MODELS FOR RETAIL INVENTORY MANAGEMENT WITH DEMAND LEARNING

Zhe Wang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Kenan-Flagler Business School, Operations Area.

Chapel Hill
2015

Approved by:

Adam J. Mersereau (Chair)

Li Chen

Vinayak Deshpande

Nur Sunar

Serhan Ziya

ABSTRACT

ZHE WANG: Models for Retail Inventory Management with Demand Learning
(Under the direction of Dr. Adam J. Mersereau)

Matching supply with demand is key to success in the volatile and competitive retail business. To this end, retailers seek to improve their inventory decisions by learning demand from various sources. More interestingly, retailers' inventory decisions may in turn obscure the demand information they observe. This dissertation examines three problems in retail contexts that involve interactions between inventory management and demand learning. First, motivated by the unprecedented adverse impact of the 2008 financial crisis on retailers, we consider the inventory control problem of a firm experiencing potential demand shifts whose timings are known but whose impacts are not known. We establish structural results about the optimal policies, construct novel cost lower bounds based on particular information relaxations, and propose near-optimal heuristic policies derived from those bounds. We then consider the optimal allocation of a limited inventory for fashion retailers to conduct "merchandise tests" prior to the main selling season as a demand learning approach. We identity a key tradeoff between the quantity and quality of demand observations. We also find that the visibility into the timing of each sales transaction has a pivotal impact on the optimal allocation decisions and the value of merchandise tests. Finally, we consider a retailer selling an experiential product to consumers who learn product quality from reviews generated by previous buyers. The retailer maximizes profit by choosing whether to offer the product for sale to each arriving customer. We characterize the optimal product offering policies to be of threshold type. Interestingly, we find that it can be optimal for the firm to withhold inventory and not to offer the product even if an arriving customer is willing to buy for sure. We numerically demonstrate that personalized offering is most valuable when the price is high and customers are optimistic but uncertain about product quality.

Dedicated to my parents, my grandfather, and my girlfriend.

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been possible without the support, advice, and encouragement from a wide community of mentors, colleagues, friends, and family.

I would like to express my deepest gratitude and appreciation to my advisor, Adam J. Mersereau, for his unreserved patience, tolerance, guidance, and support in completing this research and my PhD program. He has devoted countless hours of his time and effort in helping me become not only a good researcher, but also a good person. He has served not only as a trusted mentor, but also a trusted friend. I am privileged and blessed to be able to have worked with him and to model my career and life after him for many years to come.

I would like to thank the other committee members, Li Chen, Vinayak Deshpande, Nur Sunar, and Serhan Ziya for their time and effort in supporting the completion of this research and for their helpful discussions and constructive feedback. A special thank goes to Li Chen for his invaluable support both inside and outside research.

I would like to thank the Kenan-Flagler Business School community as I have been blessed by the many people and resources available to me during my PhD Program. I would like to thank all of the faculty in the Operations area for their advice on research, career, and life. I would like to thank my fellow PhD Students, Hyun Seok Lee, Yen-Ting Lin, Vidya Mani, Karthik Natarajan, Adem Orsdemir, Aaron Ratcliffe, Gang Wang, Ying Zhang, whose friendship is among the greatest gifts I have received during my PhD life. I would also like to thank Holly Guthrie, Erin Leach, and Sharon Parks for their assistance throughout the program.

Finally, I give thanks to my family. My thanks go to my parents, Chengji Wang and Xiaorong Wang, for their unconditional love. I give thanks to my wonderful girlfriend, Shuxin Hong, for her support, tolerance, smile, and love that have carried me through the past 5 years. I celebrate this accomplishment with my beloved family members.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Retail, the business of selling goods and services to end consumers, is one of the most important industries in the U.S, and has a significant impact on the U.S. economy. A 2011 report from the National Retail Federation reveals that in 2009, the retail industry ranked third among all industries by directly adding $1.2 trillion to GDP, accounting for 8.5% of the U.S. GDP; it was also the largest private-sector employer in the country, directly providing 28.1 million full-time and part-time jobs, accounting for 24.1% of total national employment.

The retail business is challenging given its highly volatile and competitive nature. Success in retail requires successfully matching supply with demand. On the demand side, learning customer demand is crucial and has become increasingly challenging in a fluctuating economy with shortened product life cycles, prolonged lead times, and rapidly changing consumer behaviors. On the supply side, efficient and effective management of inventory, the single largest asset for most retailers, is at the heart of retail operations.

Demand learning and inventory management are intricately interrelated. On one hand, optimization of inventory decisions relies on information gathered from demand learning; on the other hand, the ultimate goal of demand learning is to minimize profit losses, including inventory costs, that are attributed to demand uncertainty. What further complicates the relationship between the two is the fact that retailers' inventory decisions may in turn obscure their demand observations, as retailers typically do not observe lost sales due to stockouts. This thesis focuses on the interactions between demand learning and inventory management in retail contexts (see Figure 1.1). In one direction, this dissertation examines how inventory decisions should respond when firms learn a dynamically changing demand

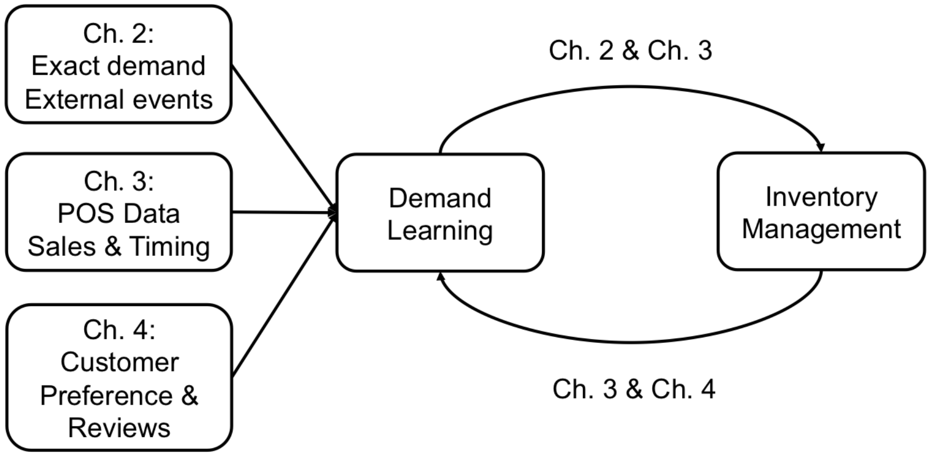


Figure 1.1: Organization of the dissertation.

(Chapter 2 and 4). In the other direction, we study the impact of inventory decisions on the effectiveness of demand learning (Chapter 3 and 4).

Demand learning also involves making the best use of available data. In response to rising trends towards business analytics and big data, this dissertation examines various sources of data for demand learning. In addition to historical demand, Chapter 2 incorporates information on past events as indicators for potential demand changes. Chapter 3 discusses the use of granular timing information of transactions on top of aggregate sales data. Chapter 4 involves learning demand from consumer-generated product reviews and from customers' personal characteristics and preference data.

We provide below an overview of Chapters 2 through 4 of this thesis. In Chapter 2, we study inventory management following a potential shift in the demand regime. The problem was motivated by the unprecedented adverse impact of the financial crisis started in 2008, which put retailers in uncharted territory in terms of revenue declines, credit availability, and demand forecasting. To analyze how retailers should manage inventory adaptively under such unpredictable circumstances, we consider a situation in which a firm is aware that the demand regime may (or may not) have changed due to some notable event and hopes to efficiently manage its inventory while also learning the actual demand trend. In the periods soon after such events, the manager can rely on historical demand to carefully estimate possibly obsolete demand parameters, discard the historical demand data and instead re-estimate demand parameters based on a limited history, or do something in between. The tradeoff inherent to this problem is between the precision brought by a long (but possibly out-of-date) history and the responsiveness that comes from relying on a recent (but limited) history.

We formulate the problem as a multi-period Bayesian inventory model with a mixture of two priors—one summarizes the inventory manager's historical demand information and the other reflects the manager's belief on the potential demand change. Although a theoretical characterization of some structural properties of the optimal inventory policy is possible, computing the exact policy remains challenging. As an alternative approach, we construct bounds on the optimal costs and develop associated heuristics. Constructing the lower bounds involves finding an "information relaxation" (Brown et al., 2010) that strikes a balance between the amount of information to relax and the computational complexity of the problem after the relaxation. In deriving a new "independentized" lower bound, we consider a novel auxiliary version of the problem that relaxes the natural dependence between demand signals and inventory trajectories that makes the inventory optimization difficult. The result is a tractable, meaningful bounding approach.

An extensive numerical study not only demonstrates the performance of the bounds and heuristics but also reveals the following key insights. Managers should remain wary of potential shifts in demand, as a demand forecast that fails to account for potential demand changes can be costly. When potential demand changes are moderate, a myopic policy may be sufficiently good, suggesting that managers may prioritize demand estimation over forward-looking inventory optimization in these cases. When extreme demand changes are possible, managers may need to use sophisticated inventory policies that jointly consider demand estimation and inventory dynamics.

Chapter 3 focuses on the practice of "merchandise testing" in the retailing of fashion products. This chapter investigates the role of inventory allocation decisions in demand learning across multiple locations. "Merchandise testing," first documented and studied by Fisher and Rajaram (2000), is a strategy adopted by fashion retailers to reduce demand uncertainty caused by short product life cycles and/or long lead times. In a merchandise test, a chain retailer allocates inventory to selected stores in its network to gather test sales data prior to the primary selling season. The retailer then uses collected data to generate a more accurate demand forecast for the entire chain, thereby improving its ordering decisions for the main selling season.

At its core, a merchandise test involves simultaneous demand learning across multiple locations. We formulate the merchandise testing problem as a stylized, two-period, multi-store Bayesian inventory model. In the presence of a limited quantity of overall test inventory and demand censoring (i.e., the retailer does not observe lost sales due to stockouts), our analysis reveals a unique tradeoff between the quantity and the quality of demand observations collected during a test. Our results on this tradeoff contribute to the literature on Bayesian inventory control with demand censoring, which mostly considers single-location settings. This chapter also examines the implication of increased visibility into demand information, which is relevant to the increased attention being paid to analytics in the retail industry. In particular, we consider cases in which the retailer does, or does not, observe the timing of each sales transaction, following a recent work by Jain et al. (2015). This sales timing information is usually available from point-of-sale systems equipped by most modern retailers. The characterization of the optimal test inventory allocations involves combining classic operations theories with the statistical literature on comparisons of experiments (Blackwell 1953). We also develop two near-optimal heuristics for computing test inventory allocations under general demand processes. Among many other analytical and numerical results, the key findings of this chapter are:

- The allocation of test inventory can significantly impact the value of demand learning through a merchandise test;
- The retailer's visibility into demand information has a pivotal impact on test inventory allocation decisions: when sales timing information is observable, retailers' priority is to achieve as many sales as possible during the test; When sales timing information is unobservable, retailers should maintain a sufficiently high service level in each test store before seeking to increase the number of stores to test.

In Chapter 4, we consider the problem of personalized offering when consumers generate and learn product quality from public product reviews. The motivation comes from online retailers' increasing capability to collect consumer preference data, to customize product offerings, and to collect and monitor consumer-generated product reviews The fact that an online retailer cannot offer customers with hands-on product information before purchases amplifies the importance of customer reviews. We aim to answer the fundamental research question of how a retailer should dynamically offer its inventory for sale to individual customers whose product reviews may influence future demand. This chapter incorporates two central elements: the consumers' ability to collectively learn product quality through reviews generated by their peers, and the firm's ability to personalize product offering based on its knowledge about individual customers' preferences.

In particular, we consider a firm that sells an experiential product at an exogenous, constant price over a finite selling season. For each customer, the gross utility from consuming the product comprises two parts—an *ex ante* observable part that we refer to as customer preference and an *ex post* observable part that we refer to as product quality. The quality of the product is known to the firm but is unknown and learned by customers. The customer base is heterogeneous and customers' preferences for the product follow a random distribution. We assume the firm may be able to identify the preference of an arriving customer (by analyzing the customer's past purchasing and online behaviors) and choose whether to offer the product to that particular customer without incurring additional costs. Once offered, the customer purchases a unit if her *ex ante* expected net utility is positive.

We model consumers' review generation and quality learning process by a stylized quasi-Bayesian social learning process. Consumers form a belief on the unknown quality of the product and update it as they observe reviews posted by previous buyers. Each arriving customer bases their purchasing decision on their *ex ante* expected net utility. Once they purchase, customers generate reviews based on their *ex post* net utilities, namely, utilities received after they have purchased and experienced the product. Customers are not fully rational and are subject to selection biases: they update their belief in a Bayesian fashion except that they ignore the potential selection biases and treat reviews *as if* they are randomly sampled from the entire population, instead of those who purchase.

We formulate the firm's personalized offering problem as a finite-horizon dynamic program. We show that the optimal product offering policy is a threshold-type policy—the firm should only offer the product for sale to customers with a higher-than-threshold preference. We demonstrate that it can be optimal for the firm *not* to offer the product to an arriving customer with a low preference in order to avoid a bad review that will negatively impact future sales, even when it is certain that the customer will buy the product if offered. While our base model assumes no capacity or inventory constraints, we extend our analysis to the setting in which the firm has a limited inventory upfront.

We study in a numerical analysis the impact of price and consumers' mean belief and uncertainty about product quality on the firm's optimal product offering decisions and on the potential value of personalized offerings. We find that compared with a benchmark policy that offers the product to every arriving customer who is willing to purchase, personalized offering may significantly improve profit, especially in settings in which the product price is high and customers are moderately optimistic but uncertain about product quality.

CHAPTER 2

Bayesian Inventory Management with Potential Change-Points in Demand

2.1 Introduction

In most real-world inventory control problems, demand changes over time and the true underlying demand distribution is never fully known to the inventory manager. The manager makes dual use of historical demand data to populate the current demand distribution and also to detect fundamental changes in the demand-generating process.

We provide two data examples in Figure 2.1 to illustrate the complexity of the manager's task. Figure 2.1(a) shows seasonally adjusted monthly sales by motor vehicle dealers in the United States before and after September 2001. Imagine the situation faced by an automobile dealer in the autumn of 2001. While a reasonable dealer would expect the September 2001 attacks to impact consumer demand for automobiles, the direction and magnitude of the impact would have been difficult to predict from data available at the time. In October 2001 sales spiked substantially, but was this just a temporary surge or an indicator of a new regime in automobile sales? Was pre-October historical data still useful for understanding demand in October and beyond? History shows that demand eventually fell back close to its pre-September levels, but this might have been unclear at the time.

Figure 2.1(b) shows seasonally adjusted monthly sales for U.S. women's clothing stores in 2008 and 2009. Uncertainty in the financial markets reached a crescendo in September 2008 with the backruptcy of investment bank Lehman Brothers. Even if a women's clothing retailer at the time anticipated a negative impact on garment sales, the magnitude and persistence of the impact would have been harder to anticipate. It turns out that adjusted women's clothing demand bottomed out in December 2008 and stayed close to its December 2008 levels for over a year afterwards. In hindsight, we see that the Lehman Brothers bankruptcy marked a distinct change in women's clothing demand that rendered the previous demand history unsuitable for understanding new demand levels.

These two examples illustrate what we believe is a common challenge faced by retail and other managers, namely how to respond to external events that have the potential to change the demand environment. While September 2001 and the Lehman Brothers bankruptcy are well-known events that impacted many firms across many industries, demand-changing events can also be local. For example, the start of a new marketing campaign, the entrance of a new competitor, the release of a new product version, and the opening of a nearby attraction can all potentially usher in new demand regimes for a firm. The introduction of a new product could also be interpreted as a potential demand-changing event when historical demand or sales data from a similar product are available for generating a reference forecast. All of these events have in common that their timing is known but their impact is not. In the periods soon after such events, the manager can rely on historical demand to carefully estimate possibly obsolete demand parameters, discard the historical demand data and instead re-estimate demand parameters based on a limited history, or do something in between. The tradeoff inherent to this problem is between the precision brought by a long (but possibly out-of-date) history and the responsiveness that comes from relying on a recent (but limited) history.

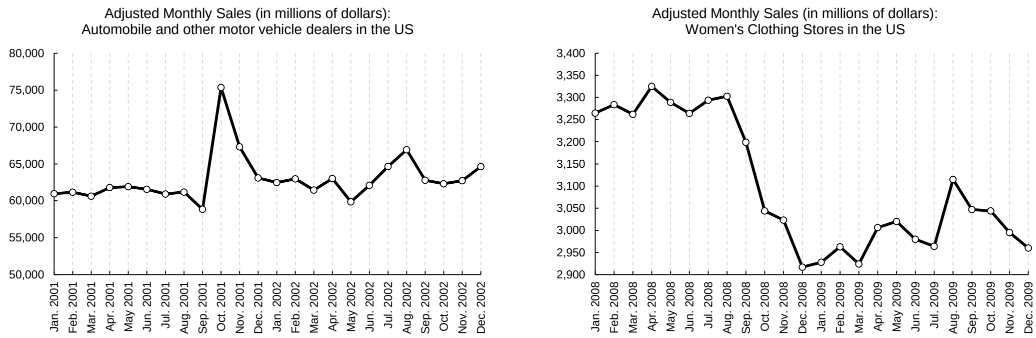
We refer to such events as *potential change-points* in demand, and we present and analyze an inventory control model that explicitly allows for potential change-points. We focus on the case in which there is a single potential change-point in the recent past, which is relevant to the examples of Figure 2.1 and to other examples in which change-points occur relatively infrequently. We seek to understand the structure and behavior of the optimal policy, and we look for computationally tractable bounds and heuristics.

