Pricing to Accelerate Demand Learning in Dynamic Assortment Planning

Masoud Talebian,* Natashia Boland, Martin Savelsbergh

University of Newcastle, Australia

March 16, 2013

Abstract

Retailers, from fashion stores to grocery stores, have to decide what range of products to offer, i.e., their product assortment. New business trends, such as mass customization and shorter product life cycles, make predicting demand more difficult, which in turn complicates assortment planning. We propose and study a stochastic dynamic programming model for simultaneously making assortment and pricing decisions that incorporates demand learning using Bayesian updates. We analytically show that it is profitable for the retailer to give price discounts early on the sales season to accelerate demand learning. Our computational results demonstrate the benefits of such a policy and provide managerial insights that may help improve a retailer's profitability.

keywords: assortment planning, demand learning, Bayesian updating, stochastic dynamic programming, retailing

1 Introduction

The retail industry plays a central role in connecting manufacturers with consumers. Retailers are at the end of the supply chain and form an essential element in a manufacturer's distribution strategy. Assortment planning is integral to a retailer's business and have a significant impact on a retailer's bottom line.

Because of its importance, assortment optimization have received significant attention in the operations management and operations research literature. The majority of research in this area assumes that the relationship between a retailer's assortment decisions and the consumers' purchase decisions is known. In other words, demand – the major factor affecting assortment decisions – is assumed to be a known function of the assortment. However, this full information setting is becoming increasingly uncommon in practice. Mass customization and shorter product life cycles because of rapidly changing consumer preferences, are just a few factors driving the frequent introduction of new products, for which it is hard to predict demand. Because it is becoming more common that there is insufficient historic data for forecasting future demand, it becomes necessary to put greater emphasis on "exploration"

---

*Corresponding author, Email:Masoud.Talebian@newcastle.edu.au, Tel:+61(2)4921-5525

of the market as opposed to "exploitation" of the market. That is, demand learning has to become an integral part of effective assortment management, which is the focus of our research. With learning also comes dynamic decision making. In the presence of full information making decisions once works well, but when demand is learned over time decisions have to be reviewed and revised frequently.

We focus on the size of the market, or the customer arrival rate, as the unknown parameter of the demand function. The size of the market represents the number of people who are interested in a product and are considering purchasing it. Whether they purchase the product or not depends on the price of the product. Our premise is that the distribution of market size can be learned by observing sales, unlike the exact size of the market which is inherently uncertain and cannot be learned. Thus, we advocate and analyze a loop in which at the end of a period demand learning occurs based on observed sales in the period, and the estimate of the demand function is updated before decisions about the assortment for the next period are made. In practice, a substantial amount of uncertainty about the demand process is resolved through early sales information.

One of the most prevalent learning methods is Bayesian updating. Bayesian updating can be used in situations where observations come from a fixed distribution and are used to update information, represented in the form of a prior distribution. In the retailer's context, a manager's belief about the demand function is updated after observing new sales data. Thus, Bayesian updating utilizes both the manager's initial estimate of demand and the observed sales data to revise the demand forecast. We use parametric Bayesian, where the shape of the demand function is assumed to be known, but some of its parameters are unknown.

More specifically, we consider a retailer who has the option to sell a number of different product families and has to decide what assortment of product families to offer and what prices to charge for them. Since we focus on assortment planning at the product family level, we assume that demands are independent and there is no need to consider substitution. This seems reasonable as product families are well differentiated and customers are likely to go to a competitor if a product family is not offered. An assortment decision still has to be made, because insufficient space is available to sell all product families.

The retailer faces a dilemma when making assortment and pricing decisions if it wants to learn more about demand at the same time. On the one hand, the retailer wants to maximize the revenue based on its current belief about demand by charging the optimal price and choosing the most profitable product families, i.e., it wants to exploit the market. On the other hand, the retailer wants to maximize the learning of the true size of the market by manipulating prices and offering different product families, i.e., it wants to explore the market. We develop a stochastic dynamic programming model that optimally balances the exploitation and exploration of the market.

Our main innovation is the use of pricing to accelerate demand learning. Our computational experiments demonstrate that offering carefully chosen price discounts for the express purpose of speeding up demand learning can outperform state-of-the-art demand learning strategies. That is, we investigate the benefits of using pricing to influence the observed sales. Note that there is full information about price-response function and, as a result, the optimal price. However, we show that the retailer is better off charging prices below the optimal price to learn about the market size.

We carefully distinguish between passive and active learning. In passive or off-line learning, the retailer observes sales and uses the observed sales to adjust its belief about the

demand and then uses this adjusted belief about demand to optimize assortment and price decisions. In this setting, no assortment or price decisions are taken with the specific intent to learn about demand. Learning takes place, as observed sales are used to update the belief about demand, but it happens in a passive way. In active or on-line learning, assortment and price decisions may be taken with the specific intent to learn about demand, e.g., including a product family in the assortment to observe the effect on sales or setting a lower price to observe the effect on sales. Our analysis and empirical study shows that both passive and active learning are effective strategies in environments in which there is uncertainty about the size of the market, that active learning is more effective.

The remainder of the paper is organized as following. In Section 2, we discuss assortment, pricing, and learning in more detail and review relevant literature. In Section 3, we propose a stochastic dynamic programming model, discuss its approximate solution, and derive some asymptotic results. In Section 4, we present results of a computational study. Finally, in Section 5, we present final remarks and discuss future research opportunities. For convenience, we will use the term product instead of product family from now on.

2   Related Literature

Assortment planning, also known as product line selection and product portfolio optimization, is concerned with the problem of choosing which products to offer or display or "put on the shelf". Assortment planning is a key element in retail merchandizing, and as Alan (1993) indicates in an early review, it is a vital factor in the final profitability of retailers. Displaying or offering a larger variety of products increases market share, as it attracts a more heterogenous set of customers and satisfies customers' variety-seeking tendencies (see for example Tang (2006)). The need to choose arises because there is a limit on the number of products that can be offered or displayed, i.e., there is "limited shelf space".

Assortment planning is not limited to traditional bricks-and-mortar stores, which have to decide which products to carry in the store. It is crucial too for modern on-line retailers, which have to decide how to allocate the available screen space on their websites. Similar decision situations arise when space to hold safety stocks is limited or when trained and knowledgeable sales staff is in short supply. In the airline industry, and more generally service industries, assortment planning manifests itself in the selection of fare classes to offer. Of course in this case it is not only the shelf space that is limited, i.e., a limited number of fare classes can be offered, but product inventory, i.e., the seats, itself is bounded.

The value of assortment planning is clearly illustrated by Kök and Fisher (2007), who develop an optimization-based methodology and report that their recommendations for a grocery store chain, when compared with the existing approach, result in profit increases of more than 50%. Similarly, Rajaram (2001) use a non-linear integer programming model for assortment planning in a large catalog retailer specializing in women's apparel, and report a profit increase of 40%.

Russell and Urban (2010) extend assortment planning by not only deciding a product's allocated space, but also it's location: it has been shown that the location of a product affects its sales. They consider a setting in which products are categorized as part of a family, and the integrity of a family should be maintained.

An important aspect of the research in the above-mentioned papers, and much of the assortment literature, is modeling product substitution. Product substitution occurs when a customer's preferred product is not offered and the customer decides to purchase a different,

but similar product. See van Ryzin and Mahajan (1999), Mahajan and van Ryzin (2001), Li (2007), and Gallego et al. (2011) for more on product substitution using Multinomial Logit models, Smith and Agrawal (2000) for more on product substitution using exogenous models, and Gaur and Honhon (2006) for more on product substitution using locational choice models. Most of the research related to product substitution in assortment planning assumes that the demand distribution is known in advance and can thus be categorized as *static* assortment planning. We refer to Kök et al. (2009) for an extensive review of the static assortment planning literature with an emphasis on its practical aspects as they arise in retail supply chain management.

In the above papers, it is assumed that the retailer has full knowledge about the demand function. Even though there are new and improved methods to forecast sales, e.g., using inventory and gross margins data as illustrated in Kesavan et al. (2010), the assumption of full information is far from reality in settings where there is limited historic data available, and the retailer needs to learn about demand during sales horizon. When assortment planning is coupled with demand learning, it is categorized as *dynamic* assortment planning. Caro and Gallien (2007) develop a stylized dynamic assortment planning model to study the sale of fashion items, where the size of the market is unknown. They employ a finite horizon multi-armed bandit model with several plays per stage and Bayesian learning. To solve the model, they use Lagrangian relaxation to handle the weakly coupled dynamic programs, an approach that we will also use. The results of the model are converted into a desirability index that can be used to choose the products to display in each period. In related research, Caldentey and Caro (2011) study an assortment planning setting in which a retailer can choose from basic and fashion items and where the vogue is modeled as a stochastic process. Rusmevichientong et al. (2011) pursue a similar line of research and consider an environment in which a retailer chooses an assortment of products so as to maximize profit subject to a capacity constraint. They model the demand under substitution and study both a static case, where the parameters of the demand function are known, and a dynamic case, where the parameters of the demand function are unknown. They develop an adaptive policy that learns the unknown parameters from past data, and that, at the same time, optimizes the profit.

The problem discussed in the recent paper by Posen and Levinthal (2011) has many of the characteristics of dynamic assortment planning. The paper shows how an organization chooses, or should choose, its strategy in a changing environment. (The problem can be viewed as assortment planning with a shelf space of size one.) Learning, more specifically exploration to generate new knowledge about the environment, is incorporated to enhance the quality of decision making. The main message of the work, in the terminology of dynamic assortment planning, is that care has to be taken when incorporating mechanisms to learn the demand when that demand itself may be changing.

