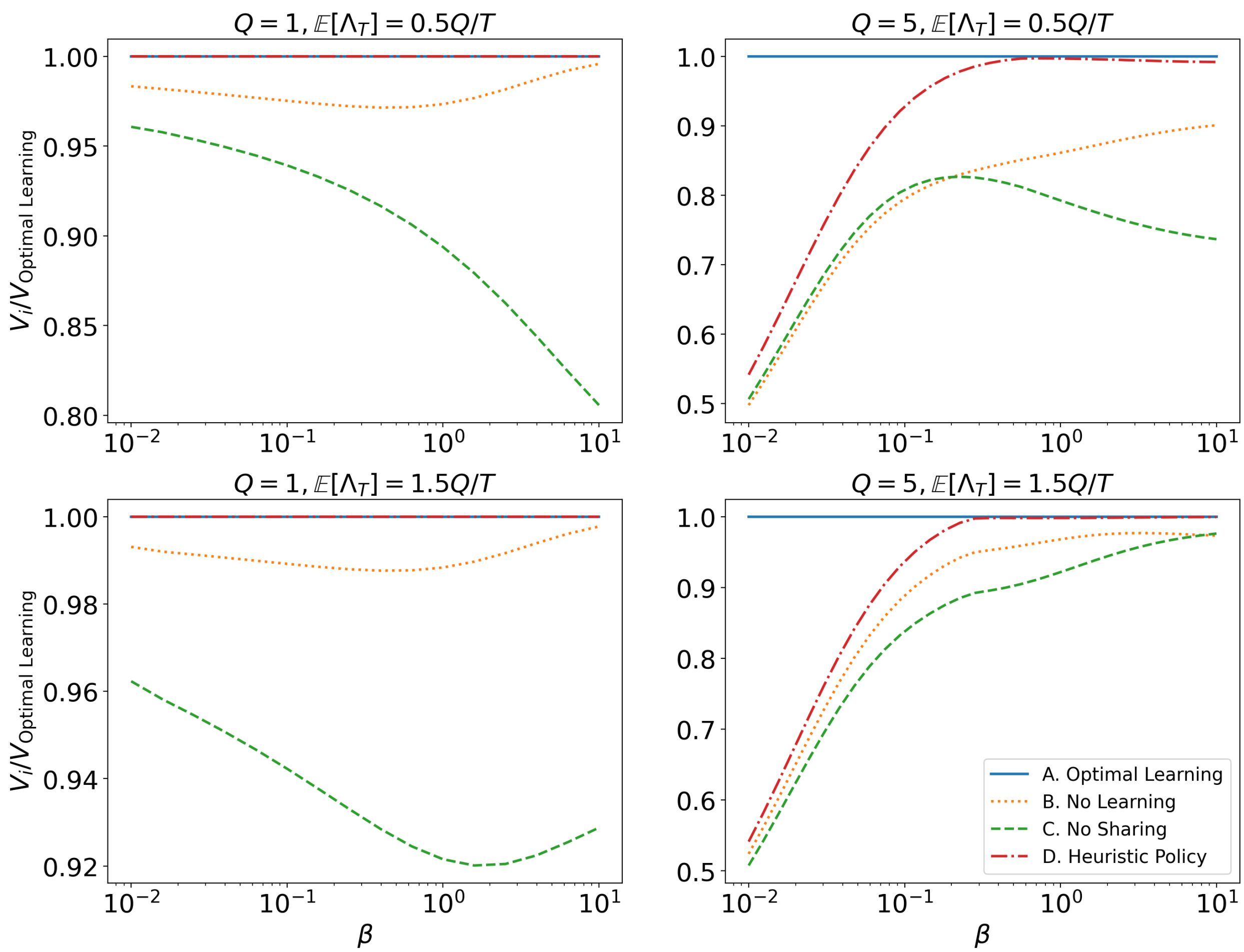

Figure 6 **The performance of the no-learning policy, of the no-sharing policy, and of the heuristic policy against the optimal learning policy with Bayesian demand learning; for high and low $\mathbb{E}[\Lambda]$, high and low Q, $r_L = 0.6$, $\theta_L = 4$, $T = 6$.**

6. Discussion

We conclude the paper with a broader discussion of the approach followed and the results and insights obtained, in an attempt to put them in perspective.