We model the evolution of the manager's belief on the demand process using a Bayesian framework, extending the model pioneered by Scarf (1959) to allow for an unknown demand parameter to be distributed according to a mixture of a "historical" prior distribution and a "change" prior distribution. We leverage the structure of our demand model to characterize the effects of observed demand and the manager's belief on the optimal (state-dependent) base-stock levels.

The optimal policy remains challenging to compute. Scarf (1959) and Azoury (1985) show that the stationary Bayesian inventory problem can be solved efficiently using a dimensionality reduction approach for particular assumptions on the prior and demand distributions, but these assumptions do not hold when the unknown parameter is described by a mixture of distributions. We pursue heuristic policies coupled with cost lower bounds specific to our setting. Our most sophisticated bounding approach is novel in its formulation of an "independentized" problem that relaxes the dependence between physical demand and demand signals. A particular information relaxation of the demand signal information yields efficient subproblems that are solutions to stochastic multiperiod inventory problems with known demand distributions. In contrast to the dimensionality reduction approach of Scarf and Azoury, this approach can be applied for a broad set of belief and demand distributions. An extensive numerical analysis reveals that this bound and a look-ahead policy derived from it achieve small gaps. The numerical study also reveals that a myopic policy that accounts for potential change-points (but that ignores future inventory dynamics) works well except in extreme instances.

We also consider the sensitivity of our inventory policies to misspecification of the parameters of the manager's Bayesian prior. Taking a maximin profit perspective, we show that a conservative manager worried about profit downside will follow a policy that assumes the smallest prior (in a sense we will make precise) among a set of candidates.

The remainder of this chapter is organized as follows. We review related literature in §2.2. In §2.3, we formulate our Bayesian demand model and associated inventory control problem, and we present structural properties of the optimal inventory policy. In §2.4, we develop lower bounds for the optimal expected cost, and we introduce heuristic policies derived from these lower bounds. We numerically study these bounds and policies and measure their performance in §2.5. In §2.6, we discuss the estimation of model parameters and sensitivity to parameter misspecification. We conclude in §2.7.



(a) (b)

Figure 2.1: Examples of potential change-points in demand include (a) the terrorism events of September 2001 and (b) the Lehman Brothers bankruptcy in September 2008. (Source: U.S. Census Bureau)

2.2 Literature Review

This chapter relates to the inventory control literature dealing with nonstationary and/or partially observed demand processes. For situations in which the demand is nonstationary but the demand distributions are known, Karlin (1960) analyzes a dynamic inventory system in which demands are stochastic and may vary from period to period and proves the optimality of state-dependent base-stock policies. Song and Zipkin (1993, 1996) propose a continuous-time Markov-modulated Poisson demand framework to model inventory management problems in fluctuating demand environments. They assume that the demand distribution changes regime according to a known Markov chain and that the demand distribution in each regime is also fully known. Under these assumptions, they establish the optimality of state-dependent (s, S) policies. Sethi and Cheng (1997) show similar results in a generalized discrete-time inventory model with Markov-modulated demands. Graves (1999) characterizes the behavior of an adaptive base-stock policy under an ARIMA demand process. Iida and Zipkin (2006) and Lu et al. (2006) study approximate solutions for inventory planning problems with demand forecasting based on the martingale model of forecast evolution (MMFE).

Using a Bayesian framework, Scarf (1959) pioneers the study of optimal inventory policies under a stationary demand process with an unknown demand distribution parameter. Our work extends this framework to general demand distributions with a more flexible belief structure. Scarf (1960) and Azoury (1985) provide conditions under which the dimensionality of the problem can be reduced and the optimal base-stock levels can be obtained by solving a one-dimensional dynamic program. Our heuristics make possible the computation of approximate solutions to problems with more general prior and demand distributions. Azoury and Miyaoka (2009) study a Bayesian inventory problem where demand in each period depends on side information through a linear regression model. All of these works assume, as we do, that demand is fully observable and backlogged. There is another stream of research on inventory management problems when lost sales are unobserved and demand is therefore censored, assuming stationary demand. See, for example, Lariviere and Porteus (1999), Ding et al. (2002), Chen and Plambeck (2008), Bensoussan et al. (2007, 2008), Chen (2010), and Huh et al. (2011). Chen and Mersereau (2015) include a survey of this literature.

The demand process we consider is also related to that of Treharne and Sox (2002), who assume a Markov-modulated demand process in which state transitions are unobserved but the manager knows the transition probability matrix and maintains a belief of the underlying Markov state. They evaluate several heuristics, including limited lookahead policies, numerically. Brown et al. (2010) apply information relaxation bounds to an extended version of Treharne and Sox (2002)'s model with non-stationary cost parameters. Our model differs from these in two important respects. First, we assume a single potential shift in the past. This simplification yields structure that we exploit in deriving new results and bounds. Second, we model component demand distributions that are learned over time, whereas Treharne and Sox (2002) assume the demand distribution within each Markov state is known and fixed. We believe that our model brings distinct advantages in flexibility and parsimony. For further discussion, see §2.3.2. While the bounds we develop in §2.4 make use of results in Brown et al. (2010), we believe our "independentized" bound to be new.

Inasmuch as our work considers a change in demand regime, it also relates to Besbes and Zeevi (2011), in which a decision-maker seeks to detect and exploit a potential change in customers' willingness-to-pay distribution through dynamic pricing.

Our work is also related to a large stream of the statistics literature on change-point detection — detecting departures of a stochastic process from a known model by monitoring observations drawn from the process over time. We refer readers to Basseville and Nikiforov

(1993), Lai (1995), and the recent text Tartakovsky et al. (2015). This literature most commonly seeks to identify *when* a change occurs, focusing on the tradeoff between detection delay and the risk of false alarm. Our interest is not in declaring when a change-point occurs; rather, we formulate a dynamic optimization problem built on a stochastic model involving a potential change-point. Our model specializes typical sequential change-point formulations in that we assume that the timing of our potential change-point is known. In our model, a key unknown is whether or not the change actually occurs.

2.3 Model and Analysis

In this section we model an inventory management problem over a finite horizon following a potential change in the demand process. We present several structural properties, including certain structure inherited from well-studied inventory problems, which we use in our algorithm development in §2.4.

2.3.1 Inventory Management Following a Single Potential Change-point

Consider a single-item, T-period inventory system. At the beginning of period t, the decision maker (DM) observes the inventory position, x_t, and can place an order to bring the inventory position up to $y_t \geq x_t$ at a linear purchasing cost $c \geq 0$. We assume zero lead time such that the order is instantaneously delivered. Demand, denoted by a random variable D_t with realized value d_t, is then realized and satisfied by the inventory on hand. If at the end of the period the DM still has leftover inventory, i.e., $y_t - d_t > 0$, a linear holding cost h is charged; otherwise (i.e., $y_t - d_t \leq 0$), the excess demand is fully backlogged and incurs a linear shortage cost p. The discount factor is $\alpha \in (0, 1]$ each period. We assume $p > c(1 - \alpha)$ to avoid trivial solutions. The salvage value for leftover inventory at the end of period T is assumed to be zero. We shall omit the subscript t whenever it is clear from the context.

We assume the DM fully observes past demands without censoring, as does Scarf (1959). This assumption is driven in part by analytical tractability (as is our assumption of inventory backlogging), but we believe it is reasonable in practice when changes in demand are likely to impact a whole department, firm, or industry at the same time. This is the case, for example, for the September 2001 and Lehman Brothers bankruptcy contexts described in §2.1. In such cases, the firm can use data across stock-keeping units to correct for demand censoring.

We extend the Bayesian framework of Scarf (1959). The distinctive feature of our model is how we model the demand process. We assume that the demands D_t are independently drawn according to a density function $f(\cdot|\theta)$, where $\theta \in \Theta$ is an unknown parameter. The DM has a *historical prior* π^h (h stands for "history") on θ which reflects his prior knowledge of the demand parameter based on historical information. A *potential* changepoint occurs in period 1; thereafter, the DM is uncertain about whether the historical prior π^h continues to apply or whether the demand process has changed. The DM has a second prior distribution π^c on θ conditional on a change occurring (the superscript c stands for "change"). The *change probability* γ represents the DM's initial belief that a change has indeed occurred in period 1.

In practice, it is reasonable for the DM to estimate the historical prior using historical demand. However, it may be less obvious how to estimate the change prior π^c and the probability γ. We provide a full discussion of this in §2.6, where we perform a sensitivity analysis and suggest robust choices for these parameters.

Let π_t denote the DM's prior belief on the unknown parameter θ at the beginning of period t, then $\pi_1(\theta) = (1-\gamma)\pi^h(\theta) + \gamma\pi^c(\theta)$ by definition, and π_{t+1} is the posterior distribution obtained by updating π_t based on d_t, the demand realization in period t, using Bayes rule. That is,

$$\pi_{t+1}(\theta|\pi_t, D_t = d_t) = \frac{f(d_t|\theta)\pi_t(\theta)}{\int_\Theta f(d_t|\omega)\pi_t(\omega)\mathrm{d}\omega}. \tag{2.1}$$

We will show in §2.3.4 that this update has a particular structure that enables our analysis. The predictive demand density in period t given belief π_t is defined by $\phi(\xi|\pi_t) = \int_\Theta f(\xi|\theta)\pi_t(\theta)\mathrm{d}\theta$. A natural generalization of our model allows for multiple change priors. Most of our results directly extend to this case. (The main exception is Proposition 2.4 in §2.3.4, which requires further clarification on how priors are ordered and how to handle multi-dimensional change probabilities.)

The DM's objective is to minimize the Bayesian expected discounted total cost over a finite horizon based on his prior belief on the demand process by choosing an order quantity in each period. We use $(a)^+$ to denote $\max\{a, 0\}$ for a real number a. Given inventory position y after ordering and a demand realization d, the holding and shortage cost incurred in a single period is

$$l(y, d) = h(y - d)^+ + p(d - y)^+,$$

and the expected cost in period t with initial inventory position x and belief π_t is given by

$$\mathsf{E}_{D_t|\pi_t}\left[c(y - x) + l(y, D_t)\right] = c(y - x) + L(y|\pi_t),$$

where $L(y|\pi_t) := \mathsf{E}_{D_t|\pi_t}\left[l(y, D_t)\right] = \int_0^\infty l(y, \xi)\phi(\xi|\pi_t)\mathrm{d}\xi$.

Let $C_t(x|\pi_t)$ be the optimal expected cost for periods $t, t + 1, \ldots, T$. We can formulate the problem as a Bayesian dynamic program with the following optimality equations for $t = 1, \ldots, T$:

$$C_t(x|\pi_t) = \min_{y \geq x}\{c(y - x) + L(y|\pi_t) + \alpha\mathsf{E}_{D_t|\pi_t}\left[C_{t+1}(y - D_t|\pi_t \circ D_t)\right]\}, \qquad (2.2)$$

where $\pi_t \circ D_t := \pi_{t+1}(\cdot|\pi_t, D_t)$ as defined by (2.1). The terminal cost is given by $C_{T+1}(\cdot|\cdot) = 0$.

2.3.2 Discussion of our Demand Model

Our choice of a mixture model as a prior distribution for the unknown demand parameter is driven by our interest, as discussed in §2.1, in situations in which the DM has reason to believe a change in demand regime *may* have just occurred but is uncertain about whether a fundamental change has really transpired and, if so, about its extent. Our mixture model explicitly models this uncertainty. Such problems are most relevant and interesting in the few periods just after the potential change, and our choice of a parametric Bayesian model permits meaningful demand learning even with a few observations.

We use our model to illustrate numerically in Figure 2.2 the core demand learning tradeoff we seek to capture. The left panel of Figure 2.2 corresponds with a single demand

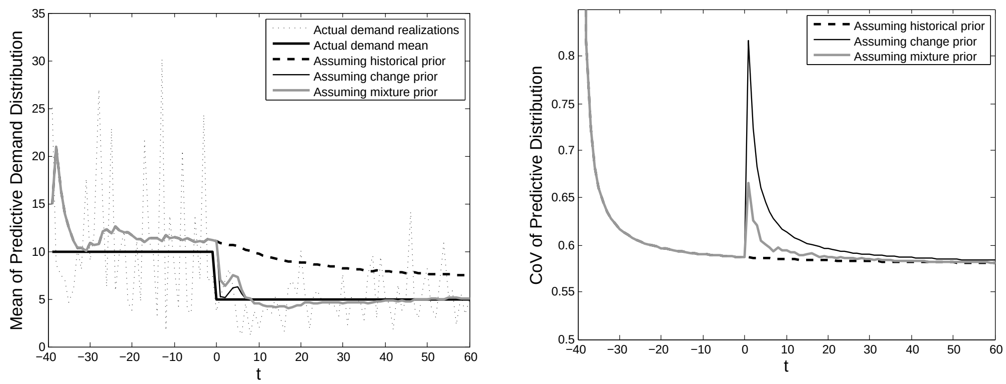


Figure 2.2: Illustration of behavior of various Bayesian demand learning models for two demand paths.

path involving a change in the demand mean from 10 to 5 occuring at time $t = 0$, while the right panel corresponds to a demand path drawn from a stationary demand process with demand mean 10 units.[1] The gray curves show statistics of the predictive demand distribution under our mixture model. For comparison, we also plot predictive demand statistics for models that use the historical prior alone and the change prior alone from time $t = 0$ on.

In the left-hand plot, we see that the mixture model more quickly learns the changed demand mean compared with the model using the historical prior alone, while (as expected) not quite as quickly as the model that assumes a change definitely occurred. In the right-hand plot, we see that the coefficient of variation (CoV) for our mixture model jumps considerably less and stabilizes more quickly than the model that relies on the change prior alone. We conclude that our "mixture" model of demand learning achieves a robust balance of responsiveness (in the event a change actually occurs) and stability (in the event no change occurs).

Further testing (omitted to conserve space) shows the necessity of allowing for the component distributions (in particular, the change distribution) to be learned from data rather than fixed *a priori* in situations where the DM has uncertainty around the post-

[1] The specific instance is similar to those in §2.5.3 (i.e., gamma demand distribution). Using notation to be introduced later, we assume an initial change prior of $\gamma_0 = 0.5$, we assume a known "shape" parameter $k = 3$ for gamma demand, and we assume a "change" gamma prior with $a^c = 3$, $S^c = 5$. The "historical" gamma prior is generated based on the observations from time $t = -40$ to $t = 0$ starting from $(a_{-40}, S_{-40}) = (3, 10)$.

change demand parameter. Fixing and mis-specifying θ may prevent the mixture model from converging to the true mean and variance of demand.

Finally, we have considered alternative modeling approaches for modeling change-points. Hypothesis testing-based approaches to change-point detection (e.g., Tartakovsky et al., 2015) do not naturally lead to forward-looking distributional forecasts that we require for multiperiod inventory control. Non-parametric methods (e.g., Huh et al., 2011) offer no concise state representation for use in a forward-looking dynamic optimization formulation.

2.3.3 Structure of the Optimal Policy

Although the demand process described in §2.3.1 is complicated by the potential change-points, it is still independent of the ordering decisions. Because of this, the cost functions are convex and a state-dependent base-stock policy is optimal. We state the following result for completeness, but we omit the proof because the result can be obtained by a straightforward modification of proofs in Scarf (1959) and Treharne and Sox (2002).

Proposition 2.1. *(a)* $C_t(x|\pi_t)$ *is a convex function of* x *for all* π_t*.*

(b) The optimal policy takes the form of a state-dependent base-stock policy. There exists a sequence of nonnegative functions $\{y_t^*(\pi_t)\}$ *such that it is optimal for the DM to order* $\min\{y_t^*(\pi_t) - x_t, 0\}$ *at the beginning of period* t *given inventory position* x_t *and belief* π_t*.*

We do not have closed-form expressions for the optimal policy, and given previous research it is unlikely that the optimal policy can be easily computed, much less simply expressed. We discuss the computability of the optimal policy in §2.4. However, as is often possible in finite horizon, non-stationary inventory problems (see Theorem 9.4.2 of Zipkin, 2000; also Karlin, 1960; Morton and Pentico, 1995), we are able to bound the optimal base-stock levels by easily computed myopic base-stock levels, which has the potential to reduce the search space for an optimal policy. The myopic policy is one in which the DM considers neither the evolution of future demand forecasts nor the carry-over of inventory across periods. The DM therefore treats each period as a single-period newsvendor problem. In our case, let $\Phi(\cdot|\pi_t)$ be the cumulative distribution function representing the DM's prediction of period t demand given belief π_t, i.e., $\Phi(d_t|\pi_t) = \int_0^{d_t} \phi(\xi|\pi_t)\mathrm{d}\xi$. Then, the base-stock level for period t under a myopic policy is given by $y_t^M(\pi_t)$ such that

$$\Phi(y_t^M(\pi_t)|\pi_t) = \begin{cases} \frac{p-c(1-\alpha)}{p+h} & ,t = 1, \ldots, T-1, \\ \frac{p-c}{p+h} & ,t = T. \end{cases} \tag{2.3}$$

The following proposition shows that this myopic policy upper-bounds the optimal policy. Proofs appear in an appendix unless otherwise indicated.