Price optimization, or pricing, is concerned with the problem of choosing what to charge for a product when demand is price-sensitive. A high price may turn some customers away, but a low price reduces the profit margin. A major portion of the academic literature focuses on *dynamic* pricing which, adopting the definition of Weatherford and Bodily (1992), arises in situations where a perishable and nonrenewable product has a stochastic demand over a finite period of time. That is, in situations where a given stock of items has to be sold by a certain deadline, and where demand is stochastic and a function of the offered price. The problem is to dynamically adjust the price to maximize the total revenue over the sales horizon. In such a setting, if the demand function is known fully and in advance: the changes

in the optimal price arise only because of the limited inventory and the stochasticity of the demand. Furthermore, the analysis of various models has shown that dynamically changing the price is not necessary to achieve high profits; a fixed-price policy is near-optimal. We refer to Gallego and van Ryzin (1994, 1997) for a detailed analysis. Another reason for adjusting prices is demand learning. We refer to Araman and Caldentey (2009) and Gallego and Talebian (2012) for two recent studies about dynamic pricing with demand learning. For an example of non-parametric learning, we refer the reader to Besbes and Zeevi (2009).

Our paper is different from the above stream of research in the sense that the retailer is not uncertain about the optimal price; the retailer has full information about customers' sensitivity to price. However, the retailer uses price discounts as a mechanism to increase the number of observations and thus to learn more about the market size. The idea of decreasing prices to accelerate information acquisition has been studied in the marketing literature, although it has been primarily focused on a single product (we refer the reader to Braden and Oren (1994) for an example). To the best of our knowledge, we are the first to consider pricing as a leverage to accelerate demand learning in a setting where there exist multiple products and an assortment decision needs to be made.

Finally, there is some existing literature on joint assortment and pricing. Chen and Hausman (2000), for example, formulate a nonlinear integer optimization problem, where the nonlinearity is due to considering a sum of linear ratios. Ratios correspond to the probability of purchasing a product and is derived by dividing each product's attractiveness by sum of the products' attractiveness. Fractional programming properties are used to solve the model. Schön (2010a) extends this work in two respects. First, by assuming that the seller can practice price discrimination and therefore has to decide the price to charge different groups of customers. Second, by introducing fixed costs associated with choosing a product line. The analysis is restricted to the case where prices have to be chosen from a set of discrete values. This restriction is removed in Schön (2010b). Aydin and Ryan (2000) study joint pricing and assortment planning under substitution. Rodriguez and Aydin (2011) study joint assortment and pricing for configurable products under demand uncertainty. Cachon and Kök (2007) study joint assortment and pricing in a competitive setting with two retailers. They compare the prices and variety levels under decentralized management and under centralized management, and show that decentralized management results in higher prices and less variety. The study is extended in Kok and Xu (2011), where it is shown that the order of consumer decisions, i.e., whether consumers first choose the product type or first choose the brand, has a critical effect on the optimal management policy.

From a methodological perspective, we use discrete-time finite-horizon dynamic programming (see Wright (2011) for a short but rigorous introduction and Powel (2007) for a review of approximation methods). The exploration vs. exploitation trade-off is well-known in dynamic programming, where information about a state can only be obtained by visiting it. Therefore, a state can be visited because it is profitable (exploitation), or to gain information about it (exploration). In a landmark paper, Gittins and Jones (1974) show that learning, or what they call the information acquisition problem, and the exploration vs exploitation tradeoff can be reduced to a series of one-dimensional problems using an index policy. Our research extends their work in two respects. First, we allow visiting of more than one state at each stage, corresponding to shelf space larger than one. Second, we allow influencing the profit of a state by deciding on prices.

3 Model and Analysis

We consider a retailer that has the option to sell N products over a sales horizon T. We assume that the customer arrival rates for products are independent and follow a Poisson distribution. We represent the customer arrival rate for product n by θ^n; we can think of θ^n as the size of the market for product n. We represent the normalized price-response function of product n by $d^n(p^n)$, and assume this function is known for all products. We use k_t^n and p_t^n to denote the retailer's assortment and price decisions at time t. That is, k_t^n is a binary variable indicating whether product n is in the assortment at time t ($k_t^n = 1$) or not ($k_t^n = 0$), and $p_t^n \geq 0$ is the price charged for product n at time t. Thus, the expected demand for product n at time t is $\theta^n d^n(p_t^n)$. The number of products that can be part of the assortment is limited to r. We will frequently refer to this limit as the available "shelf space". The retailer's objective is to maximize its total revenue over the sales horizon. For convenience and when it is clear from the context, we drop superscripts and subscripts.

Our model assumes the following:

1. The sales horizon is finite, i.e., the products are perishable;

2. The demand is stationary;

3. The market size for a product is stochastic and follows a Poisson distribution;

4. The variable costs associated with sales are negligible;

5. The decision maker is risk-neutral; and

6. The revenue function $(pd(p))$ is strictly concave.

These assumptions are commonly made in the pricing literature and their supporting arguments can be found there. We refer the reader to Bitran and Caldentey (2003), who present an overview of pricing models, and the references therein. Assumption 4 implies that maximizing revenue is an appropriate objective function and Assumption 5 assures that working with expected revenues is justified. Assumption 6 guarantees that there exists a (unique) revenue maximizing price.

Under *full information*, the assortment and price optimization problem is as follows:

$$R = \max_{p,k} \sum_{t=1}^{T} \sum_{n=1}^{N} k_t^n p_t^n \theta^n d^n(p_t^n)$$

such that

$$\sum_{n=1}^{N} k_t^n \leq r \quad t = 1, ..., T \tag{1}$$

$$p_t^n \geq 0, \; k_t^n \in \{0, 1\} \quad n = 1, ..., N, \; t = 1, ..., T.$$

This problem can be solved easily in two steps. In the first step, the optimal price for each product is determined. The optimal price is independent of the arrival rate and equal to $\hat{p}^n \in \operatorname*{argmax}_p\{pd^n(p)\}$; i.e., $\hat{p}$ denotes the optimal price under the full information setting.

In the second step, the assortment is determined, which involves nothing more than selecting the r products with the highest revenues, i.e., with highest values $\theta^n \hat{p}^n d^n(\hat{p}^n)$.

3.1 Demand learning under an uncertain market size

We investigate the variant in which the size of the market is uncertain, i.e., the parameter θ of the demand function is uncertain, and the retailer has the opportunity to adjust its initial belief about the size of the market during the sales horizon. We represent the retailer's belief about the size of the market at time t as a prior distribution and use the observed sales s_t at time t to update this belief. We assume that the retailer's belief about θ^n is a *Gamma* distribution with parameters (a^n, b^n), and thus $E[\theta^n] = \dfrac{a^n}{b^n}$ and $C.V.[\theta^n] = 1/\sqrt{a^n}$. We choose the Gamma distribution to represent the retailer's belief since it is a conjugate distribution for Poisson and results in closed-form Bayesian updating. We denote the retailer's belief (a^n, b^n) at time t by $(a^n(t), b^n(t))$, and therefore the retailer's belief about θ^n at time t is distributed as $\Gamma(a^n(t), b^n(t))$.

Lemma 1 *If a retailer's belief about the market size for a given product in a given period is distributed as $\Gamma(a, b)$ and the price for the product has been set to p, then after observing sales s in the period and applying Bayesian updating, the retailer's belief will update to:*

$$\Gamma(a + s, b + d(p)).$$

Proof: All proofs can be found in the appendix. ■

Note that the retailer's knowledge concerning the demand function is affected both by its initial belief (demand estimation) and by the observed sales (demand learning). Since we have to capture uncertainty and information being revealed over time, a stochastic dynamic programming model seems appropriate. Counting time backward, so that t represents the remaining time, we let $(a(T), b(T))$ represent the retailer's initial belief and let $R_t(a(t), b(t))$ represent the expected revenue from period t to the end of the time horizon, conditioned on the current belief. Assuming $R_0(a(0), b(0)) = 0$ for all states, the Bellman's equation corresponding to the stochastic dynamic program is as follows:

$$R_t(a(t), b(t)) = \max_{p_t, k_t}\left(\sum_{n=1}^{N} k_t^n p_t^n \frac{a^n(t)}{b^n(t)} d^n(p_t^n) + E_{s_t}[R_{t-1}(a(t-1), b(t-1))]\right)$$

such that

$$\sum_{n=1}^{N} k_t^n \leq r \quad t = 1, ..., T$$
$$p_t^n \geq 0, \ k_t^n \in \{0, 1\} \quad n \in N, \ t = 1, ..., T \qquad (2)$$
$$a(t-1) = a(t) + k_t s_t$$
$$b(t-1) = b(t) + k_t d(p_t).$$

For each product n, s_t^n is taken to follow a Poisson$\left(\dfrac{a^n(t)}{b^n(t)} d^n(p_t^n)\right)$ distribution if $k_t^n = 1$ and to be zero otherwise.

The objective function consists of two parts. The first part corresponds to the revenue at time t and the second part corresponds to the expected revenue, with respect to the observed sales, in the remainder of the sales horizon. The last two transition equations capture the updating of the retailer's belief.

We consider three options for "solving" this assortment and price optimization problem with demand learning.

The simplest, but most naive, option is for the retailer to assume that its initial belief is correct and use $a(T)$ and $b(T)$ throughout the sales horizon. We refer to this option as the "no-learning" policy. The retailer focuses on exploiting the market under the assumption that its belief about the market size is correct, i.e., the retailer reverts back to solving problem 1 and chooses the r products with the largest values of $\dfrac{a^n(T)}{b^n(T)}\hat{p}^n d(\hat{p}^n)$. In this option, products are chosen greedily based on the initial belief.

The second, more intelligent, option is for the retailer to update its belief about the market size at time t, i.e, $\dfrac{a(t)}{b(t)}$, using observed sales s_t, and adjust the assortment and prices based on the updated belief about the market size, i.e., for time $t-1$ choose the r products with the largest values of $\dfrac{a^n(t-1)}{b^n(t-1)}\hat{p}^n d(\hat{p}^n)$. We refer to this option as the "passive learning" policy. The retailer updates its belief based on observed sales, and thus learns, but does not actively seek new knowledge and does not incorporate any exploration. The retailer's decisions are not influenced by a desire or intent to learn more about the market size; learning is passive. In this option, products are chosen greedily too, but based on the current belief.