Rapid advances in Information Technology during the last decades have generated an abundance of data in a variety of firms and industries. This has stimulated intense research activity on dynamic learning problems, whether in the supply or in the demand management side of things, combining

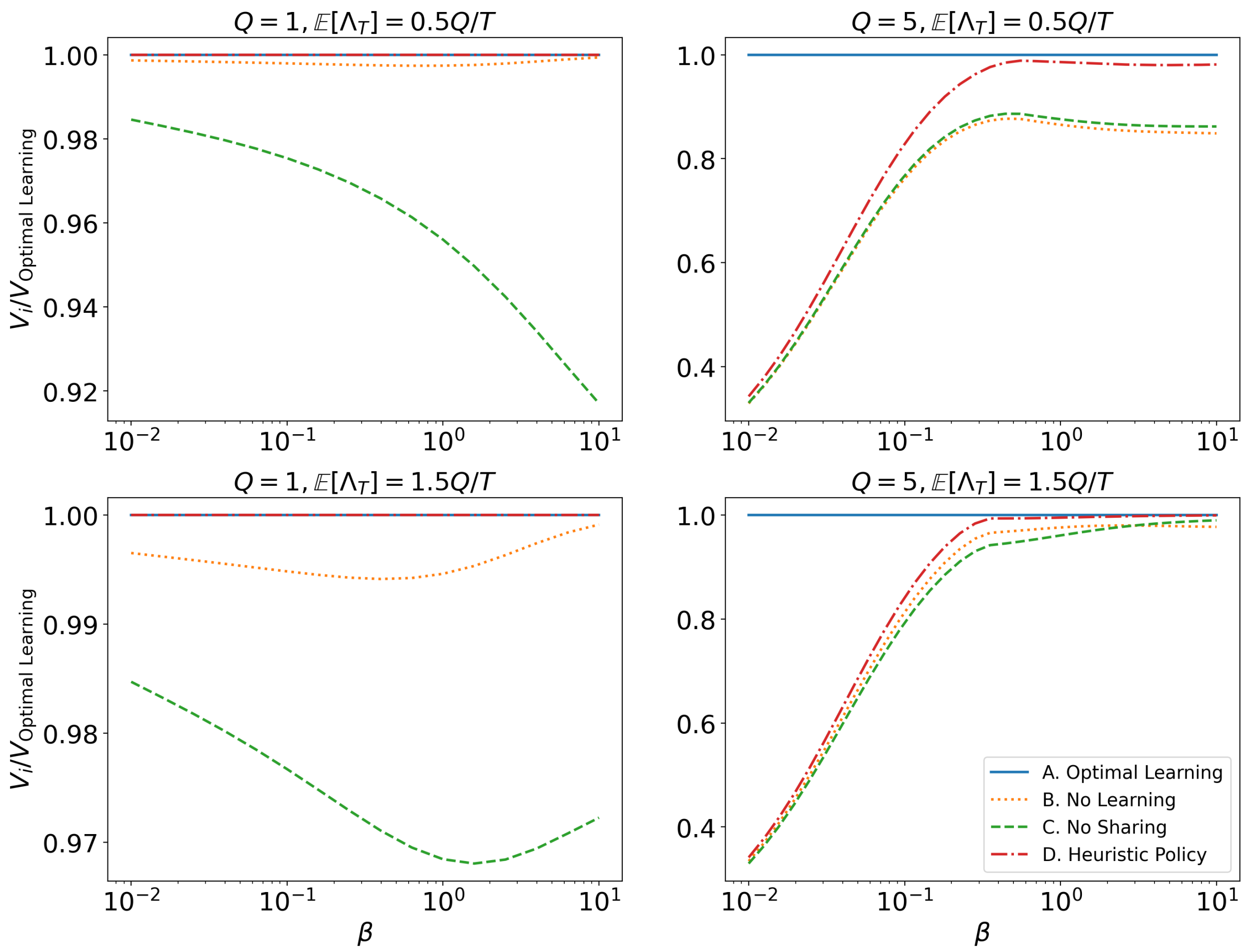

Figure 7 **The performance of the no-learning policy, of the no-sharing policy, and of the heuristic policy against the optimal learning policy with Bayesian demand learning; for high and low $\mathbb{E}[\Lambda]$, high and low Q, $r_L = 0.3$, $\theta_L = 8$, $T = 6$.**

tools from Operations Research with concepts from Economics and methods from Statistics and Machine Learning. RM is one of the academic fields where this interdisciplinary effort has been particularly fruitful. Characteristic of many approaches in this thread of literature is the principle of "estimate, then optimize" whereby the exploration and exploitation phases that the solution to every dynamic learning problem must include, are largely decoupled. While these approaches have been quite successful in different business settings, their applicability to RM problems with limited capacity, which are the ones that motivate our work, is questionable: the opportunity cost of each unit of capacity is very high which, in turn, makes phases of "pure exploration" very costly.

We argue that an exact analysis via a Dynamic Programming/Optimal Control formulation, coupled with Bayesian demand learning, is the correct approach for RM problems where there is limited capacity, limited sales season, and high uncertainty regarding the demand statistics. Capacity allocation problems, e.g., in luxury apparel and resort hotels, have these characteristics.

From an academic standpoint, an argument against the approach that we favor is that Bayesian formulations of dynamic learning problems are typically intractable analytically, and rarely give rise to efficient computational methods. Consequently, we consider among the main contributions of this paper the fact that we identify a special case of the problem, in our regime of interest (limited capacity), where we can provide a closed-form solution to the dynamic learning problem at hand. We leverage this analytical result in three ways: by deriving managerial insights through comparative statics on the optimal solution obtained; by building on it an efficient heuristic policy for the general case, which performs very well in our numerical experiments; and by studying the monetary value of demand learning.