Proposition 2.2. *For all $t = 1, 2, \ldots, T$, $y_t^M(\pi_t) \geq y_t^*(\pi_t)$.*

We remark that both Proposition 2.1 and 2.2 extend to models with multiple potential change-points in both the past and future, as long as the timing of the potential change-points and their associated change priors and change probabilities are all known to the DM. Chen and Plambeck (2008) also show that a DM may want to stock less than the myopic inventory level when inventory is perishable, albeit in a different Bayesian inventory setting than ours (with stationary demand and censored observations).

2.3.4 Monotonicity Properties of Optimal Base-Stock Levels

We explore in this subsection some monotonicity properties of the optimal base-stock levels with respect to demand history, the historical and change priors, and the change probability. Some definitions are needed here before we proceed.

Likelihood Ratio Order. Let $f(\cdot)$ and $g(\cdot)$ be two probability density functions. f is larger than g *in the likelihood ratio order*, denoted by $f \geq_{lr} g$, if for all $d_1 > d_2$, $f(d_1)/g(d_1) \geq f(d_2)/g(d_2)$.

Monotone Likelihood Ratio Property (MLRP). A distribution family $f(\cdot|\theta)$ with a parameter $\theta \in \Theta$ is said to have the *Monotone Likelihood Ratio Property (MLRP)* if $f(\cdot|\theta_1) \geq_{lr} f(\cdot|\theta_2)$ for all $\theta_1 \geq \theta_2$. Many common distributions, such as normal with known variance, binomial, Poisson, gamma, and Weibull, have MLRP (see Karlin and Rubin, 1956).

Hereafter we assume that the demands are independent and from a distribution family $f(\cdot|\theta)$ with parameters $\theta \in \Theta$, and that $f(\cdot|\theta)$ has MLRP. The underlying implication of the MLRP assumption is that if a larger demand occurs, it becomes more likely that the underlying demand distribution $f(\cdot|\theta)$ has a higher θ parameter.

Scarf (1959) shows a monotonicity result in his setting with respect to the observed demand history. Specifically, the optimal base-stock level is increasing in the demand observation if the underlying demand process is stationary and the demand distribution $f(\cdot|\theta)$ is from the exponential family of the form $f(\xi|\theta) = \beta(\theta)e^{-\theta\xi}r(\xi)$ (with $r(\xi) = 0$ for $\xi < 0$). We can view our single change-point model as a variant of Scarf's model with MLRP demand and an initial prior being a mixture of distributions. The following proposition shows that we inherit Scarf's monotonicity result by generalizing his result to the case of MLRP demand.

Proposition 2.3. *Let $y_t^*(\pi_t)$ be the optimal base-stock level in period t ($t = 1, \ldots, T$) given belief π_t, where π_t ($t \geq 2$) is updated over π_{t-1} based on demand realization d_{t-1}. If the demand distribution family $f(\cdot|\theta)$ has MLRP, then the following hold:*

(a) $y_t^*(\pi_t) \leq y_t^*(\pi_t')$ for $\pi_t \leq_{lr} \pi_t'$;

(b) $y_t^*(\pi_t)$ is increasing in d_τ, for all $t \geq 2$, $\tau < t$.

Proposition 2.3(a) characterizes the behavior of the optimal base-stock level with respect to the DM's belief on the demand process. Intuitively, a larger (smaller) belief (in the likelihood ratio ordering) indicates a larger (smaller) demand parameter, which further implies a stochastically higher (lower) demand, which finally leads to a higher (lower) optimal base-stock level. Proposition 2.3(a) paves the way for establishing monotonicity properties of the optimal base-stock levels with respect to π^c, π^h, and γ in what follows. We use a closely related result when deriving the independentized lower bound in §2.4.1.2. Proposition 2.3(b) guarantees that it is always optimal to order more (less) in the next period if a higher (lower) demand is observed during the previous periods. We note that these results do not require specific assumptions on the initial belief π_1; it need not have a mixture form and can be any general distribution over the parameter space Θ. We present an example in Appendix A.2 showing the necessity of the MLRP assumption on $f(\cdot|\theta)$.

As mentioned previously, our model is distinguished by its particular prior structure. The prior is a mixture of two distinct distributions. The following lemma establishes that this structure survives the DM's belief updating procedure.

Lemma 2.1. *In the single potential change-point problem, let* $\mathbf{d}_t = (d_1, \ldots, d_t)$ *be any demand history up to period* t, $t = 1, \ldots, T$. $\pi_t(\cdot|\mathbf{d}_{t-1})$ *is then given by*

$$\pi_t(\theta|\mathbf{d}_{t-1}) = (1 - \gamma_t(\mathbf{d}_{t-1}))\pi_t^h(\theta|\mathbf{d}_{t-1}) + \gamma_t(\mathbf{d}_{t-1})\pi_t^c(\theta|\mathbf{d}_{t-1}),$$

where $\pi_t^h(\cdot|\mathbf{d}_{t-1})$ *is updated over* π^h *based on* $\mathbf{d}_{t-1}$, $\pi_t^c(\cdot|\mathbf{d}_{t-1})$ *is updated over* π^c *based on* $\mathbf{d}_{t-1}$, *and* $\gamma_t(\cdot)$ *is a function of* $\mathbf{d}_{t-1}$.

Lemma 2.1 shows that the belief updating procedure can be decomposed into two parts: one separately updates the beliefs conditioned on there being a change and on there being no change; the other updates the change probability. The belief is still in the form of a linear mixture distribution of those two updated beliefs, with the updated change probability as the weight. The following Proposition 2.4 uses this result to establish a relationship between the corresponding optimal base-stock levels. In §2.4 we will use the structure in Lemma 2.1 to derive an easily computed cost lower bound.

Proposition 2.4. *In the single potential change-point problem, let* $y_t^*(\pi_t)$ *be the optimal base-stock level in period* t $(t = 1, \ldots, T)$. *Let* $y_t^h(\pi_t^h)$ $(y_t^c(\pi_t^c))$ *be the corresponding optimal base-stock level when the change probability* $\gamma = 0$ *(respectively,* $\gamma = 1$*). The following hold:*

(a) If $\pi^h \leq_{lr} \pi^c$, $y_t^h(\pi_t^h) \leq y_t^*(\pi_t) \leq y_t^c(\pi_t^c)$*; otherwise if* $\pi^c \leq_{lr} \pi^h$, $y_t^c(\pi_t^c) \leq y_t^*(\pi_t) \leq y_t^h(\pi_t^h)$*;*

(b) If $\pi^h \leq_{lr} \pi^c$, $y_t^*(\pi_t)$ *is increasing in* γ*; otherwise if* $\pi^c \leq_{lr} \pi^h$, $y_t^*(\pi_t)$ *is decreasing in* γ.

Proof. We only show the proofs of the first parts of (a) and (b). The proofs of the second parts follow from a straightforward modification.

It is easy to verify that if $\pi^h \leq_{lr} \pi^c$ and $\pi_1(\theta) = (1-\gamma)\pi^h(\theta) + \gamma\pi^c(\theta)$ for some $\gamma \in [0, 1]$, then $\pi^h \leq_{lr} \pi_1 \leq_{lr} \pi^c$. Lemma 2(c) of Chen (2010) further guarantees that $\pi_t^h \leq_{lr} \pi_t \leq_{lr} \pi_t^c$ for all t. The first part of (a) follows directly from this result and from Proposition 2.3(a).

Now define γ' such that $\gamma < \gamma' \leq 1$, and let $\pi'_1(\theta) = (1-\gamma')\pi^h(\theta) + \gamma'\pi^c(\theta)$. Because we can write π'_1 as a convex combination of π^h and π^c, it follows that $\pi_1 \leq_{lr} \pi'_1$, and thus $\pi_t \leq_{lr} \pi'_t$ for all t. The desired result $y^*_t(\pi_t) \leq y^*_t(\pi'_t)$ then follows from Proposition 2.3(a). $\square$

Proposition 2.4 provides sufficient conditions for the optimal base-stock levels of the single potential change-point problem to be bounded by those of the two degenerate problems — one with $\gamma = 0$ and the other with $\gamma = 1$. The result is intuitive: if an increase in demand is possible, the DM should order more than if the demand remains stable, and less than if the demand is guaranteed to increase. Moreover, the DM should order more as the change probability increases.

Proposition 2.4 may reduce the search space for optimal policies. It also motivates simple and computable heuristic ordering policies. In particular, for certain choices of π^h and π^c, the optimal solutions to the two degenerate problems can easily be computed by applying the dimensionality reduction technique in Scarf (1960) and Azoury (1985). A base-stock level in the form of a convex combination of these two solutions is an appealing heuristic policy. We have found such a policy to perform reasonably well, though we do not pursue it in the following section because it is outperformed by a related policy, which is greedy with respect to a convex combination of cost-to-go functions for the two degenerate problems.

2.4 Bounds and Policies

The usual approach to evaluate the performance of an inventory policy is to compare its expected cost with that of the optimal policy. However, the complexity of the Bayesian inventory control problem with potential change-points makes it intractable to compute optimal solutions. The dimensionality reduction technique in Scarf (1959) and Azoury (1985) is in general not applicable for our model with potential change-points. The conditions for applying the technique are:

1. Suppose that S_t is a sufficient statistic for demand observations up to period t. There is a function $q_t(S_t)$ such that $\phi(\xi|S_t) = (1/q_t(S_t))\psi_t(\xi/q_t(S_t))$, where $\psi_t(\cdot)$ is a probability density function that depends only on t;

2. The function $q_t(S_t)$ satisfies $q_{t+1}(S_t \circ d) = q_t(S_t)U_{t+1}(d/q_t(S_t))$ for some continuous real valued function U_{t+1} such that $\int_0^\infty U_{t+1}(u)\psi_t(u)\mathrm{d}u < \infty$, where $S_t \circ d$ denotes an update of S_t based on demand observation d.

However, since the beliefs in our problem are linear mixtures of distributions, there do not exist q_t functions that can serve as such scale parameters for the predictive demand distributions. Therefore, it is computationally impractical to obtain the optimal policy or the optimal expected cost.

Treharne and Sox (2002) face a similar issue with an adaptive inventory control problem with similarities to our own. They point out the difficulty of computing an optimal policy even with an understanding of the policy structure, and they turn to heuristic policies. As an alternative approach, we develop lower bounds for the expected cost. Coupled with ordering heuristics derived from these bounds, we seek to bound the optimal cost as tightly as possible.

2.4.1 Bounds for Expected Cost

We develop two lower bounds in this subsection. The first makes use of the decomposition of Lemma 2.1, while the second makes use of a novel relaxation we call the "independentized" problem. Both make use of the "information relaxation" framework outlined in Brown et al. (2010).

2.4.1.1 The Mixture Lower Bound.

Lemma 2.1 implies that the DM's belief in a period can be decomposed as a convex combination of the beliefs implied by two "degenerate" information structures in which a change is known to have occurred or known not to have occurred. If the degenerate problems are easily solved (e.g., if the historical prior π^h and change prior π^c satisfy the conditions of Azoury, 1985), then the solutions can be easily employed to form an expected cost lower bound. Imagine an oracle who reveals to the DM whether or not a change has occurred. It is intuitive that the expected cost utilizing the oracle information would lower bound the true expected cost. (Given that the DM is seeking to minimize cost, the additional information revealed by the oracle can only help achieve lower cost.) This is the content of the following proposition.

Proposition 2.5. *Let $\mathbf{d}_{t-1}$, $\pi_t(\cdot|\mathbf{d}_{t-1})$, $\pi_t^h(\cdot|\mathbf{d}_{t-1})$, $\pi_t^c(\cdot|\mathbf{d}_{t-1})$ and $\gamma_t(\mathbf{d}_{t-1})$ be defined as in Lemma 2.1. For all $t = 1, \dots, T$, define the mixture lower bound $LB_t^M(x_t|\mathbf{d}_{t-1})$ by*

$$LB_t^M(x_t|\pi_t(\cdot|\mathbf{d}_{t-1})) = (1 - \gamma_t(\mathbf{d}_{t-1}))C_t(x|\pi_t^h(\cdot|\mathbf{d}_{t-1})) + \gamma_t(\mathbf{d}_{t-1})C_t(x|\pi_t^c(\cdot|\mathbf{d}_{t-1})),$$

then $LB_t^M(x_t|\pi_t(\cdot|\mathbf{d}_{t-1})) \leq C_t(x_t|\pi_t(\cdot|\mathbf{d}_{t-1}))$.

Proof. The intuition behind the result is given above. The oracle information can be viewed as an information relaxation. Therefore, the proposition follows from Lemma 2.1 in Brown et al. (2010). □

2.4.1.2 The Independentized Lower Bound.

When implementing an inventory policy with demand learning, the DM uses demand realizations in two ways: to calculate inventory positions and to update demand beliefs. We construct a lower bound using the notion of information relaxations (Brown et al., 2010) by relaxing only the information available to the DM for belief updating.

To motivate this, write as $\mathbf{D}_t = (\hat{D}_t, D_t)$ the DM's observation of demand in period t, where we artificially distinguish between the physical demand $\hat{D}_t$ that impacts inventory positions and the demand signal D_t that the DM uses to update his beliefs around θ. In the original problem, the physical demand and demand signal are one and the same and are therefore perfectly correlated. We write $\mathbf{D}_t^o$ for the original problem as $\mathbf{D}_t^o = (D_t, D_t)$. For the purpose of constructing a bound, we consider an "independentized" problem in which the physical demand and the demand signal are assumed to be independent of each other. We write $\mathbf{D}_t^\perp = (D_t^\perp, D_t)$ where both $D_t^\perp$ and D_t have a marginal density $\phi(\cdot|\pi_t)$, which is the predictive demand density implied by the belief π_t, but $D_t^\perp$ and D_t are independent, conditional on π_t.

Let $C_t(x_t|\pi_t)$ and $C_t^\perp(x_t|\pi_t)$ be the optimal expected costs of the original and the independentized problems, respectively, for periods $t, \dots, T$ given initial inventory position

x_t and belief π_t. Then we have

$$C_t(x_t|\pi_t) = \min_{y \ge x_t} \left\{ c(y - x_t) + L(y|\pi_t) + \alpha \mathsf{E}_{\mathbf{D}_t^o = (D_t, D_t)|\pi_t} \left[C_{t+1}(y - D_t|\pi_t \circ D_t) \right] \right\},$$
$$C_t^{\perp}(x_t|\pi_t) = \min_{y \ge x_t} \left\{ c(y - x_t) + L(y|\pi_t) + \alpha \mathsf{E}_{\mathbf{D}_t^{\perp} = (D_t^{\perp}, D_t)|\pi_t} \left[C_{t+1}^{\perp}(y - D_t^{\perp}|\pi_t \circ D_t) \right] \right\},$$

with terminal values $C_{T+1}(\cdot|\cdot) = C_{T+1}^{\perp}(\cdot|\cdot) = 0$.

With the notation above, we have the following proposition which shows that the optimal expected cost of the independentized problem serves as a lower bound for that of the original problem.

Proposition 2.6. *$C_t^{\perp}(x_t|\pi_t) \le C_t(x_t|\pi_t)$ for all x_t, π_t, and $t = 1, \dots, T$.*

The proof, in Appendix A.4, shows that the cost-to-go function, as a function of both the physical demand realization $d_t^{\perp}$ and demand signal realization d_t, is supermodular and that $\mathbf{D}_t^{\perp} = (D_t^{\perp}, D_t)$ is less than $\mathbf{D}_t^o = (D_t, D_t)$ in the supermodular ordering. High-level intuition is as follows. In the original problem, a small demand observation hurts the DM because it yields low revenues in the current period, but also because it implies a high end-of-period inventory position at the same time that demand forecasts are lowered. This combination of high inventory position and low demand forecast accentuates the possibility of inventory overage in the original problem. In the independentized problem, the correlation between high inventory positions and lowered demand forecasts is removed. In particular, high inventory positions and low demand forecasts are less likely to occur together.

Unfortunately, the independentized problem is not necessarily easier to solve than the original problem. To cope with this, we use the information relaxation approach proposed by Brown et al. (2010) to construct a lower bound for the expected cost of the independentized problem. The basic idea is the following. At each decision point t we assume that an oracle reveals the entire future path of demand signals $(d_t, \dots, d_T)$ to the DM. With this extra information and his current belief π_t, the DM is able to compute his future beliefs $\tilde{\pi}_{t+1}, \dots, \tilde{\pi}_T$ recursively through

$$\tilde{\pi}_t = \pi_t \qquad \text{and} \qquad \tilde{\pi}_{u+1} = \tilde{\pi}_u \circ d_u, \quad \forall u = t, \dots, T.$$

Let $\tilde{C}_t^{\perp}(x_t|\pi_t;(d_t,\ldots,d_T))$ be the optimal expected cost-to-go at period t given inventory position x_t, belief π_t and future demand signals $(d_t,\ldots,d_T)$. The independentized problem after relaxing future demand signals reduces to

$$\begin{aligned}\tilde{C}_t^{\perp}(x_t|\pi_t;(d_t,\ldots,d_T)) &= \tilde{C}_t^{\perp}(x_t|\tilde{\pi}_t,\ldots,\tilde{\pi}_T)\\ &= \min_{y\geq x_t}\left\{c(y-x_t)+L(y|\tilde{\pi}_t)+\alpha\mathsf{E}_{D_t^{\perp}|\tilde{\pi}_t}\left[\tilde{C}_{t+1}^{\perp}(y-D_t^{\perp}|\tilde{\pi}_{t+1},\ldots,\tilde{\pi}_T)\right]\right\},\end{aligned}$$

with $\tilde{C}_{T+1}^{\perp}(\cdot|\cdot)=0$. This is in fact a stochastic inventory problem with *nonstationary, known* demand distributions, the solution to which can easily be obtained as the solution to a (fully observed) MDP with a one-dimensional state space. Because the oracle information is impermissible in the independentized problem, the optimal expected cost of the reduced problem will be lower than that of the independentized one.