The third option is to solve, or approximately solve, the stochastic dynamic program (2). We refer to this option as the "active learning" policy. When making assortment and price decisions, the retailer considers both exploring the market to learn more about demand and exploiting the market to generate revenue.

Stochastic dynamic programs are notoriously hard to solve analytically and numerically. We next discuss the steps we have taken to solve the problem approximately.

3.2 Deriving the active learning policy

Limited self space is the only problem characteristic that links the products. Therefore, by dualizing the shelf space constraint and solving the Lagrangian dual problem instead, we can treat the products independently. Note that the shelf space constraint is a cardinality knapsack constraint and that a cardinality knapsack constraint has the integrality property, so solving the Lagrangian dual is likely to provide a tight approximation.

We can write the Lagrangian relaxation as follows, where $\lambda_t(a(t), b(t))$ is the Lagrangian multiplier corresponding to shelf space constraint for time t:

$$H_t^{\lambda_t(a(t),b(t))}(a(t), b(t)) =$$

$$\max_{p_t,k_t}\{\sum_{n=1}^{N} p_t^n k_t^n \frac{a^n(t)}{b^n(t)} d^n(p_t^n) + \lambda_t(a(t),b(t))(r - \sum_{n=1}^{N} k_t^n) +$$

$$E_{s_t}[H_{t-1}^{\lambda_{t-1}(a(t)+k_t s_t, b(t)+k_t d(p_t))}(a(t) + k_t s_t, b(t) + k_t d(p_t))]\} =$$

$$r\lambda_t(a(t), b(t)) + \max_{p_t,k_t}\{\sum_{n=1}^{N} \left(p_t^n \frac{a^n(t)}{b^n(t)} d^n(p_t^n) - \lambda_t(a(t),b(t))\right) k_t^n +$$

$$E_{s_t}[H_{t-1}^{\lambda_{t-1}(a(t)+k_t s_t, b(t)+k_t d(p_t))}(a(t) + k_t s_t, b(t) + k_t d(p_t))]\}.$$

The corresponding Lagrangian dual problem may be thought of as seeking $\lambda_t(a(t), b(t))$ values so as to minimize $H_T^{\lambda_T(a(T),b(T))}(a(T), b(T))$. We solve this Lagrangian dual problem

approximately using a rolling horizon approach. That is, we solve the Lagrangian dual problem sequentially, period by period, where in each period, we make decisions for the remaining periods, but only implement the decisions for the current period. Furthermore, when solving the Lagrangian dual problem for a given period, we make several approximations (i.e., each time the Lagrangian dual problem is solved for a given period, a new set of approximations is made):

1. We assume that the cost of shelf space does not depend on the retailer's belief concerning demand and does not depend on the time remaining in the sales horizon, i.e., $\lambda_t(a(t), b(t)) = \lambda$. This gives

$$
\begin{aligned}
H_t^\lambda(a(t), b(t)) &= r\lambda + \max_{p_t, k_t}\{\sum_{n=1}^{N}(p_t^n \frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda)k_t^n + E_{s_t}[H_{t-1}^\lambda(a(t) + k_t s_t, b(t) + k_t d(p_t))]\} \\
&= r\lambda + \sum_{n=1}^{N} H_{t,n}^\lambda(a^n(t), b^n(t))
\end{aligned}
$$

where

$$
H_{t,n}^\lambda(a^n(t), b^n(t)) =
$$
$$
\max\{\max_{p_t^n}\{p_t^n \frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + E_{s_t^n}[H_{t-1,n}^\lambda(a^n(t)+s_t^n, b^n(t)+d^n(p_t^n))]\}, H_{t-1,n}^\lambda(a^n(t), b^n(t))\} =
$$
$$
[\max_{p_t^n}\{p_t^n \frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + E_{s_t^n}[H_{t-1,n}^\lambda(a^n(t) + s_t^n, b^n(t) + d^n(p_t^n))]\}]^+.
$$

The value $H_{t,n}^\lambda(a^n(t), b^n(t))$ can be interpreted as the expected revenue from product n from time period t onwards, assuming a belief of $(a(t), b(t))$. The two terms that are being compared in the maximum operation represent including product n in the assortment at time t, $(k_t^n = 1)$, and not including product n in the assortment at time t, $(k_t^n = 0)$. Including the product in the assortment results in immediate revenue and future revenue. Not including the product in the assortment can only result in future revenue. In fact, with fixed shelf space cost, if it is not optimal to include the product in the assortment now, it will not be optimal to include the product in the assortment later, hence the second equality. (See Caro (2005), Chapter 3, for a rigorous proof of this assertion in a similar setting.)

We next focus our attention on

$$
\tilde{R}_{t,n}^\lambda(a^n(t), b^n(t)) = \max_{p_t^n}\{p_t^n \frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + E_{s_t^n}[H_{t-1,n}^\lambda(a^n(t) + s_t^n, b^n(t) + d^n(p_t^n))]\}.
$$

The magnitude of λ determines if $H_{t,n}^\lambda(a^n(t), b^n(t))$ is equal to $\tilde{R}_{t,n}^\lambda(a^n(t), b^n(t))$ or equals zero.

2. We approximate $\tilde{R}_{t,n}^\lambda(a^n(t), b^n(t))$ by employing a one-period look-ahead to estimate future revenue. This gives

$$
\tilde{R}_{t,n}^\lambda(a^n(t), b^n(t)) \cong \max_{p_t^n}\{p_t^n \frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + (t-1)E_{s_t^n}[(\max_{p_{t-1}^n}\{p_{t-1}^n \frac{a^n(t-1)}{b^n(t-1)}d^n(p_{t-1}^n) - \lambda\})^+]\}.
$$

9

The one-step look-ahead captures the first order trade-off between exploration and exploitation, and a cursory investigation showed us that a more detailed approximation, e.g., a two-step look-ahead, does not change the results significantly. The maximization problem inside the expectation can be simplified by observing that $a^n(t-1) = a^n(t)+s_t^n$, $b^n(t-1) = b^n(t) + d^n(p_t^n)$, and λ are all independent of p_{t-1}^n. Thus this term can be written as

$$\max_{p_{t-1}^n}\{p_{t-1}^n\frac{a^n(t-1)}{b^n(t-1)}d^n(p_{t-1}^n)-\lambda\} = \frac{a^n(t-1)}{b^n(t-1)}\max_{p_{t-1}^n}\{p_{t-1}^n d^n(p_{t-1}^n)\}-\lambda = \frac{a^n(t-1)}{b^n(t-1)}\hat{p}^n d^n(\hat{p}^n)-\lambda$$

where $\hat{p}^n$ is the (time-independent) maximizer of $p^n d^n(p^n)$. Therefore, we have

$$\begin{aligned}\tilde{R}_{t,n}^\lambda(a^n(t),b^n(t)) &\cong \max_{p_t^n}\{p_t^n\frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + (t-1)E_{s_t^n}[(\hat{p}^n\frac{a^n(t-1)}{b^n(t-1)}d^n(\hat{p}^n) - \lambda)^+]\}\\ &= \max_{p_t^n}\{p_t^n\frac{a^n(t)}{b^n(t)}d^n(p_t^n) - \lambda + (t-1)E_{s_t^n}[(\hat{p}^n\frac{a^n(t)+s_t^n}{b^n(t)+d^n(p_t^n)}d^n(\hat{p}^n) - \lambda)^+]\}.\end{aligned}$$

3. Computing the expectation in the approximation for $\tilde{R}_{t,n}^\lambda(a^n(t),b^n(t))$ involves the distribution of s_t^n, and is complicated by the $(\cdot)^+$-function (otherwise the result would simply be a linear function of the expected value of s_t^n). We deal with this difficulty as follows. Observe that $\dfrac{s_t^n}{d^n(p_t^n)}$ has a Poisson distribution with expected rate $\dfrac{a^n(t)}{b^n(t)}$, the parameters of the Gamma distribution. It can be shown, a priori, that $\dfrac{s_t^n}{d^n(p_t^n)}$ has a Negative Binomial distribution with parameters $(a^n(t), \dfrac{b^n(t)}{1+b^n(t)})$. To make analysis and numerical computations easier, we approximate the Negative Binomial distribution with a Normal distribution, because the expected value of the positive part of a Normal distribution has a closed form, and approximate s_t^n with a Normal distribution with parameters

$$(\frac{a^n(t)}{b^n(t)}d^n(p_t^n), \frac{a^n(t)(b^n(t)+1)}{b^n(t)^2}d^n(p_t^n)^2),$$

which is appropriate if $a^n(t)$ is large enough. Representing a random variable with standard Normal distribution by z, we obtain

$$\tilde{R}_{t,n}^\lambda(a^n(t),b^n(t)) \cong R_{t,n}^\lambda(a^n(t),b^n(t)) = \max_{p_t^n} g_{t,n,a^n(t),b^n(t)}^\lambda(p_t^n) \qquad (3)$$

where

$$\begin{aligned}g_{t,n,a,b}^\lambda(p) &= p\frac{a}{b}d^n(p) - \lambda + (t-1)v_{n,a,b}(p)E[(z - y_{n,a,b}^\lambda(p))^+],\\ v_{n,a,b}(p) &= \frac{\sqrt{a(b+1)}\hat{p}^n d^n(\hat{p}^n)}{b}\frac{d^n(p)}{b+d^n(p)}, \text{ and}\\ y_{n,a,b}^\lambda(p) &= \frac{(\lambda - \frac{a}{b}\hat{p}^n d^n(\hat{p}^n))}{v_{n,a,b}(p)}.\end{aligned}$$

We will refer to $R_{t,n}^\lambda(a^n(t),b^n(t))$ as the "revenue potential" of product n at time t for shelf space cost λ.