One of our main findings, counterintuitive at first sight, has to do with the fact that the optimal time to switch from the high-margin channel to the low-margin one is non-monotonic with respect to the uncertainty regarding the demand statistics. In other words, higher uncertainty regarding the mean demand may imply switching to the low margin, high volume channel earlier or later. This phenomenon may seem like a mere mathematical curiosity, but has important implications in quantity-based RM with demand learning: conventional wisdom suggests that, under high demand uncertainty, the seller should switch early to the low margin, high volume channel in order to boost sales and learn the demand faster; along the lines of the "estimate, then optimize" principle. This line of reasoning would give rise, in a more realistic setting, to progressive price/margin reductions, akin to the practice of markdowns, prevalent in retailing. Our result implies that a markdown strategy is inappropriate in our setting, because it may be optimal for the seller to reverse a mark-down with a subsequent mark-up on certain occasions.

The final question that we aim to provide an answer for in this work is whether demand learning, if done properly, brings significant monetary value to the firm, or if it should be a second-order consideration from a practical standpoint, and mostly of academic interest. As learning and revenue optimization are intertwined in RM problems with active/dynamic demand learning, the challenge is how to disentangle the value added by the practice of price- or quantity-based RM from the value added by learning. To that end, we leverage the stylized nature of our model and the closed-form solution that we obtain for the special case of (extremely) limited capacity, to answer this question in the affirmative: the monetary value of Bayesian demand learning can be of the same order of magnitude as the practice of RM itself – the tactical optimization of capacity allocation, in our case – if the demand statistics were assumed to be known. We argue, thus, that demand learning could be a first-order consideration for certain firms and industries.