We formally state the independentized lower bound as follows.

Proposition 2.7. *Let $(D_t,\ldots,D_T)$ denote the random demand signals in the independentized problem for periods $t,\ldots,T$. For all $t=1,\ldots,T$, define the independentized lower bound $LB_t^I(x_t|\pi_t)$ by*

$$LB_t^I(x_t|\pi_t)=\mathbb{E}_{(D_t,\ldots,D_T)|\pi_t}\left[\tilde{C}_t^{\perp}(x_t|\pi_t;(D_t,\ldots,D_T))\right],$$

then $LB_t^I(x_t|\pi_t)\leq C_t^{\perp}(x_t|\pi_t)\leq C_t(x_t|\pi_t)$.

Proof. The first inequality is an application of Lemma 2.1 in Brown et al. (2010). The second inequality follows from Proposition 2.6. □

We estimate $LB_t^I(x_t|\pi_t)$ in the numerical results using the following procedure. In an outer simulation, we randomly generate full demand signal paths $(d_1,\ldots,d_T)$ and calculate predictive demand distributions, $(\phi_1,\ldots,\phi_T)$, based on the generated demand signal paths. We then solve for each sequence of predictive demand distributions an inner optimization problem which is an inventory control problem with nonstationary, known demand distributions. These inner dynamic programming problems can be solved with straightforward backwards induction. The average of the resulting expected costs estimates the independentized lower bound.

To our knowledge, the "independentized" approach to bounding inventory problems with demand learning has not previously been used in the literature. An advantage of the approach over the mixture lower-bounding approach of §2.4.1.1 is that it requires efficient solutions only for inventory subproblems with known demand distributions, not for subproblems involving demand learning as required in §2.4.1.1. This widens its applicability. A drawback of the approach is that it is estimated via simulation. Due to estimation error, this means that technically we do not have a provable bound if it is based on a finite number of sample paths. In our numerical results, we estimate the bound based on a large number (100,000) of sample paths.

The approach may be useful for inventory problems involving demand learning beyond the one considered in this chapter. It is clearly applicable for other generalizations of the Scarf (1959) model. Azoury (1985) shows that Scarf's model can be efficiently solved, but only for certain assumptions on the demand distribution. Without these assumptions, the optimal policy remains difficult to compute. In §2.5.2 we demonstrate that the independentized information relaxation is capable of meaningful bounds for the classic Scarf (1959) problem, for which we can generate the optimal costs for comparison.

2.4.1.3 Penalties.

The information relaxation approach of Brown et al. (2010) also allows for the assignment of a penalty on each sample path, which potentially tightens the bound by penalizing the use of "impermissible" information in solving the inner problems. The lower bound for the optimal expected cost of the original problem is obtained by either simulation or analytical expression of the minimum expected value of the cost of the relaxed problem plus the penalty.

Unfortunately, we do not find computationally viable penalties for the two relaxations we have proposed. For the mixture lower bound, any natural penalty destroys the decomposition exploited by the information relaxation, and the inner problem becomes as difficult to solve as the original problem. For the independentized lower bound, limited-lookahead methods for computing penalties (as considered in Brown et al. (2010)) prove too time consuming to compute for the continuous prior and demand distributions we consider. As a result, in general we impose a zero penalty on our inner problems for computing the lower bounds. We leave further investigation of penalties for future work. Even with zero penalties, we see meaningfully tight bounds in our numerical results.

2.4.2 Heuristic Policies

We develop three heuristic policies for the single potential change-point problem: a myopic policy, a look-ahead policy based on the mixture lower bound, and a look-ahead policy based on the independentized lower bound. In §2.5 we evaluate these heuristics using the lower bounds in §2.4.1.

Myopic Policy. Each period the DM updates his belief based on the observed demand and then uses the single-period newsvendor solution as the base-stock level. This policy therefore forecasts demand using the potential change-point model but is not forward looking in its inventory optimization.

Look-Ahead Policy Based on Mixture Lower Bound (LA-M). This policy takes advantage of the mixture lower bound (LB^M) we have developed in the previous subsection. For each period t, the DM uses LB^M_{t+1} as an approximation for the optimal cost-to-go function in period $t+1$, $C_{t+1}(\cdot|\cdot)$, and solves the following problem:

$$C^M_t(x_t|\pi_t) = \min_{y \geq x_t} \left\{ c(y - x_t) + L(y|\pi_t) + \alpha \mathsf{E}_{D_t|\pi_t} \left[LB^M_{t+1}(y - D_t|\pi_t \circ D_t) \right] \right\}.$$

Of course, the LA-M policy is only implementable if the LB^M_{t+1} lower bound is simple to compute. Therefore, this policy is only attractive for instances in which the degenerate "change" (i.e., $\gamma = 1$) and "no change" (i.e., $\gamma = 0$) problems are easy to solve; e.g., when they conform to the assumptions of Scarf (1960) or Azoury (1985).

Look-Ahead Policy Based on Independentized Lower Bound (LA-I). This policy is very similar to the LA-M policy except that it uses the independentized lower bound LB^I_{t+1} instead of LB^M_{t+1} to approximate the optimal cost-to-go function for the next period.

More specifically, in each period t the DM solves the following problem:

$$C_t^I(x_t|\pi_t) = \min_{y \geq x_t} \{c(y - x_t) + L(y|\pi_t) + \alpha \mathsf{E}_{D_t|\pi_t}\left[LB_{t+1}^I(y - D_t|\pi_t \circ D_t)\right]\},$$

where $LB_{t+1}^I(\cdot|\cdot)$ is estimated using Monte-Carlo simulation as described in §2.4.1.2. This LA-I policy can be applied to the single change-point problem with any belief and demand distribution; however, the computation effort grows as more sample paths are used to estimate the LB^I lower bound.

2.5 Numerical Analysis

In this section we conduct numerical analyses to demonstrate the performance of the lower bounds and heuristics proposed in §2.4. Without loss of generality we normalize the purchasing cost c to zero and the unit holding cost h to one. We also assume no discounting ($\alpha = 1$) throughout the section. We have also run our experiments with discount factor $\alpha = 0.8$ and found that the results do not change qualitatively.

We make use of the gamma-gamma conjugate pair as our model of demand in our numerical results. This demand structure is amenable to the dimensionality reduction technique of Scarf (1960) and Azoury (1985) for stationary versions of our problem. Given this demand structure we can therefore easily compute the degenerate problems required to evaluate the LA-M bound.

We will first review the gamma-gamma demand model and its relevant properties in §2.5.1. We will then test the independentized lower bound against Scarf (1960)'s Bayesian inventory problem with gamma-gamma demand in §2.5.2. Unlike the potential change-point problem, we are able to solve Scarf's problem optimally and compare our bound against the known optimal solution. Finally in §2.5.3, we will perform a comprehensive numerical study on bounds and heuristics for the potential change-point problem analyzed in §2.3 and 2.4.

2.5.1 The Gamma-Gamma Demand Model

The gamma-gamma demand model is a common one for the study of inventory management with demand learning (e.g., Azoury, 1985; Scarf, 1960; Chen, 2010) because of its versatility and ease of updating. Assume that demand follows a gamma density with known shape parameter k and unknown scale parameter θ:

$$f(\xi|\theta) = \frac{\theta^k \xi^{k-1} e^{-\theta\xi}}{\Gamma(k)}.$$

We assume an initial gamma prior with parameters (a, S) around the unknown scale parameter θ:

$$\pi_1(\theta) = \pi(\theta|a, S) = \frac{S^a \theta^{a-1} e^{-S\theta}}{\Gamma(a)}.$$

Given this information structure and demand observations $(d_1, \ldots, d_{t-1})$, it is well-known that sufficient statistics for Bayes updating are

$$a_t = a_{t-1} + k = a + k(t-1) \qquad \text{and} \qquad S_t = S_{t-1} + d_{t-1} = S + \sum_{i=1}^{t-1} d_i.$$

Furthermore, the updated distribution around θ at the beginning of period t is

$$\pi_t(\theta) = \pi(\theta|a_t, S_t) = \frac{S_t^{a_t} \theta^{a_t-1} e^{-S_t\theta}}{\Gamma(a_t)},$$

and the predictive demand density can be written as

$$\phi(d|\pi_t) = \phi(d|a_t, S_t) = \frac{1}{S_t}\phi_t\left(\frac{d}{S_t}\right),$$

where $\phi_t(u) = \frac{\Gamma(a_t+k)}{\Gamma(a_t)\Gamma(k)} u^{k-1}(1+u)^{-(a_t+k)}$. A result of Scarf (1960), extended in Azoury (1985), is that the optimization (2.2) can be written as a one-dimensional dynamic program:

$$v_t(x) = \min_{y \geq x} \left\{ c(y-x) + L_t(y) + \alpha \int_0^\infty (1+u) v_{t+1}\left(\frac{y-u}{1+u}\right) \phi_t(u) \mathrm{d}u \right\}, t = 1, \ldots, T,$$

with $v_{T+1}(\cdot) = 0$. Let y_t^* denote its optimal solution for period t. Then we have

(i) $C_t(x|S_t) = S_t v_t(x/S_t)$,

(ii) $y_t^*(S_t) = S_t y_t^*$.

Property (i) greatly simplifies calculation of our policies and bounds, in particular the mixture lower bound. Assuming that the change belief π_t^c for θ in period t is gamma with parameters (a_t^c, S_t^c) and that the no-change belief π_t^h is gamma with parameters (a_t^h, S_t^h), the mixture bound can be computed as

$$\begin{aligned} LB_t^M(x_t|\pi_t) = LB_t^M(x_t|\gamma_t, S_t^h, S_t^c) &= (1-\gamma_t)C_t(x|S_t^h) + \gamma_t C_t(x|S_t^c) \\ &= (1-\gamma_t)S_t^h v_t^h(x/S_t^h) + \gamma_t S_t^c v_t^c(x/S_t^c). \end{aligned}$$

2.5.2 Applying the Independentized Lower Bound to a Classical Problem

In this subsection we use the classic Bayesian inventory problem with gamma-gamma demand from Scarf (1960) to explore the behavior and quality of the independentized lower bound. This problem is a special case of our change-point problem with change probability equal to zero (or one), and it can be solved using a dimension reduction technique (as previously discussed). Therefore, it qualifies as a reasonable testbed for understanding the potential tightness of the independentized lower bound.

We make two observations about the results, which cover 36 instances in a full factorial design. First, we are able to estimate the lower bounds precisely, resulting in standard errors no more than 0.5% of the optimal cost for each of the instances. Second, the independentized bounding method produces meaningful lower bounds for most of the instances. We find the average gap over the 36 instances to be 0.73% (negative gaps are truncated to zero), and smaller than 2% for 33 out of the 36 instances. We observe that the gap is relatively larger for larger critical ratios (i.e., for large p) and for large spread in the prior (i.e., small a).

2.5.3 Bounds and Heuristics for the Change-Point Problem

In this subsection we numerically examine the performance of three heuristic policies – Myopic, LA-M and LA-I – for the single change-point problem introduced in §2.3 by comparing their expected costs with the lower bounds. In order to compute both the LB^M and LB^I lower bounds and their corresponding one-period look-ahead policies LA-M and LA-I, we assume a gamma-gamma conjugate demand structure. The demands are from a gamma distribution with parameters (k, θ). We only report the results for $k = 3$ here since we have observed results for $k = 1$ and 5 to be qualitatively similar. If the demand does not change at the beginning of the planning horizon, θ follows a gamma distribution with parameters (a^h, S^h); otherwise it follows a gamma distribution with parameters (a^c, S^c). We choose the shape parameters of the two prior distributions to be $a^h = 48$ and $a^c = 3$. We therefore have $a^h > a^c$, which implies that the DM is more uncertain about the demand distribution if the demand does change. This seems representative of practice, where the DM would have an accurate demand forecast based on an abundant demand history but would only have a coarse one following a potential demand shock. We fix $S^h = 160$ such that the no-change prior mean is $a^h/S^h = 48/160 = 0.3$. We vary S^c such that $S^c = 1, 5, 10, 15$ and 19, indicating extremely downward, downward, stationary, upward, and extremely upward potential changes in demand. We label the $S^c = 5, 10, 15$ cases as "moderate change" cases and the $S^c = 1$ and 19 cases as "extreme change" cases in which potential demand changes are quite large. We vary the initial change probability γ such that $\gamma = 0.2, 0.5$ and 0.8. The unit shortage cost p is set to be 4 and 9, indicating critical fractiles of 0.8 and 0.9, respectively. To examine the effect of the length of the planning horizon T, we let $T = 5$ and 10. Therefore, we have $5 \times 3 \times 2 \times 2 = 60$ instances in total in our full-factorial design.

For each instance, we compute the LB^M bound, estimate the LB^I using Monte-Carlo simulation with 100,000 demand signal paths, and estimate the expected costs of the Myopic, LA-M, and LA-I policies using simulations with an identical set of 10,000 demand paths. We also use the same set of demand paths to estimate the expected costs of two additional naïve policies – optimal policies as if $\gamma = 0$ (denoted by OPTNOCHG) and as if $\gamma = 1$ (denoted by OPTCHG) – as performance benchmarks. The OPTNOCHG policy would be adopted if the DM ignores the potential change-point and only uses the historical demand information for forecasting and inventory decisions. At the other extreme, the OPTCHG policy would be employed if the DM ditches all the historical demand information and starts fresh with a belief reflecting a change in demand.

Due to space limitations, we refer readers to the electronic companion for detailed tables of results. We find the estimated independentized lower bound LB^I to be tighter than the mixture lower bound LB^M across all 60 instances, and therefore we use our estimated LB^I to evaluate optimality gaps of the various policies. We calculate the deviation of each policy's estimated expected costs from the estimated independentized lower bound LB^I (computed by $(\text{Cost} - LB^I)/LB^I \times 100\%$) for each instance. The percentage gaps, averaged over parameter levels, are summarized in Table 2.1 (for "moderate change" cases) and Table 2.2 (for "extreme change" cases).

We make several observations about the results in Table 2.1 and Table 2.2. First, for both moderate and extreme scenarios, myopic, LA-M and LA-I policies nearly always perform significantly better than the OPTCHG and OPTNOCHG policies. Intuitively, as change probability γ increases, the performance of OPTCHG gets better while that of OPTNOCHG gets worse. But even in their best instances (i.e., $\gamma = 0.2$ for OPTNOCHG and $\gamma = 0.8$ for OPTNOCHG), they yield larger gaps than the three heuristics. This highlights the danger of ignoring uncertainty around whether a demand change may or may not have happened.

Second, myopic, LA-M and LA-I policies have nearly the same performance under moderate scenarios, achieving average gaps of 1.16%, 1.15% and 1.16%, respectively. This suggests that the myopic policy may be an appealing choice except when extreme demand changes are possible, especially given its simplicity for implementation in practice. Other authors have found myopic policies to perform well in inventory contexts with demand learning (e.g., Lovejoy, 1990, 1992). A managerial insight is that intelligent demand estimation may merit more attention than forward-looking optimization when a (moderate) demand shift may have recently occurred.

The myopic policy still performs reasonably well along with the two look-ahead heuristics when there has been a large potential increase in demand ($S^c = 19$). However, when there has been a potential extreme downward change in demand ($S^c = 1$), all the three heuristics exhibit larger gaps relative to the lower bound. The myopic policy yields an average gap of 15.74%. The LA-M and LA-I policy have much smaller average gaps (5.72% and 5.74%, respectively) than the myopic policy. This observation suggests that more sophisticated policies bring significant benefits over the myopic policy when relatively extreme changes

Table 2.1: Mean percentage gaps for moderate change cases, averaged over parameter levels.

Parameter		Myopic	LA-M	LA-I	OPTCHG	OPTNOCHG
S^c	5	1.00%	0.99%	0.99%	5.79%	14.32%
	10	0.90%	0.97%	1.00%	3.41%	21.73%
	15	1.58%	1.50%	1.50%	7.67%	39.78%
p	4	0.74%	0.79%	0.81%	4.73%	19.83%
	9	1.58%	1.51%	1.51%	6.52%	30.72%
γ	0.2	0.83%	0.91%	0.90%	11.09%	11.89%
	0.5	1.53%	1.55%	1.58%	4.54%	26.49%
	0.8	1.12%	1.00%	1.00%	1.23%	37.45%
T	5	1.50%	1.50%	1.51%	6.95%	25.39%
	10	0.82%	0.81%	0.81%	4.29%	25.16%
Overall		1.16%	1.15%	1.16%	5.62%	25.27%

Note. Negative gaps are truncated to zero before averaging.

are possible. The gaps discussed here reflect the deviation of the policies' expected costs only from the cost lower bound rather than the optimal cost. Therefore, these gaps are conservative in that they overestimate the optimality gaps.