As an aside, we note that the value of a Negative Binomial distribution with parameters (n, p) at point x can be interpreted as the probability of observing x failures before observing n successes when the probability of success is p. There exists evidence in the literature to support a Negative Binomial distribution for arrivals, e.g., Smith and Agrawal (2000) who study static assortment planning under full knowledge. The negative binomial also results from a sequence of Poisson arrivals, where each arrival demands a random quantity according to logarithmic distribution.

We are now in a position to approximately solve the Lagrangian dual, because we have the following proposition.

Proposition 1 *A product's revenue potential $R_{t,n}^{\lambda}(a^n(t), b^n(t))$ is well-defined, as the maximum in (3) exists and is decreasing in the shelf space cost λ.*

Because of Proposition 1 and because the maximization in (3) is relatively easy to compute, we can approximately solve the Lagrangian dual by increasing λ, starting from zero until at most r products have nonnegative revenue potential. In other words, we choose the cost of shelf space, λ, so that all available shelf space is used. If there are fewer than r products with positive revenue potential for $\lambda = 0$, then it is not optimal to use all available shelf space. Note that while available shelf space is exogenous in our model, the optimal value of λ, i.e., the shadow price of shelf space, can guide the retailer in decisions regarding the increase of shelf space.

Our active learning approach can thus be defined as follows. For a given t, and given state $(a(t), b(t))$, we choose $\lambda_t(a(t), b(t)) \geq 0$ as large as possible so that at least r products have non-negative revenue potential, i.e. for at most r of the n's, $R_{t,n}^{\lambda_t(a(t),b(t))}(a^n(t), b^n(t)) \geq 0$, where if this is not possible, we take $\lambda_t(a(t), b(t)) = 0$. Include in the assortment all products with strictly positive revenue potential, and, if possible, make up the remainder up to r with any products that have zero revenue potential, chosen arbitrarily. Now for each product n in the assortment, choose its price $(p_t^n)^*$ to be the value of p_t^n which achieves the maximum in (3) for $\lambda = \lambda_t(a(t), b(t))$.

The stochastic dynamic program (2) is hard to solve analytically, since the optimal policy does not have a closed form. It is hard to solve computationally, since the states a and b are continuous. The three different policies – the no-learning, passive learning, and active learning policies – can be viewed as three different approximations for solving the stochastic dynamic program (2). In the first policy, the no-learning policy, the transition equations are ignored and (a, b) are approximated as fixed over time. In this policy, the assortment and the prices do not change. The passive and active learning policies differ in how the revenue potential is approximated. The passive and active learning policies employ zero-period and one-period look-ahead estimations, respectively. This results in fixed prices and dynamic assortment under the passive learning policy, because new information only changes retailer's belief about size of the market which does not affect the optimal price to be charged. However, under active learning both prices and assortment are dynamic; where prices dynamic is due to the retailer's active seek for new information.

3.3 Analytic results and insights on active learning

In what follows, we study the properties of the function $g_{t,n,a,b}^{\lambda}(p)$ introduced in (3) and of its maximization over p, providing consequent insights about the active learning prices that

result and their sensitivity to parameters. We consider an arbitrary product n and so take n to be fixed. We use p^* to denote the maximizer of $g^{\lambda}_{t,n,a,b}(p)$, for fixed parameters λ, t, a, and b (and n). We also use $\hat{p}$ to denote the price under the full information setting, i.e. to denote the maximizer of $pd^n(p)$.

Proposition 2 *Prices under the active learning policy (p^*) are lower than prices in the full information setting $(\hat{p})$, i.e. $p^* \leq \hat{p}$.*

Proposition 2 confirms our intuition that a retailer may give price discounts under active learning to accelerate exploration of the size of the market. In other words, the discounts reflect the price of exploration for the retailer. A lower price is more informative because it results in a higher demand and more observations, and, as a result, in less uncertainty regarding the updated belief.

The following propositions provide insight into the asymptotic behavior of revenue potential and price for a single product at a specific time and with a specific shelf space cost. In particular, we explore the behavior of $R^{\lambda}_{t,n}(a,b) = \max_{p} g^{\lambda}_{t,n,a,b}(p)$ and the maximizer p^*, as each of the remaining time t, the market size C.V. $1/\sqrt{a}$ and the expected market size a/b increase. We illustrate each with a sample product for which $d(p) = e^{-.5p}$, and hence $\hat{p} = 2$, with $\lambda = 2$ fixed.

Proposition 3 *As the remaining sales horizon (t) goes to infinity, the price (p^*) converges to zero and the revenue potential (R) converges to infinity.*

Figure 1 shows the sensitivity of price and revenue potential to the time remaining, t, for the sample product, with $\frac{a}{b} = 2$ and $a = 2$ fixed. We see that price is a non-increasing function of the remaining sales horizon; as the time remaining increases so does the price discount. This is not surprising, because when there is more time remaining, the retailer should invest more in exploring the market size so as to increase future revenues. This is also reflected in the product's revenue potential.

Proposition 4 *As the C.V. of the belief about arrival rate $(1/\sqrt{a})$ goes to infinity, the price (p^*) converges to the full information optimal price $(\hat{p})$ and the revenue potential (R) converges to infinity.*

Figure 2 shows the sensitivity of price and revenue potential to the arrival rate C.V., which represents the retailer's uncertainty regarding its belief of the market size, for the sample product, with $\frac{a}{b} = 2$ and $t = 4$ fixed. We see that the relationship between price and uncertainty is not monotone. As the retailer's uncertainty increases, the price initially decreases but eventually increases back to the full information optimal price. This is likely the result of two competing factors simultaneously impacting the price. On the one hand, the value of learning increases when the retailer is more uncertain, and therefore the retailer is more willing to decrease the price to accelerate learning. On the other hand, learning is easier when the retailer is more uncertain, and therefore the retailer does not need to decrease the price much to learn a lot. The first factor pushes the price down and the second factor keeps it close to the full information optimal price. At beginning the first factor is more powerful, but at some point the second factor overtakes the first one. We also observe that as the belief uncertainty increases, the revenue potential increases. This can be explained
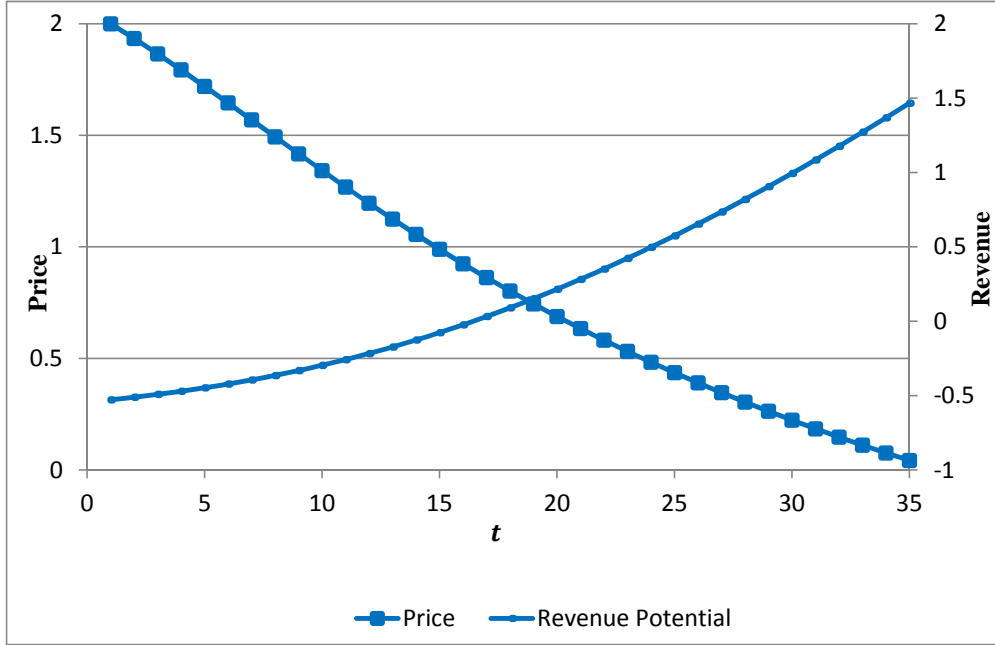
Figure 1: Sensitivity of price and revenue potential to the time remaining.

by noting that if a product turns out not to be profitable, then the retailer has the option to exclude it from the assortment. This means that when computing the expected future revenue, the retailer only needs to consider products with positive revenue potential, so a product with more uncertainty becomes more attractive.

Proposition 5 *As the arrival rate expectation* (a/b) *goes to infinity, the price* (p^*) *converges to the full information optimal price* $(\hat{p})$, *and the revenue potential* (R) *converges to infinity.*

Figure 3 shows the sensitivity of price and revenue potential to the arrival rate expectation for the sample product, with $a = 2$ and $t = 4$ fixed. We see that as the expected arrival rate increases, the retailer has a higher willingness to explore and learn about the market size. However, the relationship between price and arrival rate expectation is not monotone. Intuitively, it is clear that for an extremely large market, with infinite demand, there is no incentive for the retailer to sacrifice the optimal price in the hope of learning more about the size of the market. Similarly, for an extremely small market, with hardly any demand, there is no incentive for the retailer to sacrifice the optimal price either. In between these extremes the value of learning drives down the price.

3.4 An extension to set-up costs

In some settings, there can be costs associated with including a product in the assortment. If such costs are incurred in each period in which the product is in the assortment, they can easily be accommodated in our approach: they play a similar role to the shelf space cost.