References

Araman, V. F., and R. Caldentey. 2009. Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Research* 57 (5): 1169–1188.

Aviv, Y., and A. Pazgal. 2002. Pricing of Short Life-Cycle Products through Active Learning. Working paper, Washington University in St. Louis.

Azoury, K. S. 1985. Bayes Solution to Dynamic Inventory Models under Unknown Demand Distribution. *Management Science* 31 (9): 1150–1160.

Belobaba, P. P. 1989. OR Practice–Application of a Probabilistic Decision Model to Airline Seat Inventory Control. *Operations Research* 37 (2): 183–197.

Belobaba, P. P. 1992. Optimal vs Heuristic Methods for Nested Seat Allocation. Presentation at ORSA/TIMS Joint National Meeting.

Bertsimas, D., and A. J. Mersereau. 2007. A Learning Approach for Interactive Marketing to a Customer Segment. *Operations Research* 55 (6): 1120–1135.

Besbes, O., J. M. Chaneton, and C. C. Moallemi. 2022. The Exploration-Exploitation Trade-off in the Newsvendor Problem. *Stochastic Systems* 12 (4): 319–339.

Besbes, O., and A. Zeevi. 2009. Dynamic Pricing without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms. *Operations Research* 57 (6): 1407–1420.

Besbes, O., and A. Zeevi. 2012. Blind Network Revenue Management. *Operations Research* 60 (6): 1537–1550.

Besbes, O., and A. Zeevi. 2015. On the (Surprising) Sufficiency of Linear Models for Dynamic Pricing with Demand Learning. *Operations Research* 61 (4): 723–739.

Broder, J., and P. Rusmevichientong. 2012. Dynamic Pricing Under a General Parametric Choice Model. *Operations Research* 60 (4): 965–980.

Brumelle, S. L., and J. I. McGill. 1993. Airline Seat Allocation with Multiple Nested Fare Classes. *Operations Research* 41 (1): 127–137.

Caro, F., and J. Gallien. 2007. Dynamic Assortment with Demand Learning for Seasonal Consumer Goods. *Management Science* 53 (2): 276–292.

Caro, F., and V. Martínez-de Albéniz. 2015. Fast fashion: Business model overview and research opportunities. *Retail supply chain management: Quantitative models and empirical studies*:237–264.

Chang, S. H., and D. E. Fyffe. 1971. Estimation of Forecast Errors for Seasonal-Style-Goods Sales. *Management Science* 18 (2): 89–96.

Chen, L., and E. L. Plambeck. 2008. Dynamic Inventory Management with Learning about the Demand Distribution and Substitution Probability. *Manufacturing & Service Operations Management* 10 (2): 236–256.

Chen, Q. G., S. Jasin, and I. Duenyas. 2019. Nonparametric Self-Adjusting Control for Joint Learning and Optimization of Multiproduct Pricing with Finite Resource Capacity. *Mathematics of Operations Research* 44 (2): 601–631.

Chen, Q. G., S. Jasin, and I. Duenyas. 2021. Technical Note—Joint Learning and Optimization of Multi-Product Pricing with Finite Resource Capacity and Unknown Demand Parameters. *Operations Research* 69 (2): 560–573.

Chen, Y., and V. F. Farias. 2013. Simple Policies for Dynamic Pricing with Imperfect Forecasts. *Operations Research* 61 (3): 612–624.

Chod, J., M. G. Markakis, and N. Trichakis. 2021. On the Learning Benefits of Resource Flexibility. *Management Science* 67 (10): 6513–6528.

Cope, E. 2007. Bayesian Strategies for Dynamic Pricing in E-Commerce. *Naval Research Logistics* 54 (3): 265–281.

Curry, R. E. 1990. Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations. *Transportation Science* 24 (3): 193–204.

Drakopoulos, K., S. Jain, and R. Randhawa. 2021. Persuading Customers to Buy Early: the Value of Personalized Information Provisioning. *Management Science* 67 (2): 828–853.