Finally, although we have observed that LB^I is tighter than LB^M for all instances, the LA-M and LA-I policies (which approximate cost-to-go functions by LB^M and LB^I, respectively) have nearly the same performance under all scenarios. Recall that the LA-M policy can only be efficiently computed for cases in which the "degenerate" problems referenced in §2.4.1.1 can be solved easily. The LA-M policy is recommended for such cases; for other cases, the LA-I policy is likely to be more efficient to compute.

2.6 Parameter Estimation and Sensitivity

The demand model of §2.3 requires the specification of three inputs: a "no-change" or historical prior π^h, a change prior π^c, and a change probability γ. The no-change prior π^h, the forecast of demand in the absence of a potential change-point, can be estimated using established techniques applied to historical demand, and we do not elaborate on it here. However, in many contexts it may be less obvious how to estimate the parameters π^c and γ.

Table 2.2: Mean percentage gaps for extreme change cases, averaged over parameter levels.

Parameter		Myopic	LA-M	LA-I	OPTCHG	OPTNOCHG
S^c	1	15.74%	5.72%	5.74%	35.06%	73.97%
	19	1.73%	1.56%	1.55%	12.09%	53.38%
p	4	6.79%	2.96%	2.97%	18.55%	61.74%
	9	10.68%	4.32%	4.32%	28.60%	65.60%
γ	0.2	5.16%	3.97%	3.96%	43.12%	21.92%
	0.5	11.74%	4.94%	4.96%	20.89%	53.49%
	0.8	9.30%	2.01%	2.02%	6.72%	115.60%
T	5	8.96%	4.32%	4.32%	29.43%	60.73%
	10	8.51%	2.97%	2.97%	17.72%	66.62%
Overall		8.73%	3.64%	3.64%	23.58%	63.67%

Note. Negative gaps are truncated to zero before averaging.

Selecting the change prior π^c entails predicting the direction and magnitude of a potential change. This represents a new demand regime for the firm by definition, but in many cases it may be a regime with past precedents. Imagine a retailer facing the entrance of a new competitor at one of its locations. It is likely to have faced similar entrances in the past at other locations. When potential change-points are driven by changes in the state of the economy (e.g., the example of women's clothing following the 2008 financial crisis, discussed in our introduction), financial markets (or forecasts thereof) may offer signals that can be used to inform demand forecasts (Osadchiy et al., 2013). If neither of these two approaches is applicable, a firm might generate π^c by inflating the variance of π^h and/or inflating or deflating its mean by percentages determined by expert opinions.

The change probability γ is particularly challenging to estimate because it is arguably most situation-specific and least amenable to estimation from historical data. Fortunately, we have found that the performance of our policies is relatively insensitive to mis-specification of γ. Figure 2.3 plots results from a numerical study similar to §2.5.3 except that we allow for misspecification of the change probability γ. The manager employs the LA-M heuristic, but computes forecasts and stocking decisions using a γ parameter that may differ from the parameter used to simulate the underlying demand process.

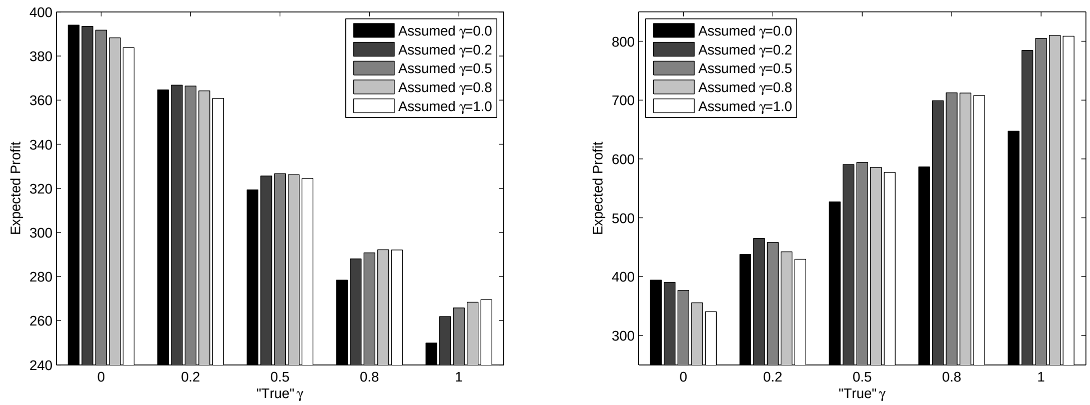


Figure 2.3: Sensitivity to misspecification of the change probability γ when the change prior represents a moderate decrease (left) and moderate increase (right) in demand. The bars in each chart represent expected profits when the manager assumes $\gamma = 0.0$ (white), 0.2, 0.5, 0.8, and 1.0 (black).

In contrast to our earlier development, we take a profit perspective here because changing the "true" value of γ changes expected demand, making a comparison between costs meaningless. Specifically, we translate expected costs into expected profit in the natural way, defining expected single-period profit as $\mathsf{E}\left[p\min\{y, D\} - c(y-x) - h(y-D)^+\right] = -\mathsf{E}\left[c(y-x) + h(y-D)^+ + p(D-y)^+\right] + p\mathsf{E}\left[D\right]$. The results in Figure 2.3 assume $c = 0$, $h = 1$, $p = 9$, $T = 5$, historical prior π^h given by a Gamma(48,160), and change prior π^c given by either Gamma(3,5) (i.e., "moderate decrease"), or Gamma(3,15) (i.e., "moderate increase"), where the parameters have the same interpretations as in §2.5.3. We have found consistent results across a broader set of instances.

We observe from Figure 2.3 that the expected profits naturally vary with the true underlying demand process, and that the profits are always highest for each instance when the assumed γ matches the "true" one used to generate the demand data. We also observe that the expected profit for each instance remains relatively flat as we move the assumed change probability γ from 0.2 to 0.5 to 0.8. In particular, the policy assuming $\gamma = 0.5$ exhibits robust performance across all of the instances we tried.

We also observe the least variation in profits across instances for policies that assume demand will be low. That is, when the change prior indicates a possible downward change, the "flattest" profits are obtained by the policy assuming γ equal to 1.0 or 0.8. When the change prior indicates a possible upward change in demand, the flattest profits are obtained by the policy assuming γ equal to 0 or 0.2. This suggests that a conservative decision-maker worried about downside risk may wish to choose π^c and γ by erring on the side of underestimating demand.

Proposition 2.8 below formalizes this finding for a Bayesian repeated newsvendor setting in which there is no inventory carryover across periods. Consider a T-period Bayesian newsvendor problem with unit selling price r, unit purchasing cost $c < r$, and inventory that perishes at the end of each period with zero salvage value. As before, demands are i.i.d. with density $f(\cdot|\theta)$. Let $G(y|\pi) = \mathsf{E}_{D|\pi}[r\min\{y, D\}] - cy$ be the single-period expected profit given order quantity y and prior π. We denote by $\mathbf{y} = (y_1, \ldots, y_T)$ a non-anticipative inventory policy. In general, y_t may be a function of all the information that the DM has up to period t. Let $\mathbf{D}_t = (D_1, \ldots, D_t)$ be demand until period t. For any initial prior π, let $V_T(\pi)$ denote the optimal expected profit for this problem, i.e.,

$$V_T(\pi) = \max_{\mathbf{y}} \sum_{t=1}^{T} E[G(y_t|\pi \circ \mathbf{D}_{t-1})|\pi], \tag{2.4}$$

where $\pi \circ \mathbf{D}_{t-1}$ is the posterior updated based on demand history. Suppose that a conservative DM has a bounded set $\mathcal{P}$ that contains all candidate priors on θ, and there exists a "smallest" prior $\underline{\pi} \in \mathcal{P}$ such that $\underline{\pi} \leq_{lr} \pi$ for all $\pi \in \mathcal{P}$. The objective is to maximize the worst-case expected profit, which translates into a max-min version of problem (2.4):

$$R_T(\mathcal{P}) = \max_{\mathbf{y}} \min_{\pi \in \mathcal{P}} \sum_{t=1}^{T} E[G(y_t|\pi \circ \mathbf{D}_{t-1})|\pi].$$

Proposition 2.8. *Suppose that $f(\cdot|\theta)$ has MLRP. Then $R_T(\mathcal{P}) = V_T(\underline{\pi})$.*

The proposition says that the DM can obtain the optimal policy for the max-min problem by simply solving (2.4) for $\pi = \underline{\pi}$. We note that Proposition 2.8 is a fairly general statement about the choice of prior beliefs, and the intuition can be applied to the selection of π^c as well as γ.

To summarize this section, we have suggested a few ways for a manager to think about choosing the parameters π^c and γ. We show evidence that the results of our heuristics are relatively insensitive to the specification of the change probability, particularly if a change prior is chosen away from the extremes 0 and 1. We also find both analytically and numerically that a max-min formulation is solved by assuming the smallest change prior structure among a set of candidates. Therefore, a manager concerned about downside profit risk may choose to "play it safe" by erring on the side of underestimating demand.

2.7 Conclusions

Our numerical study yields several insights on inventory management in uncertain demand environments. First, if a change in demand regime is suspected, managers can recover significant costs by accounting for this uncertainty. That is, a manager should remain wary of demand changepoints. Second, a Bayesian myopic policy may be sufficiently good in many cases, suggesting that a manager may be justified in prioritizing demand estimation over forward-looking inventory optimization in these cases. Third, more sophisticated policies may be needed when extreme demand changes are possible. Fourth, a manager worried about profit downside may opt for lower demand estimates.

Several extensions of our model may merit further research. First, it would be interesting to examine the case with censored demand, which makes the DM's future observations dependent on current ordering decisions. A conjecture is that the "stock more" result of Lariviere and Porteus (1999), Ding et al. (2002), and others may be accentuated in the presence of potential upward changes in demand. Second, interesting questions arise when a potential change-point is anticipated in the future (as opposed to the assumption we have made in most of the present paper that the potential change-point is at a known time in the past). Third, it would seem relevant to inventory management practice to allow for uncertainty in the timing of potential change-points in order to model demand shifts that occur for unobserved reasons. Fourth, we believe that the independentized bound idea may merit further investigation for other inventory models involving demand learning.

CHAPTER 3

Optimal Merchandise Testing with Limited Inventory

3.1 Introduction

Fashion products, which are characterized by highly uncertain demand, short life cycles, and often long lead-times, pose great challenges for retailers trying to match supply with demand. With only limited replenishment opportunities and inaccurate demand forecasts, retailers often end up with significant losses in profit due to either lost sales from stockouts or heavy price discounts needed to clear excess inventory at the end of the selling season. The past two decades have seen various initiatives by retailers to streamline inventory management for fashion products: for example, quick response (Fisher and Raman, 1996), backup agreements with manufacturers (Eppen and Iyer, 1997), and advanced booking discount programs (Tang et al., 2004), to name a few.

An effective strategy used in practice to improve initial demand forecasts is so-called "merchandise testing," wherein a retailer gathers demand information about new products by offering inventory for sale in selected stores of its network during a short testing period before the primary selling season starts. The retailer then uses the information gained during the test to construct more accurate demand forecasts and thereby to make improved ordering decisions in preparation for the main selling season (Fisher and Rajaram, 2000). There are three key decisions involved in such a merchandise test: (1) store selection—which store to select for the test? (2) inventory allocation—how to allocate inventory across the test stores? and (3) demand learning—how to construct demand forecasts based on the test sales data?

Fisher and Rajaram (2000) formulate the merchandise testing problem as a store clustering problem based on historical sales data of similar products and apply their methods to a large women's apparel retailer and two major shoe retailers in the United States. Designed for practical implementation, their model requires a few simplifying assumptions. First, Fisher and Rajaram implicitly assume retailers adopt a "depth test;" that is, they assume that ample inventory is placed at each test store to meet demand during the test period. Second, they assume that store-level demand forecasting for the selling season is accomplished using a linear function of test sales that is calibrated using historical sales data. In this chapter, we complement their work by relaxing these two assumptions. Our goal is to obtain insights into the interplay between limited inventory and (Bayesian) demand learning in a multi-location setting.

We consider a model in which a retailer, with multiple stores and a fixed quantity of available test inventory, manages a testing period followed by a selling period. Demands at each store in both the testing and selling periods are dependent on an unknown parameter that is common across stores and periods. Demands are independent once conditioned on this unknown parameter. Between the two periods, the retailer updates its demand forecast in a Bayesian fashion (over some prior distribution on the unknown demand parameter) based on its observations from the testing period, and it chooses inventory quantities for the selling period, during which we assume no further replenishment opportunities. The retailer's objective is to allocate the test inventory to stores at the beginning of the testing period such that the ex-ante expected total profit in the selling period is maximized. We assume no fixed costs of testing at a store and no inventory carryover, choosing instead to focus on the statistical forces at play.

When spreading test inventory across multiple locations, a retailer must choose not only how many (and which) stores to include in the test, but also the service level to target at each of the test stores. There is a natural incentive for the retailer to test at a large number of stores, as this yields a large *quantity* of demand observations. On the other hand, spreading a limited inventory among many stores compromises the service level targeted at each store during the test, and these service levels can impact the *quality* of information gathered in the test. How exactly service levels impact information quality depends on the retailer's visibility into the demand process. When the retailer has access to the timing of each sales transaction within a period (e.g., Jain et al., 2015), the available inventory at a test store limits the maximum number of transactions that can be observed. However, when the sales timing information is unobservable, the retailer only observes the total sales during the test period at each store. It is known that such sales information is an imperfect demand observation that is "censored" by the amount of inventory at each store (e.g., Lariviere and Porteus, 1999; Ding et al., 2002). In this case, the retailer has an incentive to concentrate inventory in fewer stores to achieve higher service levels, reduce censoring, and thereby enhance the quality of the demand observations.

We begin in §3.4 with the case in which sales timing information is observable. We first analyze a base case in which stores have stochastically identical demands. For general demand processes with a general prior, we prove that (in the absence of fixed costs) an optimal policy will never omit a store from a merchandise test. We further show that an "even-split" policy, which allocates the test inventory to all stores as evenly as possible, is always optimal under a Poisson demand process with a gamma prior. These results suggest the opposite of the traditional practice of a depth test which tends to avoid stockouts during the test. In fact, a high service level in the testing period is no longer a necessity when the retailer has access to sales timing information. We then extend our analysis to the non-identical-store case, in which stores may have diverse demand volumes. We characterize the monotone structure of the optimal allocation policy with respect to the relative demand volumes among stores.

A key intuition underlying the above results is that the availability of data on the timing of sales transactions largely ensures that store-level observations are of high quality, freeing the retailer to primarily consider observation quantity when allocating test inventory. Moreover, "quantity" in this context is best interpreted not in terms of the number of stores but rather in terms of the total number of sales observations—following Jain et al. (2015), each sales transaction can be viewed as an (exact) observation of an inter-arrival time in the underlying demand process. Therefore, an effective merchandise test is one that tends to maximize the quantity of sales transactions in the testing period across the store network. Motivated by this insight, we propose a "Max-Sales" heuristic allocation policy which maximizes the expected sales during the testing period.

We obtain contrasting results when the retailer does not have access to sales timing information. In this case, the form of the optimal allocation policy becomes contingent upon the amount of test inventory as well as the shape of the demand distribution, and it can be complex to characterize. Our analysis of the case with stochastically identical stores suggests that (1) when the amount of test inventory is small, a retailer should follow a "single-store" policy which allocates all of the test inventory to only one store; (2) when the amount of test inventory is large, an even-split policy is optimal. These results are established analytically assuming a continuous gamma-Weibull demand structure (with shape parameter exceeding one), and they are corroborated numerically for the case of Poisson demand with a gamma prior. Moreover, we find examples with a moderate amount of available test inventory in which the optimal allocation policy may stock unbalanced positive quantities in each store even though stores are otherwise identical.

These findings reveal a delicate tension between the quantity of stores included in the test and the quality of observations obtained from each one. Our results show that improving the quality of each demand observation is a higher priority than seeking a large observation quantity when the total test inventory is tightly constrained. This encourages the retailer to consolidate inventory in fewer stores to increase service levels in the testing period so as to avoid the negative impact of censoring on demand learning. This may be one justification (in addition to operational fixed costs) for adopting a concentrated test at a small number of stores. Motivated by these insights, for cases with heterogeneous stores we propose a "Service-Priority" heuristic that allocates test inventory to achieve a target service level during the testing period at as many stores as the inventory budget allows.

We evaluate our heuristic allocation policies in a numerical study by comparing their performance with the optimal policies. We consider two- and three-store problems in which the optimal allocations of the test inventory can be obtained through an exhaustive enumeration. Our numerical study indicates that when timing information is observable, the Max-Sales policy yields allocations that are extremely close to the optimal solutions; in fact, the maximum optimality gap in our study is 0.01% across both two-store and three-store instances. When timing information is unobservable, the Service-Priority policy also appears near-optimal in our numerical experiments, resulting in an average gap of 0.05% and a maximum gap of 1.02% for two-store instances, and an average gap of 0.08% and a maximum gap of 1.30% for three-store instances. Furthermore, we find that using inefficient allocations—for example, using the Max-Sales and Service-Priority heuristics in the wrong settings—can result in significantly suboptimal performance.