In other situations, such costs are incurred only the first time a product is included in the assortment, e.g., contract negotiation and legal fees or the acquisition of product expertise. In these situation, we need to keep track of whether a product has been included in the assortment before or not, with the addition of a new state space variable. Representing this
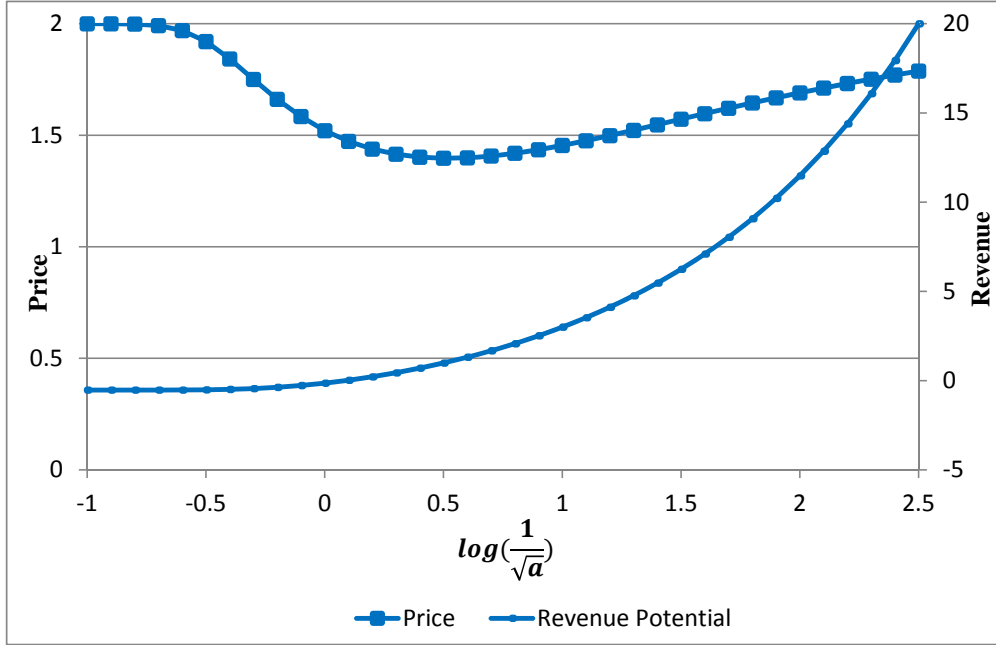
Figure 2: Sensitivity of price and revenue potential to the arrival rate C.V.

variable by $f^n(t)$, taken to be 1 if product n has not been included in the assortment in periods prior to t (or $t = T$), and 0 otherwise, we have

$$R_{t,n}^\lambda(a(t), b(t), f(t)) = \max_{p_t^n}\{p_t^n \frac{a^n(t)}{b^n(t)} d^n(p_t^n) - \lambda - f^n(t)F^n + (t-1)vE[(z-y)^+]\},$$

where F^n is the set-up cost for product n.

We can also model situations where the fixed costs need to be paid in periods which a product is included in assortment after not being in the assortment, like decoration costs. Similar variables to the $f^n(t)$ can be defined to represent whether a product has been included in the assortment in the last period or not.

Whilst our solution approach can be applied in either of these two set-up cost settings as described above, the assumption of a fixed shelf space cost becomes less realistic as the competition for shelf space increases over time as more and more products try to recapture the fixed cost incurred.

4 Policy Comparison

We use numerical simulation to analyze computationally the value of the active learning and providing price discounts.

Consider a situation in which the retailer has the option to sell three products with the characteristics given in Table 1.

Furthermore, let the shelf space available allow at most one product to be offered in each period. We see that the retailer believes that the size of the market for Product 3 is smaller than the size of the market for Product 1 and 2, and that the retailer's confidence in its belief is the same for all products. We also note that the demand for Product 1 is less price
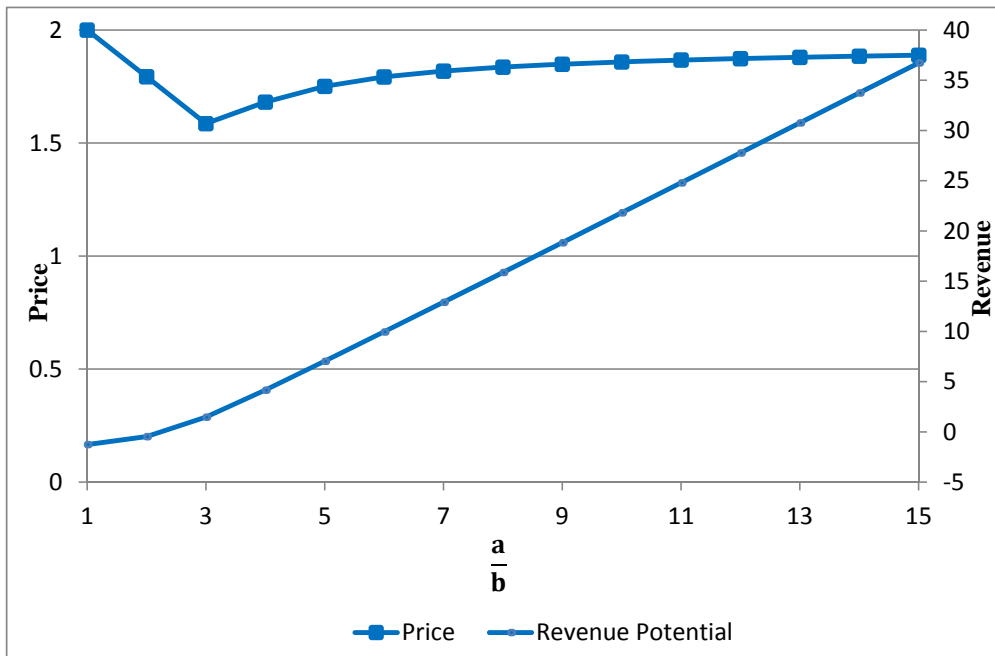
14

Figure 3: Sensitivity of price and revenue potential to the arrival rate expectation.

Table 1: Product characteristics

Product	$d(p)$	$a(T)$	$b(T)$	$a(T)/b(T)$	$\hat{p}$
1	$e^{-.5p}$	.5	.025	20	2
2	e^{-p}	.5	.025	20	1
3	e^{-p}	.5	.05	10	1

sensitive than the demand for Product 2 and Product 3. We have chosen a modest level of uncertainty, and expect that the benefit of learning increases as the initial uncertainty increases.

If the retailer is completely confident in its current belief, and therefore has no reason to explore and learn more about the market size and to change its decisions over time, it will choose to offer Product 1 in all periods at optimal price of \$2. If the true market size happens to be (9,20,30), then the expected revenue will be \$6.62 per period, whereas with the optimal decision, i.e., offering Product 3 at a price of \$1, the expected revenue will be \$11.04 per period. Thus, the lack of information (i.e., the unfounded confidence) is costing the retailer about 40% in terms of expected revenue.

To investigate this situation further, we have conducted the following simulation experiment. Based on the true market size of (9,20,30), we create 10,000 sample paths of realized sales over the sales horizon and observe assortment and pricing decisions under different policies. Figure 4 focuses on the revenues and the assortment and pricing decisions through time under active learning for a sales horizon of 12 periods. In the top graph, we see that the average revenue increases substantially in initial periods and slightly in subsequent periods. This is due to both an improving understanding of the market size, and thus of expected

revenues, and less exploration towards the end of the sales horizon, which implies smaller price discounts. In the middle graph, we see that in all sample paths Product 1 is chosen in the first period, Product 2 is chosen in the second period, and depending on the observed sales in the first two periods, Product 3 is chosen in about a quarter of sample paths in the third period and its share increases to about half by the last period. We also observe that active learning recognizes the revenue potential of Product 3 and switches to it over time. This switch will not happen with passive learning, because the retailer's initial belief about the profitability of Product 3 is less than the true profitability of Product 2. Furthermore, we see that a planning horizon of 12 periods is not long enough for the assortment to converge to Product 3 in all sample paths. In the bottom graph, we see that whenever a product is offered for the first time, a high discount is offered (a discount with respect to the optimal price) and that the discount decreases as the retailer learns more about the market size and as the remaining sales horizon gets shorter.

Figure 5 compares revenues of different policies. In addition to policies already introduced, we also report the revenue of the variant of the active learning policy in which the price is always taken to be the optimal price. This policy corresponds to the policy of Caro and Gallien (2007). We observe that using pricing to accelerate learning results in higher revenues. Active learning results in a 13.7% increase in revenue over passive learning, and active learning with pricing to accelerate learning results in a 16.3% increase in revenues over passive learning. Note that we report revenues and that because of the small profit margins in the retail industry, the differences will be more pronounced when considering profits.

Figure 6 shows the effectiveness of active learning in terms of closing the gap between the revenue in the no learning and full information settings. In other words, the graph shows the ability of the active learning policy to recapture the cost of lack of information. We see that as the sales horizon increases and the retailer has more time to learn, the benefit of active learning increases.

The above observations obviously depend on the true market size and for a different market size the results will be different. Therefore, in the second set of simulation experiments, we assume that the retailer's belief about the market size, represented by a Gamma distribution with parameters a and b, is correct, we draw a market size from that distribution (which will be unknown to the retailer) and we observe a single sample path of realized sales over the sales horizon and repeat this process 10,000 times. When presenting results, we average over the 10,000 repetitions. The simulation results are summarized in Figure 7. As before, we see that the value of learning increases as there is more time to learn. For a sales horizon of 20 periods, the active learning policy gives average revenues of $18.15, which corresponds to recapturing 80% of the cost of the lack of information.

A different way of evaluating the benefits of learning is to look at how the retailer's belief evolves over time. Figure 8 shows how the error of the retailer's belief evolves over time under the active learning policy for given levels of initial belief uncertainty, where the error of the retailer's belief is measured as the absolute difference between its belief about the market size and the true market size. As we observe, as the retailer becomes less certain about its initial belief the learning becomes less beneficial.