Eppen, G. D., and A. V. Iyer. 1997. Improved Fashion Buying with Bayesian Updates. *Operations Research* 45 (6): 805–819.

Farias, V. F., and B. Van Roy. 2010. Dynamic Pricing with a Prior on Market Response. *Operations Research* 58 (1): 16–29.

Fisher, M., and A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations research* 44 (1): 87–99.

Gallego, G., and G. Van Ryzin. 1994. Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management science* 40 (8): 999–1020.

Harrison, J. M., N. B. Keskin, and A. Zeevi. 2012. Bayesian Dynamic Pricing Policies: Learning and Earning under a Binary Prior Distribution. *Management Science* 58 (3): 570–586.

Iglehart, D. L. 1964. The Dynamic Inventory Problem with Unknown Demand Distribution. *Management Science* 10 (3): 429–440.

Jasin, S. 2015. Performance of an LP-Based Control for Revenue Management with Unknown Demand Parameters. *Operations Research* 63 (4): 909–915.

Johnson Ferreira, K., D. Simchi-Levi, and H. Wang. 2018. Online Network Revenue Management Using Thompson Sampling. *Operations Research* 66 (6): 1586–1602.

Karlin, S. 1960. Dynamic Inventory Policy with Varying Stochastic Demands. *Management Science* 6 (3): 231–258.

Lariviere, M. A., and E. L. Porteus. 1999. Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales. *Management Science* 45 (3): 346–363.

Lautenbacher, C. J., and S. Stidham Jr.. 1999. The Underlying Markov Decision Process in the Single-Leg Airline Yield Management Problem. *Transportation Science* 34 (2): 136–146.

Lee, T. C., and M. Hersh. 1993. A Model for Dynamic Seat Inventory Control with Multiple Seat Bookings. *Transportation Science* 27 (3): 252–265.

Littlewood, K. 1972. Forecasting and Control of Passenger Bookings. Proceedings of the Twelfth Annual AGIFORS Symposium.

Lovejoy, W. S. 1990. Myopic Policies for Some Inventory Models with Uncertain Demand Distributions. *Management Science* 36 (6): 724–738.

Martínez-de Albéniz, V., C. Pinto, and P. Amorim. 2022. Driving Supply to Marketplaces: Optimal Platform Pricing When Suppliers Share Inventory. *Manufacturing & Service Operations Management* 24 (4): 2367–2386.

Murray, G. R., and E. A. Silver. 1966. A Bayesian Analysis of the Style Goods Inventory Problem. *Management Science* 12 (11): 785–797.

Papanastasiou, Y., and N. Savva. 2017. Dynamic Pricing in the Presence of Social Learning and Strategic Consumers. *Management Science* 63 (4): 919–939.

Robinson, L. W. 1995. Optimal and Approximate Control Policies for Airline Booking with Sequential Nonmonotonic Fare Classes. *Operations Research* 43 (2): 252–263.

Scarf, H. 1959. Bayes Solutions of the Statistical Inventory Problem. *The Annals of Mathematical Statistics* 30 (2): 490–508.

V. den Boer, A., and B. Zwart. 2014. Simultaneously Learning and Optimizing Using Controlled Variance Pricing. *Management Science* 60 (3): 770–783.

V. den Boer, A., and B. Zwart. 2015. Dynamic Pricing and Learning with Finite Inventories. *Operations Research* 63 (4): 965–978.

van Ryzin, G., and J. I. McGill. 2000. Revenue Management without Forecasting or Optimization: an Adaptive Algorithm for Determining Seat Protection Levels. *Management Science* 46 (6): 760–775.

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 (2): 318–331.

Wollmer, R. D. 1992. An Airline Seat Management Model for a Single Leg Route when Lower Fare Classes Book First. *Operations Research* 40 (1): 26–37.