The remainder of this chapter is organized as follows. After a review of relevant literature in §3.2, §3.3 describes a general modeling framework for the merchandise testing problem with limited test inventory. §3.4 characterizes the structure of optimal allocation policies for test inventory when sales timing information is observable to the retailer. In §3.5, we analyze the case in which the retailer does not have sales timing information. We numerically evaluate the performance of our proposed heuristics in §3.6. §3.7 concludes the chapter with discussions of managerial insights and future research directions.

3.2 Literature Review

By studying the merchandise testing problem, our work contributes to a broad literature studying strategies for retailers to learn about demand for products with short life cycles and high demand uncertainty. Other examples include the "quick response" strategy of Fisher and Raman (1996), the "advanced booking discount" program modeled by Tang et al. (2004), and models allowing for advanced demand information that is updated over time (e.g., Wang et al., 2012).

There is a well-established body of research that jointly considers demand estimation and inventory optimization when unmet demand is lost and unobservable, or in other words, when demand observations are "censored." For a survey, we refer readers to Chen and Mersereau (2015). The majority of this literature focuses on single-location settings. Our work belongs to a substream of this literature that uses a Bayesian framework for demand estimation. Lariviere and Porteus (1999) analyze the Bayesian inventory problem with censored demand. To achieve tractability, they assume that the underlying demand distribution is from a family of so-called "newsvendor distributions" defined by Braden and Freimer (1991) and that a gamma prior is used. The dimensionality reduction scaling technique in Scarf (1960) and Azoury (1985), both assuming fully backlogged and exactly observable demand, is extended to the censored demand case under a Weibull demand distribution with a gamma prior. The problem receives continued exploration in Ding et al. (2002), Bensoussan et al. (2007), Lu et al. (2008), Chen and Plambeck (2008), and Chen (2010). Recently, Bisi et al. (2011) closely revisit the Bayesian inventory problem with censored observations and newsvendor distribution demand and confirm that Weibull is the only member of the newsvendor family for which optimal solutions are scalable. A common insight from this stream of literature is that the retailer should stock more than myopic order-up-to levels to better learn future demand information.

The recent paper of Jain et al. (2015) extends the literature on demand learning with censored observations by incorporating the timing of individual sales transactions. (Interestingly, Jain et al. (2015) use as a motivating example the Middle Eastern cosmetics brand Mikyajy, which uses merchandise testing at a single store to make profitable purchasing decisions prior to a full product launch.) In a parsimonious Bayesian multiperiod newsvendor framework, they prove that the "stock more" result continues to hold with the additional timing information. Furthermore, their numerical study shows that the use of timing observations significantly reduces losses in expected profit due to censoring. While their scope is again limited to a single-location setting, our research further extends their framework to a multi-location setting, which leads to tradeoffs that are non-existent in a single-location model.

A novel aspect of our work is the focus on demand learning (for a single product) across multiple locations, which is different from Caro and Gallien (2007), who consider a dynamic assortment problem with demand learning (at a single location) for multiple products. In this regard, our model is conceptually related to Harrison and Sunar (2014), who consider a firm choosing among several modes to learn the unknown value of a project for optimizing investment timing. The cost and quality of each learning mode in Harrison and Sunar (2014) are exogenously given, while our model seeks to maximize the value of demand learning subject to a resource constraint on the amount of test inventory available.

This chapter is also related to a large number of papers concerning inventory management at a warehouse serving multiple retail locations. A detailed review can be found in Agrawal and Smith (2009). Only a small subset of this literature is applicable to fashion products with short life cycles and high demand uncertainty. Two notable examples, in addition to Fisher and Rajaram (2000), are Agrawal and Smith (2013) and Gallien et al. (2015). Agrawal and Smith (2013) consider a two-period inventory model in which the retailer has multiple non-identical stores that share a common unknown parameter and uses a Bayesian scheme to update demand forecast. Gallien et al. (2015) also develop a two-period stochastic optimization model to determine initial shipments to stores at fashion retailer Zara, accounting for the allocation of leftover and replenished stock at a central warehouse to stores in the second period. Our work differs in several important ways. First, in the previous two contexts, the widespread rollout of a product occurs at the very beginning of the first period, which typically involves allocating a large amount of inventory to a large number of stores; however, the first period in our merchandise testing problem only involves distributing a very limited amount of inventory to a relatively small set of stores. Second, neither paper explicitly considers demand censoring when updating demand forecasts based on observations in the first period. Demand censoring is at the core of our study, as it leads to the quantity versus quality tradeoff at the heart of our research questions. Finally, neither of these papers considers using the timing of sales occurrences for demand learning.

At a high level, this chapter is related to research on the value of information and its structure in problems involving collecting information with limited resources, examples of which come from multiple disciplines including economics, simulation optimization, computer science, and decision science. For example, Frazier and Powell (2010) consider the Bayesian ranking and selection problem in which the decision maker allocates a measurement budget to choose the best among several alternatives. They find that spreading the measurement budget equally among alternatives can be paradoxically non-optimal when the prior is identical for each alternative, due to lack of concavity of the value of information. In our merchandise testing context, we also find that the value of information is not necessarily concave, in particular, when timing information is unobservable.

3.3 Model

In this section, we describe a general framework for the merchandise testing problem. Consider a retailer that tests and sells a single product through a chain of N stores. We model two periods, labeled 1 and 2 respectively, where period 1 represents the testing period and period 2 the main selling period. At the beginning of period 1, the retailer has $Q \in \mathbb{Z}_+$ units of inventory available in total to allocate to N stores for the merchandise test. We denote a feasible allocation of the test inventory by a vector $\mathbf{q} \in \mathcal{Q} = \{(q_1, \ldots, q_N) : \sum_{n=1}^{N} q_n \leq Q, q_n \in \mathbb{Z}_+\}$.

We assume that the testing period has length T and that demand arrives at each store n according to a renewal process, denoted by $\{D_n(t|\theta), 0 \leq t \leq T\}$, where $\theta \in \Theta$ is a parameter that is common to all N stores but unknown to the retailer. Let t_n^i denote the arrival time of the i-th demand at store n and $\tau_n^i = t_n^i - t_n^{i-1}$ the inter-arrival time between the i-th and the $(i-1)$-th demand. We assume that τ_n^i has probability density function (pdf) $\psi_n(\cdot|\theta)$, cumulative distribution function (cdf) $\Psi_n(\cdot|\theta)$, and complementary cdf $\overline{\Psi}_n(\cdot|\theta)$, and is independent of the demand processes at other stores if conditioned on θ. Cumulative demand $D_n(t)$ until time t has probability mass function $f_n(\cdot|t,\theta)$ and cdf $F_n(\cdot|t,\theta)$ (complementary cdf $\overline{F}_n(\cdot|t,\theta)$). We will use D_n and $D_n(T)$ interchangeably to denote the total demand in period 1 at store n. A Bayesian framework is employed to model demand learning and we assume that the retailer has a prior density $\pi(\theta)$ representing its initial belief about the unknown demand parameter θ.

In order to highlight the value of demand learning associated with the allocation of test inventory, we make the following simplifying assumptions. First, we assume no fixed costs of including a store in the merchandise test. It is intuitive that fixed costs would create an incentive to consolidate inventory; we focus instead on the statistical incentives to consolidate inventory in a few stores versus spreading it among many stores. Second, we assume that the revenue generated from sales in period 1 is negligible, as the testing period is typically short compared to the primary selling season. Finally, we do not consider inventory carryover from period 1 to period 2 for tractability reasons; in other words, the amount of leftover inventory from the test period is assumed to be negligible compared to the substantially larger order quantities for the main selling period.

At the end of period 1, the retailer makes an observation $X(\mathbf{q})$ which may depend on the test demand realization D_n, the allocated test inventory level q_n, and in some cases the timing of demand epochs $\{t_n^1, t_n^2, \ldots, t_n^{D_n}\}$ (or equivalently the inter-arrival times $\{\tau_n^1, \ldots, \tau_n^{D_n}\}$), at each store n. The retailer can obtain different types of observations during period 1 and we defer the details to §3.3.1. Let $l(X(\mathbf{q})|\theta)$ denote the likelihood of observing $X(\mathbf{q})$ for some θ and test inventory allocation $\mathbf{q}$. The retailer uses Bayes rule to update its knowledge about θ based on observation $X(\mathbf{q})$ over prior π as follows:

$$\hat{\pi}(\theta) = \pi(\theta) \circ X(\mathbf{q}) = \frac{l(X(\mathbf{q})|\theta)\pi(\theta)}{\int_\Theta l(X(\mathbf{q})|\omega)\pi(\omega)\mathrm{d}\omega}, \tag{3.1}$$

where $\hat{\pi}(\theta)$ is the updated posterior density.

Period 2 models the retailer's operations in the primary selling season. In essence, the retailer solves a newsvendor problem to choose the ordering quantity y_n for each store n to maximize the expected total profit generated by the entire chain based on its updated knowledge $\hat{\pi}$ about θ. Let $\hat{D}_n$ be the period 2 demand at store n. Our most general model is flexible in the demand structure of period 2 in that we only assume that $\hat{D}_n$ is distributed according to some cdf $\hat{F}(\cdot|\theta)$ which also depends on the unknown demand parameter θ and is independent of demand at other stores once conditioned on θ. We assume a unit selling price p and a unit procurement cost $c < p$, both of which are exogenously determined and apply universally to all N stores. The expected total profit in period 2 with respect to ordering quantities $\mathbf{y} = (y_1, \ldots, y_n)$ under belief $\hat{\pi}$ is thus given by

$$\hat{\Pi}(\mathbf{y}|\hat{\pi}) \triangleq E\left[\sum_{n=1}^{N} p\min\{\hat{D}_n, y_n\} - cy_n \middle| \hat{\pi}\right].$$

Denote by $\hat{\Phi}_n(\cdot)$ be the unconditioned cdf of demand at store n in period 2, i.e., $\hat{\Phi}_n(x) = \int_\Theta \hat{F}(x|\theta)\hat{\pi}(\theta)\mathrm{d}\theta$. It is straightforward to see that the optimal order quantity y_n^* for store n is given by the well-known newsvendor order quantity $y_n^* = \hat{\Phi}_n^{-1}(\frac{p-c}{p})$, where $\hat{\Phi}_n^{-1}(\cdot)$ is the inverse unconditioned cdf, i.e., $\hat{\Phi}_n^{-1}(r) = \min\{x : \hat{\Phi}_n(x) \geq r\}$.

Let $\hat{V}(\hat{\pi}) = \max_{\mathbf{y}} \hat{\Pi}(\mathbf{y}|\hat{\pi})$ be the optimal expected total profit in period 2. The retailer's problem at the beginning of period 1 is to find the optimal allocation $\mathbf{q} \in \mathcal{Q}$ of the total Q units of test inventory that maximizes its *ex-ante* expected profit $\Pi(\mathbf{q}|\pi) = E[\hat{V}(\pi \circ X(\mathbf{q}))|\pi]$, with the anticipation of an observation $X(\mathbf{q})$ being made after period 1.

3.3.1 Types of Demand Observations in Period 1

We present in this subsection two types of demand observations the retailer may receive during period 1.

Observations without Timing Information ($X^{NT}(\mathbf{q})$). This is the type of observations assumed by the majority of the literature: at the end of period 1, the retailer observes only the sales quantity at each store and whether a stockout has occurred. We denote by $s_n = \min\{D_n, q_n\}$ the sales quantity at store n and by $e_n = \mathbf{1}\{D_n \geq q_n\}$ a binary indicator of the store's stockout status at the end of period 1. The overall observation $X^{NT}(\mathbf{q}) = \{\mathbf{s}, \mathbf{e}\}$ is simply a collection of two vectors where $\mathbf{s} = (s_1, \ldots, s_n)$ and $\mathbf{e} = (e_1, \ldots, e_n)$. The superscript NT is for "No Timing." For each store n, the likelihood of observing sales quantity s_n at time T given some demand parameter θ is $f(s_n|T,\theta)$ if there is excess test inventory (i.e., $e_n = 0$) and is $\overline{F}(q_n - 1|T,\theta)$ otherwise (i.e., $e_n = 1$). Recall that we assume independent demand processes among N stores for any fixed θ. As a result, the likelihood of observing $X^{NT}(\mathbf{q})$ for some θ is given by

$$l(X^{NT}(\mathbf{q})|\theta) = \prod_{i=1}^{N} \left[(1-e_n) \cdot f(s_n|T,\theta) + e_n \cdot \overline{F}(q_n - 1|T,\theta)\right]. \tag{3.2}$$

Observations with Timing Information ($X^{T}(\mathbf{q})$). This type of observation is considered by Jain et al. (2015) in a single-store setting and we extend their definition to our multi-location setting. It contains not only stores' sales quantities and stockout statuses but also the timing of all sales occurrences. Let $\vec{\tau}_n = (\tau_n^1, \ldots, \tau_n^{s_n})$ denote the observed sequence of inter-arrival times between sales at store n. Let $X_n^T(q_n) = \{s_n, e_n, \vec{\tau}_n\}$ be the retailer's observation at store n where the superscript T stands for "Timing." If the retailer decides not to test at store n, i.e., $q_n = 0$, then it automatically stocks out (i.e., $e_n = 1$) and of course sees no sales (i.e., $s_n = 0$). Otherwise, if $q_n > 0$, the likelihood of it observing $X_n^T(q_n) = (s_n, 0, \vec{\tau}_n)$ at store n is $\prod_{i=1}^{s_n} \psi_n(\tau_n^i|\theta) \cdot \overline{\Psi}_n(T - \sum_{i=1}^{s_n} \tau_n^i|\theta)$ and of it observing $X_n^T(q_n) = (s_n, 1, \vec{\tau}_n)$ is $\prod_{i=1}^{s_n} \psi_n(\tau_n^i|\theta)$. Overall, the retailer's observation, $X^T(\mathbf{q}) = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$, is a collection of sales quantities, stockout statuses, and times between consecutive sales at all stores where we define $\vec{\tau} = (\vec{\tau}_1, \dots, \vec{\tau}_n)$. The likelihood of observing $X^T(\mathbf{q})$ for some θ is therefore given by

$$l(X^T(\mathbf{q})|\theta) = \prod_{n=1}^{N} \left[e_n \cdot \prod_{i=1}^{s_n} \psi_n(\tau_n^i|\theta) + (1 - e_n) \cdot \prod_{i=1}^{s_n} \psi_n(\tau_n^i|\theta) \cdot \overline{\Psi}_n(T - \sum_{i=1}^{s_n} \tau_n^i|\theta) \right], \quad (3.3)$$

where we use the conventions $\prod_{i=1}^{0} \psi_n(\tau_n^i|\theta) = 1$ and $\sum_{i=1}^{0} \tau_n^i = 0$ for the case $s_n = 0$.

3.3.2 Marginal Value of Learning of an Additional Unit of Test Inventory

It is intuitive that the retailer would always prefer to allocate all Q units of test inventory in period 1 so as to acquire as much demand information as possible. We formalize this intuition in the following lemma which shows that the retailer's ex-ante expected profit is increasing in the test inventory quantity allocated to any store. In other words, the marginal value of learning from an additional unit of total test inventory is always nonnegative.

To facilitate our presentation throughout the rest of this chapter, we introduce $\delta_n = (0, \dots, 0, 1, 0, \dots, 0)$ as an N-dimensional vector with only the n-th element being one and all other elements zero. We use this notation mainly to describe allocation modifications. For instance, allocation $\mathbf{q} + \delta_i - \delta_j$ modifies allocation $\mathbf{q}$ by sending one more unit of test inventory to store i and one less to store j.

Let $\Pi^T(\mathbf{q}|\pi)$ and $\Pi^{NT}(\mathbf{q}|\pi)$ denote the ex-ante expected profits for the cases with and without timing information.

Lemma 3.1. *The following hold for all π, $\mathbf{q} \in \mathbb{Z}_+^N$, and $n = 1, \dots, N$:*

(a) $\Pi^T(\mathbf{q}|\pi) \leq \Pi^T(\mathbf{q} + \delta_n|\pi)$;

(b) $\Pi^{NT}(\mathbf{q}|\pi) \leq \Pi^{NT}(\mathbf{q} + \delta_n|\pi)$;

(c) There exists an optimal allocation $\mathbf{q}$ *such that* $\sum_{i=1}^{N} q_n = Q$.

All proofs can be found in the appendix. We prove Lemma 3.1 using results from the statistics literature on comparisons of experiments (Blackwell, 1951, 1953). To prove Lemma 3.1(a), we define merchandise tests with allocations $\mathbf{q} + \delta_n$ and $\mathbf{q}$ as two statistical experiments, $\mathcal{E}^T$ and $\mathcal{F}^T$, when timing information is observable. The outcomes of the two experiments are demand observations $X^T(\mathbf{q} + \delta_n)$ and $X^T(\mathbf{q})$. We then establish that there exists a stochastic transformation from the distribution of $X^T(\mathbf{q} + \delta_n)$ to that of $X^T(\mathbf{q})$ (which is intuitive as the retailer observes more information with the additional unit of test inventory). As a result, experiment $\mathcal{E}^T$ is said to be *sufficient for* $\mathcal{F}^T$ and Lemma 3.1(a) immediately follows. The proof of part (b) uses a similar argument, and part (c) is an immediate corollary part of (a) and (b). Therefore, for the rest of the chapter, we narrow our focus to the set of allocations satisfying $\sum_{n=1}^{N} q_n = Q$ without loss of generality.