Finally, Figure 9 shows the effect of set-up costs (when set-up costs are incurred only the first time a product is offered) on demand learning. We see that as the set-up cost increases, then the effectiveness of learning in reducing the belief error decreases. As the cost of learning increases, because of the costs associated with the set-up of an as of yet unexplored product, less exploration takes place and the focus shifts to exploitation.
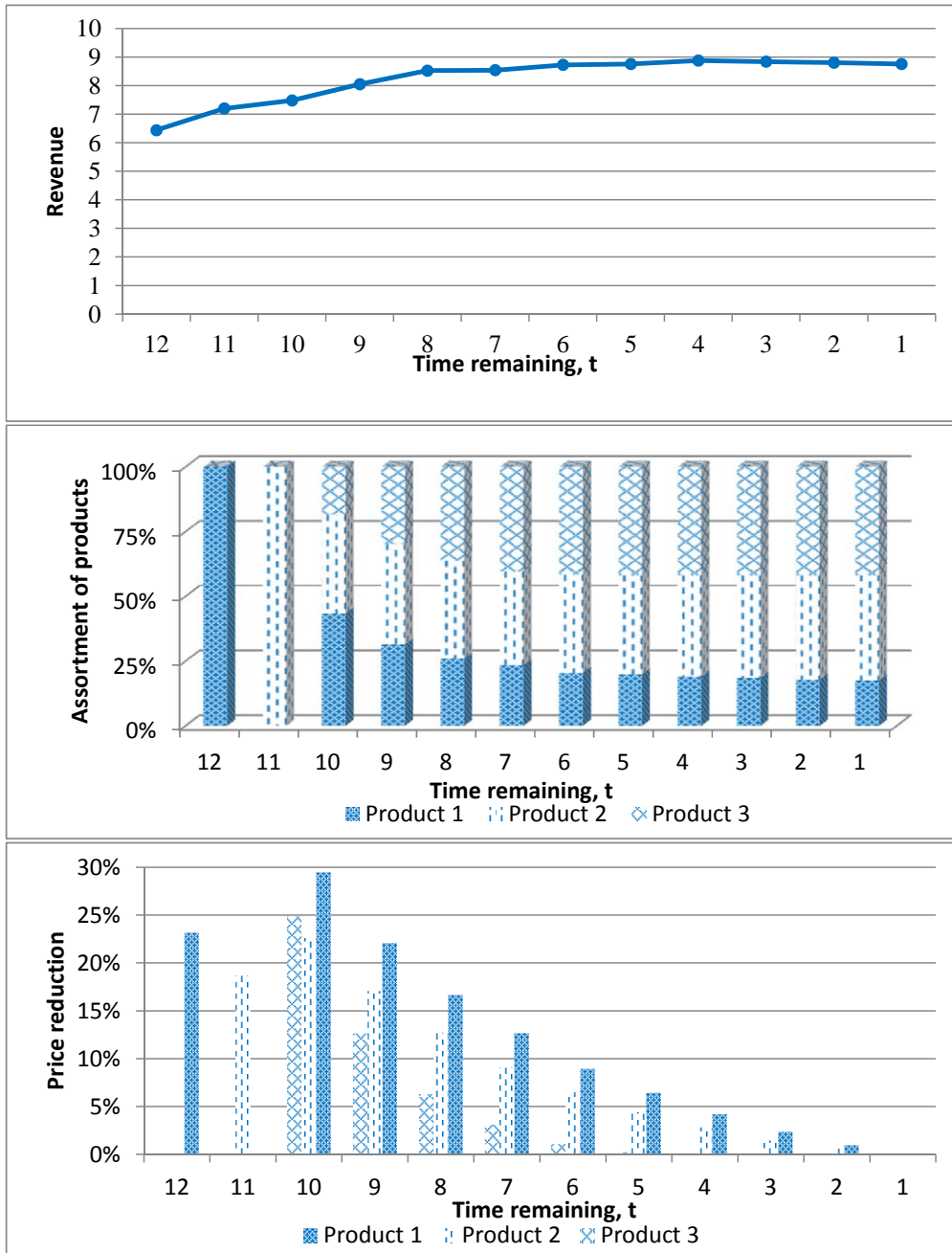
Figure 4: Revenues and assortment and price decisions

5 Conclusions and Further Research

Our research has focused on incorporating active demand learning into dynamic joint assortment and price optimization. It differs from the stream of research in which the demand parameters are estimated upfront, based on historic information, before the optimization takes place, but the optimization itself does not explicitly incorporate the possibility of learning the demand. In active learning, exploiting current knowledge to maximize profit is combined with investing in exploring to increase knowledge so as to increase future profits. Our study may initiate change in the way retail and department stores are managed and op-
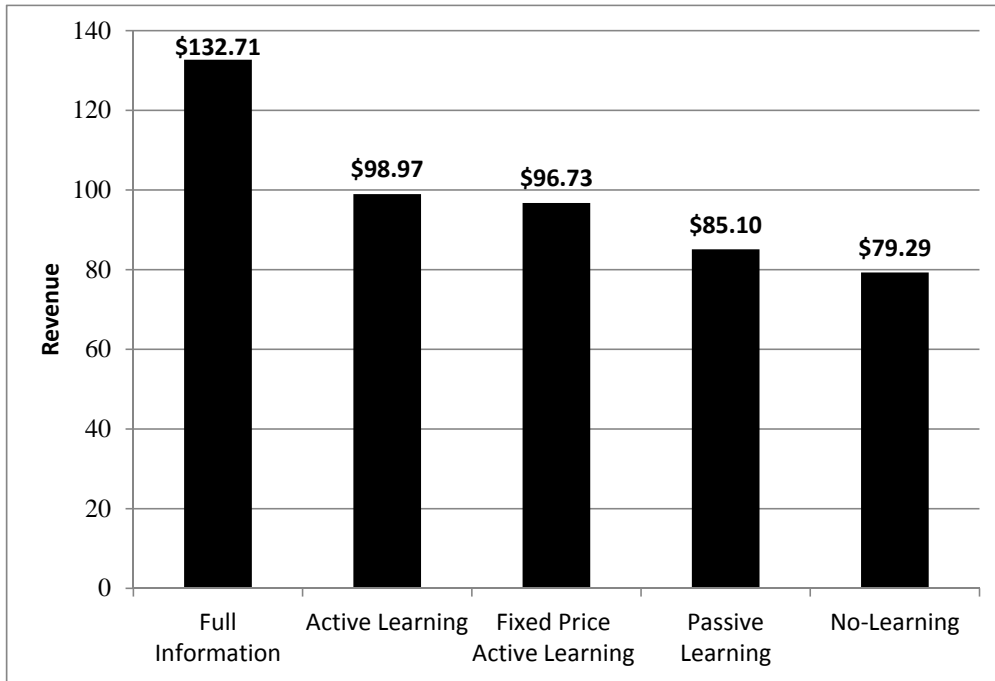
Figure 5: Revenues of four different policies and hypothetic full information case ($T = 12$)

erated, and may increase their profits by better assortment and pricing decisions. Given the importance of the retail sector, which corresponds to a large portion of developed countries' economies[1], the impact can be significant.

While the value of full information depends highly on the setting and the bias of the retailer's initial belief, our numerical study shows that both passive and active learning policies are effective in increasing revenue and recapturing loss of revenue due to the lack of full information. The value of learning increases with a longer sales horizon, less shelf space, and higher belief uncertainty.

Our investigation shows that embedding pricing in an active learning strategy results in price discounts in the early sales periods to intensify learning (lower prices increase the number of sales observations). The magnitude of the price discount increases as there is more time left until the end of the sales horizon. This confirms the intuition that product perishability should affect a retailers investment in exploration: as a product's sales horizon gets shorter, the benefits of learning reduce and the emphasis should be on exploitation. Less intuitive is the fact that as a retailer's uncertainty or as the size of the market increases, the price discounts initially increase, but then start to decrease. (If the retailer is extremely uncertain or if the market is extremely large, there is no need for any price discount.)

There are a number of extensions to the model that are beyond the scope of this paper and we plan to explore in the future. The current model assumes that there is a single unknown parameter of the demand function, i.e., the customer arrival rate. In more complex settings, it may be valuable to learn, in addition to the arrival rate, other parameters of the demand function, such as the customers' reservation price and price sensitivity. Another extension is to consider heterogenous product sizes. The current model assumes that products have the same size. If products have different sizes, then the linking constraint, i.e., the shelf