Appendix

Exact Solution for the Special Case $Q=2$

Given that $\phi_1(t,\alpha,\beta)$ is available in closed-form, Equation (10) for $q=2$ can be written as follows:

$$
\begin{aligned}
\frac{\partial\phi_2}{\partial t}(t,\alpha,\beta) =&\theta_L\frac{\partial\phi_2}{\partial\beta}(t,\alpha,\beta)\\
&+\frac{\alpha+1}{\beta}\left\{\theta_L\left(r_L+r_L\left(\frac{\beta}{\beta+t\theta_L}\right)^{\alpha+1}+\frac{r_L-1}{\theta_L-1}+\frac{1-\theta_Lr_L}{\theta_L-1}\right)\right\}\\
&-\frac{\alpha+1}{\beta}\left\{\theta_L\left(r_L+\phi_2(t,\alpha,\beta)+\frac{1-\theta_Lr_L}{\theta_L-1}\right)\right\}\\
=&\theta_L\frac{\partial\phi_2}{\partial\beta}(t,\alpha,\beta)\\
&+\frac{\alpha+1}{\beta}\left\{\theta_Lr_L\left(\frac{\beta}{\beta+t\theta_L}\right)^{\alpha+1}\right\}\\
&-\frac{\alpha+1}{\beta}\left\{\theta_L\left(r_L+\phi_2(t,\alpha,\beta)+\frac{1-\theta_Lr_L}{\theta_L-1}\right)\right\}.
\end{aligned}
$$

Using the boundary condition $\phi_2(0,\alpha,\beta)=-(1-\theta_Lr_L)/(\theta_L-1)$, we obtain the solution to the above p.d.e. in closed form:

$$
\begin{aligned}
\phi_2(t,\alpha,\beta)=&\frac{1}{t\theta_L}\left(r_L(1+\alpha)\beta\left(\frac{\beta}{t\theta_L}\right)^{\alpha}B\left(\frac{t\theta_L+\beta}{t\theta_L},1-\alpha\right)-\right.\\
&-\left(\theta_L^2t^2\alpha-\theta_L^2t^2r_L\alpha+\theta_L^2tr_L\alpha(-\theta_Lt-\beta)^{-\alpha}(-\beta)^{\alpha+1}+\right.\\
&+t\theta_L\alpha\beta-r_L\theta_Lt\alpha\beta+r_Lt\theta_L\alpha(-\theta_Lt-\beta)^{-\alpha}(-\beta)^{\alpha}\beta-\\
&-\theta_Lr_Lt\beta\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}+\theta_L^2r_Lt\beta\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}-\\
&-\theta_Lr_Lt\alpha\beta\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}+\theta_L^2r_Lt\alpha\beta\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}-\\
&-r_L\beta^2\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}+\theta_Lr_L\beta^2\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}-\\
&-r_L\alpha\beta^2\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}+\theta_Lr_L\alpha\beta^2\left(\frac{\beta}{t\theta_L+\beta}\right)^{\alpha}+\\
&+(\theta_L-1)r_L(1+\alpha)(-\theta_Lt-\beta)^{1-\alpha}(-\beta)^{\alpha}\beta\left(\frac{\theta_Lt+\beta}{2\theta_Lt+\beta}\right)^{\alpha}+\\
&+(\theta_L-1)r_Lt\theta_L\alpha(1+\alpha)(-\theta_Lt-\beta)^{-\alpha}(-\beta)^{\alpha}\beta\left(\frac{\beta+t\theta_L}{t\theta_L}\right)^{\alpha+1}\times\\
&\left.\left.B\left(\frac{2t\theta_L+\beta}{t\theta_L},1-\alpha\right)\right)\left((\theta_L-1)\alpha(t\theta_L+\beta)\right)^{-1}\right),
\end{aligned}
$$

where $B(a,b)$ is the beta function. The root of $\phi_2(\cdot,\alpha,\beta)=0$ that determines the optimal switching time is not tractable analytically though.