3.4 With Timing Information

In this section, we analyze the retailer's optimal policy for allocating test inventory when timing information is observable. We first examine in §3.4.1 the case in which stores have stochastically identical demand. Then we generalize our analysis to the case in which stores follow a more general demand structure.

3.4.1 Identical Stores

We consider a base case in which all stores are identical. More specifically, we assume that stores' demand processes in period 1 share a common inter-arrival time distribution, i.e., $\psi_n(\tau|\theta) = \psi(\tau|\theta)$ for all $n = 1, \ldots, N$. The identical-store case enables us to gain focused insights into the role of inventory allocation in gathering demand information from multiple locations. Practically, a group of identical stores may be interpreted as stores that have been clustered into a relatively homogeneous set in terms of demand or sales volume.

We first show in the following proposition that when the stores are identical and the retailer observes sales timing information, the retailer benefits from allocating a positive amount of test inventory to as many stores as possible under general renewal process demand with a general prior.

Proposition 3.1. *Suppose that stores are identical. Then for all π, the following hold when timing information is observable:*

(a) Let $\mathbf{q} = (q_1, \ldots, q_N)$ be a test inventory allocation such that $q_i \geq 2$ and $q_j = 0$ for some $i \neq j$. Then, $\Pi^T(\mathbf{q}|\pi) \leq \Pi^T(\mathbf{q} - \delta_i + \delta_j|\pi)$;

(b) There exists an optimal allocation $\mathbf{q}^* = (q_1^*, \ldots, q_N^*)$ such that $q_n^* > 0$ for $n = 1, 2, \ldots, \min\{Q, N\}$.

The main implication of Proposition 3.1 is that the retailer should cover as many stores as possible in a test without worrying about the potential to stock out at stores. This finding reveals an incentive for the retailer to deviate from a "depth test" that stocks high test inventory levels to avoid stockouts.

A formal proof appears in the appendix, but we sketch it here. We prove Proposition 3.1(a) by constructing two statistical experiments, $\mathcal{E}$ and $\mathcal{F}$, corresponding to the two inventory allocations, $\mathbf{q} - \delta_i + \delta_j$ and $\mathbf{q}$, respectively. As discussed in the sketch proof of Lemma 3.1, the result follows if we establish that there exists a stochastic transformation from the distribution of observation $X^T(\mathbf{q} - \delta_i + \delta_j)$ to that of $X^T(\mathbf{q})$. The intuition is as follows. When timing information is observable, the retailer learns the unknown demand parameter essentially through observations of inter-arrival times. Each realized sale gives the retailer an exact observation of a single inter-arrival time. Moreover, the retailer receives a censored observation of the inter-arrival time when a store *does not* stockout, as the time until the next demand epoch is truncated at the end of period 1. Therefore, by moving one unit of test inventory from store i to store j (with no inventory), the retailer increases both the probability of selling this unit and that of getting an accurate instead of a censored observation of the inter-arrival time. Both the quantity and the quality of observations collected during the test increase (in a stochastic sense), therefore the distribution of $X^T(\mathbf{q} - \delta_i + \delta_j)$ can be transformed to that of $X^T(\mathbf{q})$. Proposition 3.1(b) is an immediate corollary of part (a) given we have established in Lemma 3.1 that it suffices to consider allocation policies that distribute all test inventory to stores.

Proposition 3.1 hints at the desirability of an "even-split" policy which evenly distributes test inventory to all stores, thereby maximizing the expected sales, or equivalently, the number of uncensored inter-arrival time observations during period 1. This would be true if one could generalize Proposition 3.1(a) to any allocation $\mathbf{q}$ that has $q_i - q_j \geq 2$ without requiring $q_j = 0$. Unfortunately, the proof generally does not extend for $q_j > 0$, as a stochastic transformation from $X^T(\mathbf{q} - \delta_i + \delta_j)$ to $X^T(\mathbf{q})$ appears no longer possible—in other words, observations under allocation $\mathbf{q} - \delta_i + \delta_j$ do not always contain more information than that under allocation $\mathbf{q}$. Nevertheless, in the remainder of this subsection we present a result showing that the even-split policy is indeed optimal for an important special case.

Poisson Demand with a Gamma Prior. In the following, we assume that the inter-arrival times between consecutive demand epochs are exponentially distributed with an unknown rate parameter $\lambda > 0$, i.e., $\psi(\tau|\lambda) = \lambda e^{-\lambda\tau}$. In other words, the cumulative demand up to time t at each store n, $\{D_n(t|\lambda), t \geq 0\}$, is a Poisson process with unknown arrival rate λ, a demand process often assumed in academic research on retail inventory management. We further assume that the retailer uses a gamma prior with shape and rate parameters $\alpha > 0$ and $\beta > 0$, i.e., $\pi(\lambda) = \pi(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$. When timing information is observable, $\pi(\cdot|\alpha, \beta)$ is a conjugate prior for λ. More specifically, let $X^T = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ be a realized observation in period 1 under some allocation when timing information is observable. Then the posterior, updated based on X^T, is $\hat{\pi}(\lambda) = \pi(\lambda|\alpha, \beta) \circ X^T = \pi(\lambda|\alpha + \mathcal{S}, \beta + \mathcal{T})$, where

$$\mathcal{S} = \sum_{n=1}^{N} s_n \text{ and } \mathcal{T} = \sum_{n=1}^{N} \left[e_n \cdot \sum_{i=1}^{s_n} \tau_n^i + (1 - e_n)T \right]$$

constitute the sufficient statistics. Note that $\mathcal{S}$ and $\mathcal{T}$ are essentially the total sales quantity and the total sales duration across all stores, respectively.

The following proposition shows that the even-split policy is optimal for Poisson demand with a gamma prior when the retailer observes sales timing information.

Proposition 3.2. *Suppose that the demand at each store in period 1 is a Poisson process with unknown arrival rate λ and that the retailer has a gamma prior $\pi(\cdot|\alpha, \beta)$ with shape and rate parameters (α, β). Then, the following hold when timing information is observable:*

(a) Let $\mathbf{q} = (q_1, \ldots, q_N)$ *be a test inventory allocation such that* $q_i - q_j \geq 2$ *for some* $i \neq j$*. Then,* $\Pi^T(\mathbf{q}|\alpha, \beta) \leq \Pi^T(\mathbf{q} - \delta_i + \delta_j|\alpha, \beta)$*.*

(b) The "even-split" allocation $\mathbf{q}^* = (q_1^*, \ldots, q_n^*)$*, which allocates all* Q *units of test inventory to all* N *stores as evenly as possible, is optimal. In particular:*

(i) If $Q \leq N$*,* $q_i^* = 1$ *for* $i = 1, \ldots, Q$ *and* $q_j^* = 0$ *for* $j = Q + 1, \ldots, N$*;*

(ii) If $Q > N$*,* $q_i^* = \lfloor Q/N \rfloor + 1$ *for* $i = 1, \ldots, (Q \mod N)$ *and* $q_j^* = \lfloor Q/N \rfloor$ *for* $j = (Q \mod N) + 1, \ldots, N$*.*

The proof of Proposition 3.2(a) builds upon Proposition 3.1 and extends it to any allocation $\mathbf{q}$ that has $q_i - q_j \geq 0$ through a two dimensional induction on q_i and q_j. The induction relies on a first-step analysis which conditions on the time until the next demand arrival at either store i or j and treats the remaining testing period as a new merchandise test with a shorter testing periodand an updated prior. This first-step analysis relies on the memoryless property of Poisson processes and on the fact that $\mathcal{S}$ and $\mathcal{T}$ are sufficient statistics for the past demand information. Proposition 3.2(b) is a straightforward corollary of part (a).

The overall implication is that when sales timing information is used for demand learning in a merchandise test, the retailer need not aim for a high service level to avoid stockouts during the testing period. Instead, the retailer should allocate the limited test inventory to more stores so as to maximize the total sales, or equivalently, the quantity of exact inter-arrival time observations. The service level during the testing period is less of a concern because each sale individually reveals information about the demand distribution and has an equal value whether it is made in a store with a high or low service level.

3.4.2 Non-Identical Stores

In this subsection, we extend our analysis to the more general case where stores may be non-identical. We model non-identical demand as follows. We assume that stores' inter-arrival times are stochastically ordered in a consistent way conditioned on any value of the unknown demand parameter. Without loss of generality, we label the stores such that their demand inter-arrival times are increasing in the sense of first-order stochastic dominance. In particular, we assume that

$$\overline{\Psi}_1(\tau|\theta) \leq \overline{\Psi}_2(\tau|\theta) \leq \cdots \leq \overline{\Psi}_N(\tau|\theta) \tag{3.4}$$

for all $\tau \geq 0$ and $\theta \in \Theta$. Recall that $\overline{\Psi}_n(\cdot|\theta)$ is the complementary cdf of the inter-arrival times at store n given a fixed θ and is assumed to be known to the retailer. This assumption also implies that stores' demands are decreasing in the sense of first-stochastic dominance, i.e., $\overline{F}_1(x|T,\theta) \leq \overline{F}_2(x|T,\theta) \leq \cdots \leq \overline{F}_N(x|T,\theta)$ for all $x \geq 0$ and $\theta \in \Theta$. In this formulation, one can interpret θ as the overall market potential of the product. The retailer does not know θ, but knows the market share of each store, which may be relatively more stable and predictable than the overall demand.

The following proposition extends Proposition 3.1 and sheds light on which stores the retailer should prefer when allocating test inventory with timing information observable.

Proposition 3.3. *Suppose that stores are non-identical such that $\overline{\Psi}_1(\tau|\theta) \leq \overline{\Psi}_2(\tau|\theta) \leq \cdots \leq \overline{\Psi}_N(\tau|\theta)$ for all $\tau \geq 0$ and $\theta \in \Theta$. Then for all π, the following hold when timing information is observable:*

(a) Let $\mathbf{q} = (q_1, \ldots, q_N)$ be a test inventory allocation such that $q_i = 0$ and $q_j \geq 1$ for some $i < j$. Then, $\Pi^T(\mathbf{q}|\pi) \leq \Pi^T(\mathbf{q} + \delta_i - \delta_j|\pi)$;

(b) There exists an optimal allocation $\mathbf{q}^* = (q_1^*, \ldots, q_N^*)$ such that $q_n^* > 0$ for $n = 1, \ldots, m$ and $q_n^* = 0$ for $n > m$, where m is some number in $\{1, \ldots, N\}$.

Proposition 3.3 indicates that the retailer should always allocate test inventory to stores with higher demand before testing at stores with lower demand. This is in line with the intuition we have gained in §3.4.1 that the retailer should maximize its test sales to maximize the value of the test when timing information is observable. This result is also useful if the retailer has an additional constraint on the maximum number of stores to test, say, $M < N$ stores. In that case, instead of considering all subsets with at most M stores, the number of which is $\sum_{m=1}^{M} \binom{N}{m}$ in total, the retailer need examine only M subsets, each containing the m stores with the largest relative demand, $m = 1, 2, \ldots, M$.

Given the above proposition, it is natural to expect that the optimal quantities of test inventory allocated to stores should be ranked according to stores' relative demand volumes. That is, the retailer should send the most test inventory to store 1, the second most to store 2, and so forth. We prove that this conjecture holds for an important special case involving Poisson demand processes. Before formally stating the proposition, we first introduce our gamma-Poisson demand model for the non-identical-store setting.

Non-Identical Poisson Demand with a Gamma Prior. We assume that the demand inter-arrival times at store n are exponentially distributed with rate $\gamma_n\lambda$, i.e., $\psi_n(\tau|\lambda) = \gamma_n\lambda e^{-\gamma_n\lambda\tau}$, where $\lambda > 0$ is unknown but $\gamma_n > 0$ is known to the retailer. In other words, the cumulative demand up to time t at each store n, $\{D_n(t|\lambda), t \geq 0\}$, is a Poisson process with (partially) unknown arrival rate $\gamma_n\lambda$. In addition, we assume that $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N$. We call γ_n the *relative demand coefficient* of store n. The retailer still uses a gamma prior $\pi(\lambda|\alpha, \beta)$ on λ with shape and rate parameters $\alpha > 0$ and $\beta > 0$.

One can easily verify that the above Poisson demand model satisfies our general non-identical-store assumption in (3.4) by noting that the complementary cdf of inter-arrival times at store n is given by $\overline{\Psi}_n(\tau|\lambda) = e^{-\gamma_n\lambda\tau}$. Furthermore, the gamma distribution is still a conjugate prior for this non-identical Poisson demand process. Let $X^T = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ be a realized observation in period 1 under some allocation when timing information is observable. Then the posterior, updated based on X^T, is $\hat{\pi}(\lambda) = \pi(\lambda|\alpha, \beta) \circ X^T = \pi(\lambda|\alpha + \mathcal{S}, \beta + \mathcal{T})$, where

$$\mathcal{S} = \sum_{n=1}^{N} s_n \qquad \text{and} \qquad \mathcal{T} = \sum_{n=1}^{N} \gamma_n \cdot \left[e_n \cdot \sum_{i=1}^{s_n} \tau_n^i + (1 - e_n)T \right].$$

The sufficient statistics still have two dimensions with one being the total sales quantity and the other the "weighted" total sales duration across all stores where the weights are the relative demand coefficients.

We prove in the following proposition that in the gamma-Poisson demand model the optimal quantities of test inventory allocated to stores are ordered by stores' relative demand coefficients.

Proposition 3.4. *Suppose that the inter-arrival times at store n are exponentially distributed with rate parameter $\gamma_n\lambda$ where λ is unknown but $\gamma_1 \geq \ldots \geq \gamma_n$ are known, and that the retailer has a gamma prior on λ with shape and rate parameters (α, β). The following hold when timing information is observable:*

(a) Let $\mathbf{q} = (q_1, \ldots, q_N)$ be a test inventory allocation such that $q_i < q_j$ for some $i < j$. Then, $\Pi^T(\mathbf{q}|\alpha, \beta) \leq \Pi^T(\mathbf{q} + \delta_i - \delta_j|\alpha, \beta)$;

(b) There exists an optimal allocation $\mathbf{q}^* = (q_1^*, \ldots, q_N^*)$ such that $q_1^* \geq q_2^* \geq \cdots \geq q_N^*$.

The structure of the optimal allocation policy characterized by Proposition 3.4 provides an intuitive guideline for practitioners to distribute test inventory in merchandise testing: allocate more inventory to the stores with higher demand. However, computing the exact optimal allocation quantities remains difficult due to the combinatorial nature of the problem. We propose an easy-to-implement heuristic policy in the following subsection.

3.4.3 The Max-Sales Heuristic

In this subsection, we propose a heuristic policy named "Max-Sales" based on the intuition developed in §3.4.1 that with timing information observable, a good policy tends to maximize the sales during the testing period.

The Max-Sales policy is a greedy heuristic which sequentially allocates Q units of test inventory to N stores such that each unit of product is sent to the store having the highest (unconditioned) probability of selling that additional unit. Let $\phi_n(x)$ be the unconditioned probability mass function of demand being x at store n in period 1, i.e., $\phi_n(x) = \int_\Theta f_n(x|T, \theta)\pi(\theta)\mathrm{d}\theta$. We denote by $\Phi_n(x)$ the corresponding unconditioned cdf, i.e., $\Phi_n(x) = \sum_{u=0}^{x} \phi_n(u)$, and let $\overline{\Phi}_n(x)$ denote the unconditioned complementary cdf. The algorithm of the Max-Sales policy is as follows:

$\triangleright$ `Max-Sales Heuristic`

$q_1, \ldots, q_N \leftarrow 0$;

`for` $i \leftarrow 1$ `to` Q

```
        n* ← min{n : Φ̄_n(q_n) ≥ Φ̄_m(q_m) for all m = 1,...,N};
        q_{n*} ← q_{n*} + 1;
end
```

Proposition 3.5. *Suppose that stores are non-identical such that $\overline{\Psi}_1(\tau|\theta) \leq \overline{\Psi}_2(\tau|\theta) \leq \cdots \leq \overline{\Psi}_N(\tau|\theta)$ for all $\tau \geq 0$ and $\theta \in \Theta$. The following hold:*

(a) The Max-Sales heuristic yields an allocation $\mathbf{q}^{MS} = (q_1^{MS}, \ldots, q_N^{MS})$ that maximizes the expected total sales in period 1;

(b) $q_1^{MS} \geq q_2^{MS} \geq \cdots \geq q_N^{MS}$.

Proposition 3.5(a) shows that our greedy Max-Sales heuristic indeed maximizes the expected total sales during period 1 under the general non-identical demand introduced in §3.4.2. Proposition 3.5(b) guarantees that the Max-Sales allocation is monotonic, which which shares the same structure as the optimal policy under gamma-Poisson demand as we have proved in Proposition 3.4(b). The Max-Sales heuristic is easy-to-compute and applies to general demand processes and priors. Based on our numerical experience, this heuristic performs extremely well, coinciding with the optimal policy in almost all cases (see a detailed discussion in §3.6.2).

3.5 Without Timing Information

We discuss in this section the optimal test inventory allocation policy when the retailer does not observe sales timing information.