---

[1]The Australian Bureau of Statistics reports retail spending at about 20 billion dollar per month.
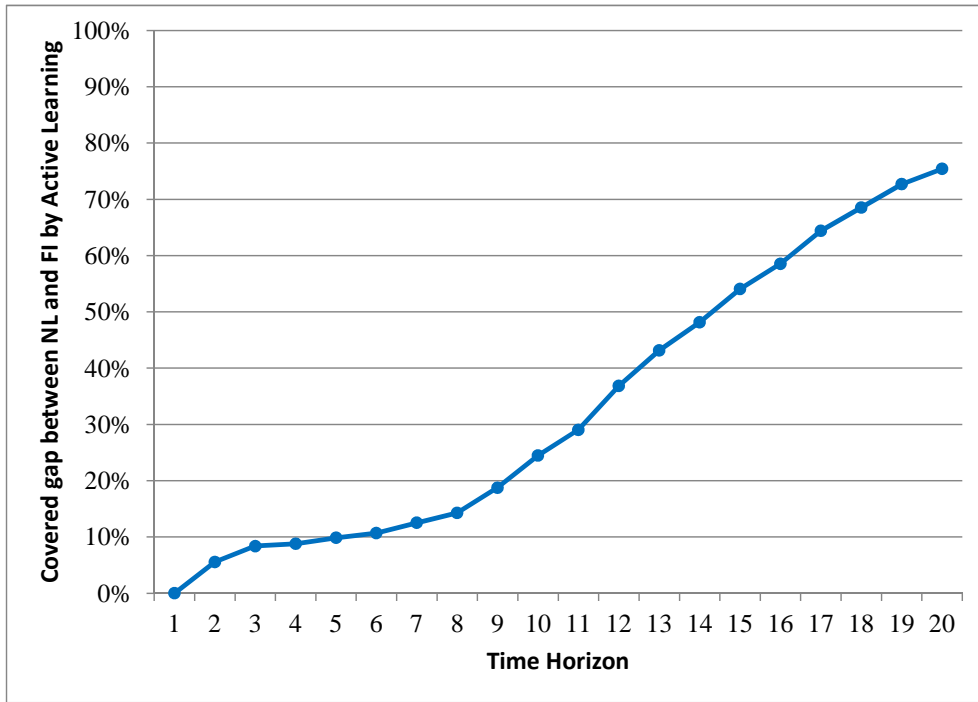
Figure 6: Revenue recovered by the active learning policy for different sales horizons
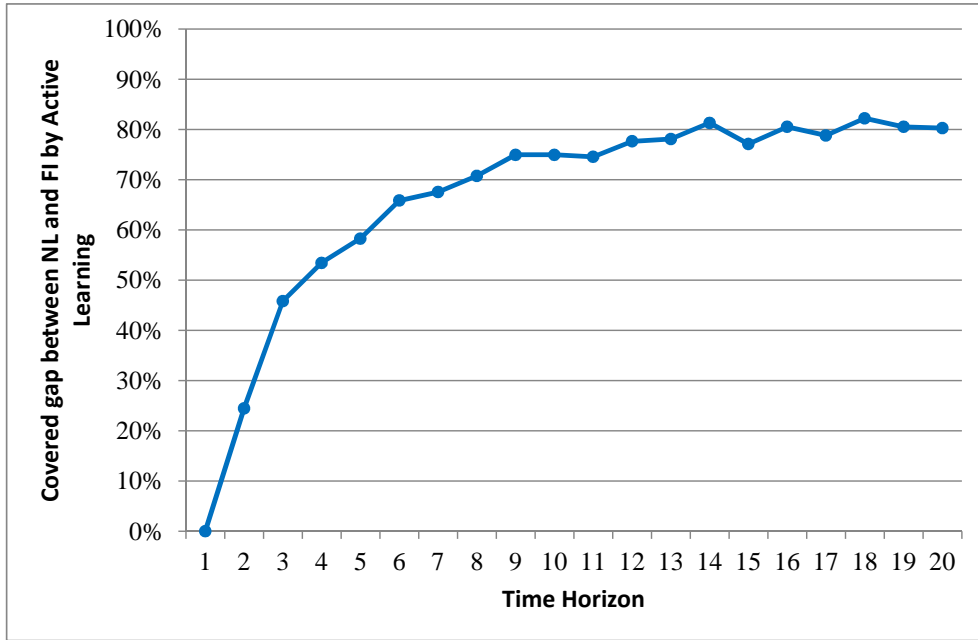

Figure 7: Revenue recovered by the active learning policy for different sales horizons

space constraint, represents a general knapsack problem rather than a cardinality knapsack problem, in which case the Lagrangian dual approximation may no longer be as strong and therefore the learning speed may decrease. Furthermore, we may also incorporate decisions on the amount of shelf space to assign to each chosen product, as studies have shown that with an increased amount of shelf space the demand for a product increases.
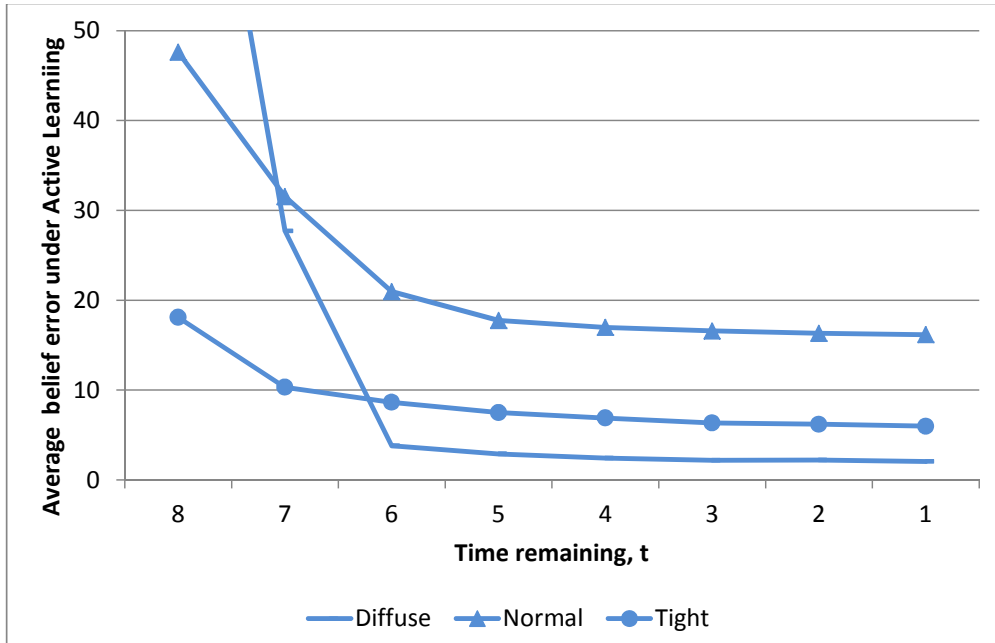
Figure 8: Belief error evolution for the active learning policy ($T = 8$)

It is also natural to investigate product substitution by modeling consumer choice. However, because of the price dynamics, substitution cannot be modeled by an exogenous matrix but needs to be price dependent, possibly a Multinomial Logit (MNL) model, which significantly increases the complexity. Note that a dynamic substitution model which incorporates prices does not allow us to treat products independently in the Lagrangian dual problem.

A substantially different model arises when we consider situations in which inventory is either scarce or needs to be managed, e.g., there is limited initial inventory, but inventory can be replenished at a cost. Recently, there has also been an increasing interest in studying *strategic customers*, which are customers that make their purchase decisions based on their expectations of the future price and availability of a product (see for example Levina et al. (2009)). Finally, a more methodological question to study is the quality of open loop policies. Our heuristic is based on constant Lagrangian multiplier and one-period look-ahead. To the best of our knowledge, there is limited knowledge about the actual quality of this widely used approximation.

References

M. Alan. Developments in implementable retailing research. *European Journal of Operational Research*, 68:1–8, 1993.

V. Araman and R. Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57, 2009.

G. Aydin and J. K. Ryan. Product line selection and pricing under the multinomial logit choice model. *Working paper*, Purdue University, 2000.
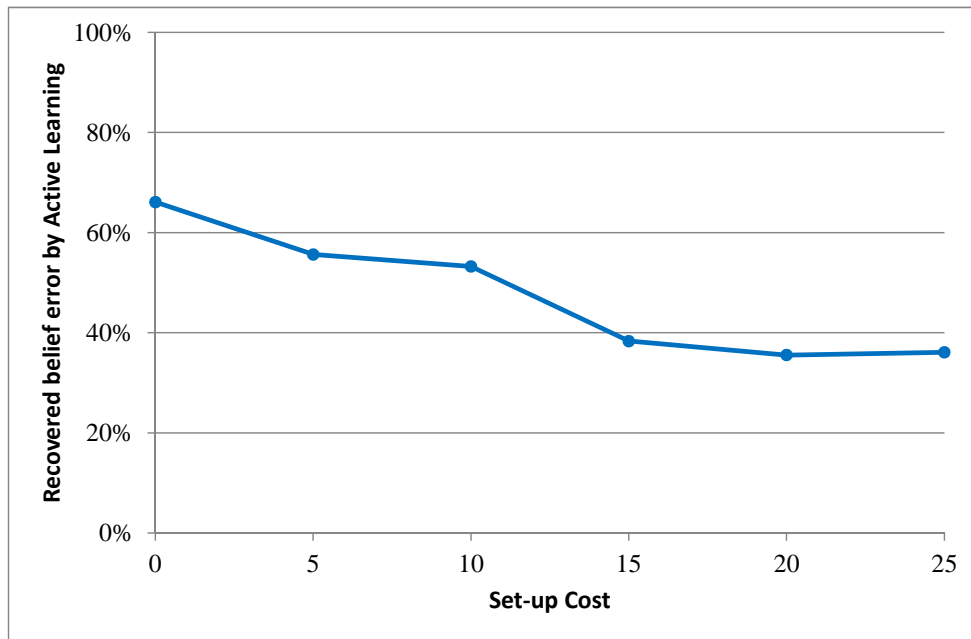
Figure 9: Belief error versus set-up cost for active learning policy in the presence of set-up costs ($T = 8$)

O. Besbes and A. Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, Advanced Publication, 2009.

G. Bitran and R. Caldentey. An overview of pricing models for revenue management. *Manufacturing Service Oper. Managemen*, 5:203–229, 2003.

D. J. Braden and S. S. Oren. Nonlinear pricing to produce information. *Marketing Science*, 13(3), 1994.

G. P. Cachon and A. G. Kök. Category management and coordination in retail assortment planning in the presence of basket shopping consumers. *Management Science*, 53(6), 2007.

R. Caldentey and F. Caro. Dynamic assortment planning. *Working Paper*, 2011.

F. Caro. *Dynamic Retail Assortment Models with Demand Learning for Seasonal Consumer Goods*. PhD thesis, Massachusetts Institute of Technology, 2005.

F. Caro and J. Gallien. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292, 2007.

K. D. Chen and W. H. Hausman. Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science*, 46, 2000.

G. Gallego and M. Talebian. Demand learning and dynamic pricing for multi-version products. *Journal of Revenue and Pricing Management*, 11, March 2012.

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

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

G. Gallego, R. Ratliff, and S. Shebalov. A general attraction model and an effcient formulation for the network revenue management problem. *Working Paper, Columbia University, NY, www.columbia.edu/ gmg2/GAMplusSBLPv2.pdf*, 2011.

V. Gaur and D. Honhon. Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52(10), 2006.

J. Gittins and D. Jones. A dynamic allocation index for the sequential design of experiments. In J. Gani, K. Sarkadi, and I. Vincze, editors, *Progress in statistics.* North-Holland Pub. Co., 1974.

S. Kesavan, V. Gaur, and A. Raman. Do inventory and gross margin data improve sales forecasts for u.s. public retailers? *Management Science*, 56(9):1519–1533, 2010.

A. G. Kök and M. L. Fisher. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55, 2007.

A. G. Kok and Y. Xu. Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science*, 57(9):1546–1563, 2011.

A. G. Kök, M. L. Fisher, and R. Vaidyanathan. Assortment planning: Review of literature and industry practice. In N. Agrawal and S.Smiths, editors, *Retail Supply Chain Management.* Springer, 2009.

T. Levina, Y. Levin, J. McGill, and M. Nediak. Dynamic pricing with online learning and strategic consumers: An application of the aggregating algorithm. *Operations Research*, 57(2), 2009.

Z. Li. A single-period assortment optimization model. *Production and Operations Management*, 16(13):369–380, 2007.

S. Mahajan and G. J. van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49, 2001.

H. E. Posen and D. A. Levinthal. Chasing a moving target: Exploitation and exploration in dynamic environments. *Management Science*, Articles in Advance, 2011.

W. Powel. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* Wiley Series in Probability and Statistics. Wiley, 2007.

K. Rajaram. Assortment planning in fashion retailing: methodology, application and analysis. *European Journal of Operational Research*, 129:186–208, 2001.

B. Rodriguez and G. Aydin. Assortment selection and pricing for configurable products under demand uncertainty. *European Journal of Operational Research*, 210:635–646, 2011.

P. Rusmevichientong, Z.-J. M. Shen, and D. B. Shmoys. Dynamic assortment optimization with a multinomila logit choice model and capacity constraint. *Operations Research*, 58, 2011.

R. Russell and T. Urban. The location and allocation of products and product families on retaile shelves. *Ann Oper Res*, 179, 2010.

C. Schön. On the optimal product line selection problem with price discrimination. *Management Science*, 56:896–902, 2010a.

C. Schön. On the product line selection problem under attraction choice models of consumer behavior. *European Journal of Operational Research*, 206:260–264, 2010b.

S. Smith and N. Agrawal. Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48:050–064, 2000.

C. S. Tang. Perspectives in supply chain risk management. *International Journal of Production Economics*, 103:451–488, 2006.

G. van Ryzin and S. Mahajan. On the relationship between inventory costs and variety benefits in retail assortment. *Management Science*, 45, 1999.

L. R. Weatherford and S. E. Bodily. A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking and pricing. *Operations Research*, 40(5):831–844, 1992.

R. Wright. Dynamic programming. *Lecture Notes*, 2011.

Appendix – Proofs

Lemma 1

The posterior probability density function (pdf) of market size θ conditioned on observing sales s at the end of the period is:

$$
\begin{aligned}
f(\theta|s) &= \frac{\exp(-\theta d(p))\frac{(\theta d(p))^s}{s!}\theta^{a-1}\exp(-b\theta)}{\int_0^\infty \exp(-x d(p))\frac{(x d(p))^s}{s!}x^{a-1}\exp(-bx)dx} \\
&= \frac{\exp\big(-\theta(b+d(p))\big)\theta^{s+a-1}}{\int_0^\infty \exp\big(-x(b+d(p))\big)x^{s+a-1}dx} \\
&= \frac{(b+d(p))^{a+s}}{\Gamma(a+s)}\theta^{a+s-1}\exp(-(b+d(p))\theta).
\end{aligned}
$$

■

Proposition 1

Consider (3) for a given product n and period t, with $a = a^n(t)$, $b = b^n(t)$ given. In what follows, we drop these subscripts. For the first part of the proof, for notational convenience we also drop the λ superscript, which we treat as constant for this part, . Notice that:

$$
v'(p) = \frac{\sqrt{a(b+1)}\hat{p}d(\hat{p})}{(b+d(p))^2}d'(p) \le 0.
$$

Applying chain rule to the expectation term in the definition of $g(p)$, we find

$$
\frac{d}{dp}E[(z-y(p))^+] = -\overline{F}(y(p))y'(p) = \overline{F}(y(p))\frac{1}{(v(p))^2}v'(p) \le 0,
$$

where $\overline{F}(y)$ represents the complementary cumulative distribution function (ccdf) of the standard normal random variable. Thus

$$
\frac{d}{dp}(v(p)E[(z-y(p))^+]) = v'(p)E[(z-y(p))^+] + v(p)\frac{d}{dp}E[(z-y(p))^+] \le 0,
$$

showing that the last term in $g(p)$ is non-increasing. Because $pd(p)$ is strictly concave, the maximum of $(p\frac{a}{b}d(p) - \lambda + (t-1)vE[(z-y)^+])$ is in the closed interval $[0, \hat{p}]$, (for more detail see the proof of Proposition 2), so R is well-defined.

We now consider how R changes with respect to λ. First we observe (treating p as constant) that

$$
\begin{aligned}
\frac{dg^\lambda}{d\lambda} &= \frac{d}{d\lambda}(p\frac{a}{b}d(p) - \lambda + (t-1)vE[(z-y^\lambda)^+]) \\
&= -1 + (t-1)v\overline{F}(z-y^\lambda)\frac{-1}{v} \\
&= -1 - (t-1)\overline{F}(z-y^\lambda) \le 0.
\end{aligned}
$$

It is now not hard to see that $R^\lambda = \max_p g^\lambda(p)$ is also a non-increasing function of λ. ■

Proposition 2

We take parameters n, t, λ, a and b to be fixed. Thus $g(p) = f(p) + h(p)$ where $f(p) = pd(p)$ is strictly concave with maximizer $\hat{p}$ and $h(p) = -\lambda + (t-1)v(p)E[(z - y(p))^+]$. From the proof of Proposition 1, we know that $h'(p) \leq 0$ for all p. Now since p^* maximizes g, it must be that $g'(p^*) = 0 = f'(p^*) + h'(p^*)$. Since $h'(p^*) \leq 0$, it must be that $f'(p^*) \geq 0$. Now since f is strictly concave and $f'(\hat{p}) = 0$, it must be that $p^* \leq \hat{p}$. ∎

Proposition 3

Let $\theta = \dfrac{a}{b}$. Then

$$
\lim_{t\to\infty} \frac{R}{t-1} = \frac{\max_p\{p\theta d(p) - \lambda + (t-1)\frac{\sqrt{a(\frac{a}{\theta}+1)}\hat{p}d(\hat{p})}{\frac{a}{\theta}}\frac{d(p)}{\frac{a}{\theta}+d(p)}E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\frac{\sqrt{a(\frac{a}{\theta}+1)}\hat{p}d(\hat{p})}{\frac{a}{\theta}}\frac{d(p)}{\frac{a}{\theta}+d(p)}})^+]\}}{t-1}
$$

$$
= \sqrt{\theta + \frac{\theta^2}{a}}\,\hat{p}d(\hat{p})\max_p\{\frac{d(p)}{\frac{a}{\theta}+d(p)}E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\sqrt{\theta+\frac{\theta^2}{a}}\frac{d(p)}{\frac{a}{\theta}+d(p)}})^+]\},
$$

where the second equality follows from the fact that the first two terms go to zero when t goes to infinity. Next, we observe that $\lim_{t\to\infty} \dfrac{R}{t-1}$ gets maximized where $\dfrac{d(p)}{\frac{a}{\theta}+d(p)}$ is maximized, since $yE[(z-\frac{c}{y})^+]$ gets maximized when y gets maximized, and that this happens at $p^* = 0$, in which case R goes to infinity. ∎

Propositions 4 and 5

$\dfrac{1}{\sqrt{a}} \to \infty$ or $\theta \to \infty$:

$$
\lim R = \max_p\{p\theta d(p) - \lambda + (t-1)\frac{\sqrt{a(\frac{a}{\theta}+1)}\hat{p}d(\hat{p})}{\frac{a}{\theta}}\frac{d(p)}{\frac{a}{\theta}+d(p)}E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\frac{\sqrt{a(\frac{a}{\theta}+1)}\hat{p}d(\hat{p})}{\frac{a}{\theta}}\frac{d(p)}{\frac{a}{\theta}+d(p)}})^+]\}
$$

$$
= \max_p\{p\theta d(p) - \lambda + (t-1)\sqrt{\theta + \frac{\theta^2}{a}}\,\hat{p}d(\hat{p})\frac{d(p)}{\frac{a}{\theta}+d(p)}E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\sqrt{\theta+\frac{\theta^2}{a}\frac{d(p)}{\frac{a}{\theta}+d(p)}}})^+]\}
$$

$$
= \max_p\{p\theta d(p) - \lambda + (t-1)\sqrt{\theta + \frac{\theta^2}{a}}\,\hat{p}d(\hat{p})E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\sqrt{\theta+\frac{\theta^2}{a}}})^+]\}
$$

$$
= \max_p\{pd(p)\}\theta - \lambda + (t-1)\sqrt{\theta + \frac{\theta^2}{a}}\,\hat{p}d(\hat{p})E[(z - \frac{(\lambda - \theta\hat{p}d(\hat{p}))}{\sqrt{\theta+\frac{\theta^2}{a}}})^+],
$$

where the third equality follows from the fact that $\dfrac{d(p)}{\frac{a}{\theta}+d(p)}$ goes to one when a goes to zero or θ goes to infinity. Next, we observe that $\lim R$ gets maximized for $p^* = \hat{p}$, in which case R goes to infinity. ∎