3.5.1 Identical Stores

When timing information is unobservable as is commonly assumed in the classic Bayesian inventory literature with demand censoring, analyzing the optimal allocation policy becomes particularly challenging. Even in the single-location setting, it is well-known that computing the optimal inventory policy is difficult (Bisi et al., 2011). In addition, we lose conjugacy under the discrete demand assumption, which makes it challenging even to compute the Bayes update between periods.

To achieve tractability, instead of assuming discrete, renewal process demand, we turn to a continuous Weibull demand distribution with a gamma prior, a parsimonious demand model widely adopted in the Bayesian inventory control literature with demand censoring. We will corroborate our key results numerically for gamma-Poisson demand in §3.5.2. For the present analysis we assume that the demand at each store in both periods is Weibull distributed with known shape parameter $k > 0$ and unknown scale parameter $\theta > 0$, and that the retailer has a gamma distributed prior on θ with shape and scale parameters (a,S) at the beginning of period 1. Note that the length T of period 1 is rendered irrelevant under this assumption. In particular,

$$f_n(x|\theta) = \hat{f}_n(x|\theta) = \theta x^{k-1} e^{-\theta x^k}, \qquad \text{and } \pi(\theta) = \frac{S^a}{\Gamma(a)} \theta^{a-1} e^{-S\theta}.$$

The following proposition partially characterizes the optimal allocation policy under gamma-Weibull demand when timing information is unobservable.

Proposition 3.6. *Suppose that the demand at each store in both periods is Weibull with shape parameter $k > 0$ and unknown scale parameter θ, and that the retailer has a gamma prior on θ with shape and scale parameters (a,S). The following hold for all $a > \frac{1}{k}$ and $S > 0$ when timing information is unobservable:*

(a) If $0 < k \leq 1$, the even-split allocation, $q_n^* = Q/N$ for all n, is optimal;

(b) If $k > 1$:

- *(i) there exists $Q_0 > 0$ such that for $0 \leq Q < Q_0$, a "single-store" allocation is optimal, i.e., $q_i^* = Q$ for some i and $q_n^* = 0$ for all $n \neq i$;*
- *(ii) the even-split allocation, $q_n^* = Q/N$ for all n, becomes optimal as $Q \to \infty$.*

We can see from Proposition 3.6 that the form of the optimal allocation policy is generally complex when timing information is unobservable. It may depend on the total test inventory Q as well as the shape of the demand distribution.

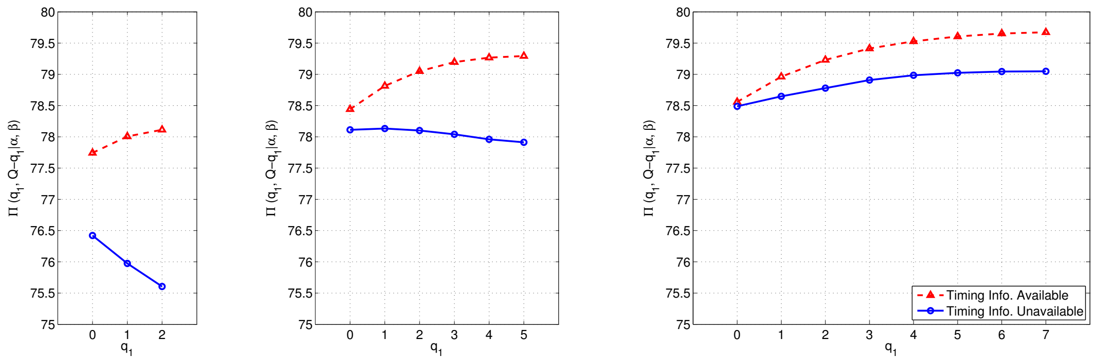
When $k \leq 1$, the demand density is strictly decreasing with a shape similar to that of an exponential density. (The exponential distribution is itself a special case of the Weibull distribution with $k = 1$.) In this case, the even-split allocation is always optimal regardless of the total test inventory Q. In fact, the ex-ante expected profit $\Pi^{NT}(\mathbf{q}|a, S)$ is jointly concave in the allocation $\mathbf{q}$. This resembles Bisi et al. (2011)'s result that the expected cost is convex in the inventory level under gamma-exponential demand in a single-store setting.

When $k > 1$, the demand density has a unimodal shape. In this case, the optimal policy is "single-store," i.e., allocating all Q units of test inventory to only one store, when Q is sufficiently small relative to demand. We note that Q_0 is a constant that depends on demand parameters k, a, and S. The even-split allocation becomes optimal for sufficiently large Q.

In contrast with the case in which timing information is observable, the pursuit of the quality and the quantity of demand observations need to be carefully balanced in the absence of timing information. Sending more test inventory to a store increases the service level and reduces the probability of demand being censored, thereby improving the quality of the demand observation obtained from the store. But with a fixed overall quantity of test inventory, this also means either sending less test inventory to some other store, which degrades the observation quality at that store, or excluding one or more stores from the test, which reduces the quantity of demand observations. When $k > 1$, Bisi et al. (2011) show in a single-store setting that the expected cost can be non-convex in the inventory level; we observe a similar phenomenon in our model, where the ex-ante expected profit $\Pi^{NT}(\mathbf{q}|a, S)$ is non-concave in each q_n. As a result, the retailer gains little demand information from a store until it stocks sufficient test inventory at the store. Spreading the test inventory equally to all stores may not be beneficial, as the increase in the total observation quantity may not compensate for the significant loss in the quality of demand observations at each store.

Proposition 3.6 suggests that the retailer may want to consolidate test inventory in a few stores to achieve a sufficiently high service level during the testing period. In other words, we provide a theoretical justification, in addition to fixed costs, for the practice of avoiding stockouts in merchandise testing. We remark that this tendency to consolidate test inventory is present even though we assume zero fixed cost of conducting a test at a store.

Figure 3.1: Ex-ante expected profit $\Pi(q_1, Q - q_1 | \alpha, \beta)$ as a function of q_1 in a two-identical-store example under Poisson demand with a gamma prior when $Q = 5, 10, 15$.



(a) $Q = 5$ (b) $Q = 10$ (c) $Q = 15$

Note. $N = 2, \alpha = 2, \beta = 0.4, p = 10, c = 1, T = 1.$

3.5.2 Identical Stores: Numerical Illustration

In the following, we numerically test and illustrate the findings of Proposition 3.6 for example merchandise testing problems with identical stores, Poisson demand, and a gamma prior. To have a contrasting comparison, we also include the case with observable sales timing information in this numerical illustration. We plot in Figure 3.1 the ex-ante expected profit for a two-store problem as a function of q_1, the units of test inventory allocated to store 1. The dashed lines with triangle markers are for the case in which timing information is observable, whereas the solid lines with circle markers are for the case in which timing information is unobservable. Lemma 3.1 implies that it is sufficient to consider allocations $\mathbf{q} = (q_1, q_2)$ that have $q_2 = Q - q_1$. The shape and rate parameters of the retailer's prior are $\alpha = 2$ and $\beta = 0.4$, so for each store the expected arrival rate $E[\lambda] = \alpha/\beta = 5$. The profits are plotted for $Q = 5$, 10, and 15, respectively. For each Q value, we only plot q_1 from 0 to $\lfloor Q/2 \rfloor$. Given that the two stores are identical, the profits for $q_1 \in \{\lfloor Q/2 \rfloor + 1, \ldots, Q\}$ mirror those shown.

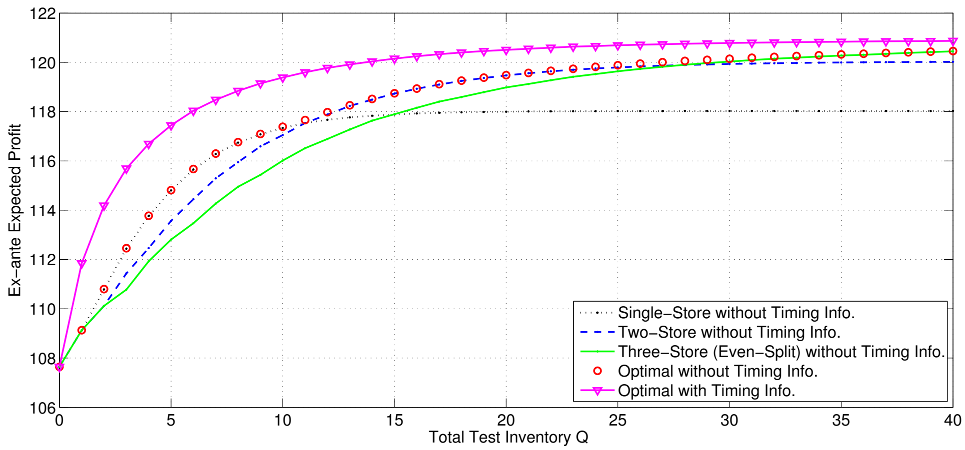
We observe that even-split allocations (i.e., (2,3) when $Q = 5$, (5,5) when $Q = 10$, and (7,8) when $Q = 15$) are always optimal when timing information is observable, consistent with Proposition 3.2. However, when timing information is not observable, the structure of the optimal allocation may differ as the total test inventory Q varies. The results in

Figure 3.1(a) and 3.1(c) are consistent with the extreme cases of Proposition 3.6: if Q is small compared to the demand ($Q = 5$), the single-store allocation (0,5) maximizes profits; if Q is large ($Q = 15$), the even-split allocation (7,8) maximizes profits. Figure 3.1(b) highlights the complexity of allocating test inventory without timing information: if Q is at a moderate level ($Q = 10$), an unbalanced allocation (1,9), which allocates unequal, positive quantities of test inventory to stores even though the stores are identical, can be optimal. Nonetheless, the additional benefit of using the (1,9) allocation is small compared with the single-store allocation (0,10). We find that the additional benefit of an unbalanced allocation is typically small; also, the region for an unbalanced allocation to be optimal is typically very small.

Figure 3.1 also yields insights into the value of using timing information for demand learning in the merchandise test. Naturally, the added value of timing information is always positive under the same allocation of test inventory, and it decreases as Q increases. An important observation is that the additional value of timing information hinges on the allocation of test inventory. Figure 3.1(b) gives an example in which the use of timing information may bring little extra value if the retailer employs a single-store rather than the optimal even-split allocation. Interestingly, Figure 3.1(c) shows that the ex-ante expected profit of using a single-store allocation with timing information is *lower* than that of using the even-split policy without timing information. In other words, a suboptimal allocation of test inventory may completely negate the advantage of observing sales timing information.

Figure 3.2 shows the ex-ante expected profit as a function of the total test inventory Q under various allocation policies in a three-store example with identical stores, Poisson demand, and a gamma prior. The parameters are the same as those used to generate Figure 3.1 except that we increase the number of stores to $N = 3$. We consider the optimal allocation when timing information is observable (i.e., the even-split allocation) and the following four allocation policies when timing information is unobservable: (1) the single-store policy; (2) the "two-store" policy (i.e., $\mathbf{q} = (\frac{Q}{2}, \frac{Q}{2}, 0)$ if Q is even or $\mathbf{q} = (\frac{Q+1}{2}, \frac{Q-1}{2}, 0)$ if Q is odd); (3) the "three-store" policy, or equivalently, the even-split policy; (4) and the optimal allocation. We obtain the optimal ex-ante expected profit for each Q value when timing information is unobservable through an exhaustive enumeration of all allocations satisfying $q_1 + q_2 + q_3 = Q$. Again, Figure 3.2 reinforces our insights from Proposition 3.6:

Figure 3.2: Ex-ante expected profit in a three-identical-store example as a function of total test inventory Q under various allocation policies.



Note. $N = 3, \alpha = 2, \beta = 0.4, p = 10, c = 1, T = 1.$

the single-store allocation is optimal when Q is small ($Q < 10$) while the even-split allocation is optimal when Q is large ($Q > 30$). We also find that the use of timing information may increase the ex-ante expected profit, potentially by a significant margin when Q is limited. The additional value of timing information diminishes as Q increases.

We further notice in Figure 3.2 that when timing information is unobservable, the optimal ex-ante expected profit closely follows the envelope of the profits achieved by a class of "m-store" allocations, which allocate test inventory to m out of N stores as evenly as possible. The optimal allocation can be something other than an m-store allocation: e.g., neither the single-store nor the two-store allocation is optimal at $Q = 11$; similarly, both the two-store and the three-store allocation are suboptimal at $Q = 29$. However, the loss in the ex-ante expected profit is negligible in both cases if the retailer chooses either m-store allocation instead of the optimal allocation. This suggests that a retailer without access to sales timing information may start with a single-store allocation and gradually add more stores to the test as the total test inventory increases. The intuition is that the retailer need maintain a sufficient service level at test stores during the testing period to ensure the quality of the collected demand observations before seeking additional observation quantity by increasing the number of stores to test.

3.5.3 The Service-Priority Heuristic

Given the complexity of the optimal allocation policy for test inventory even in the identical-store case, computing the exact optimal allocation quantities appears to be out of reach for general demand processes when timing information is unobservable. Instead, we use the intuition uncovered in previous subsections to develop a heuristic policy for allocating the test inventory.

We have learned in §3.5.1 and §3.5.2 that the optimal allocation policy strikes a balance between observation quantity (i.e., number of stores to test) and quality (i.e., service level at each store tested). We develop our heuristic with this tradeoff in mind. The idea is to achieve a certain target service level r at as many stores as possible in period 1, where r is a tunable parameter. For this reason, we name our heuristic the "Service-Priority" policy. In particular, the heuristic allocates test inventory starting from store N, the store with the lowest relative demand. The motivating logic is that the retailer can always use less inventory to achieve the target service level r in a store with lower demand.

Let $\Phi_n^{-1}(r)$ be the inverse unconditioned cdf of demand at store n in period 1, i.e., $\Phi_n^{-1}(r) = \min\{x : \Phi_n(x) \geq r\}$. The algorithm of the Service-Priority policy is the following:

```
▷ Service-Priority Heuristic
n←N;
while Q > 0
    q_n ← min{Q, Φ_n^{-1}(r)};
    Q←Q − q_n;
    n←n − 1;
end
```

The question remains how to choose the target service level r for the Service-Priority policy. A naïve method would be to arbitrarily specify a relatively high r. One could also perform a search over a set of candidate r values to identify the r value that maximizes the ex-ante expected profit. In §3.6.3, we compare the performance of both methods for choosing r.

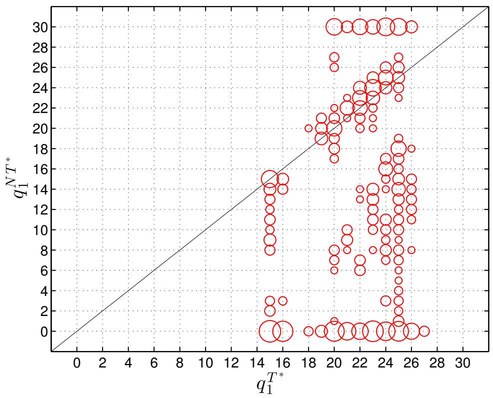
3.6 Performance of Heuristic Policies

In this section, we evaluate the performance of the Max-Sales policy and the Service-Priority policy proposed in §3.4.3 and §3.5.3 in an extensive numerical study. We consider a merchandise testing problem under non-identical Poisson demand with a gamma prior as introduced in §3.4.2. More specifically, the demand at each store n in both periods is a Poisson process with arrival rate $\gamma_n\lambda$, where λ is unknown but γ_n, the relative demand coefficient, is known. We normalize the γ_n's such that $\sum_{n=1}^{N}\gamma_n = 1$. The prior distribution of λ is gamma with shape and rate parameters (α, β).

We first report results for a set of two-store instances ($N = 2$). We choose values of the parameters to construct a large set of instances. We vary γ_1 and γ_2 such that γ_1/γ_2 takes value in {1,2,3,4,5}. The stores are identical if $\gamma_1 = \gamma_2 = 0.5$ and are non-identical otherwise. The shape parameter α reflects the degree of uncertainty about λ and takes values in $\{1, 2, 4, 8\}$, indicating a coefficient of variation of λ in $\{1, \frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{1}{\sqrt{8}}\}$. In order to illustrate the inventory allocations in a unified scale, we fix the total test inventory Q at 30 and vary the demand level. We choose the value of the rate parameter β such that the expected total arrival rate $E[(\gamma_1 + \gamma_2)\lambda] = E[\lambda] = \alpha/\beta \in \{10, 20, 30, 40, 50, 60\}$. The unit selling price is fixed at $p = 20$, and we vary the unit purchasing cost c in $\{1, 2, \ldots, 10\}$ to have a range of newsvendor ratios $(p - c)/p$ in $\{0.50, 0.55, \ldots, 0.95\}$ targeted in period 2. This gives us a set of 1,200 instances in total. We also briefly report results for a set of $N = 3$ instances in §3.6.4.

Throughout the section, we denote by q_n^σ the test inventory quantity allocated to store n under some policy σ. We use T^* and NT^* to denote the optimal allocation policy for the cases with and without timing information, respectively. We refer to the "optimality gap" (or "gap", for short) of a policy σ for a problem instance as the percentage gap with respect to the ex-ante expected profit under the optimal allocation policy. For example, the optimally

Figure 3.3: Comparison of the optimal test inventory allocation policies for the cases with and without timing information ($N = 2$).



Note. The area of a bubble is linear in the number of instances at the bubble's coordinate.

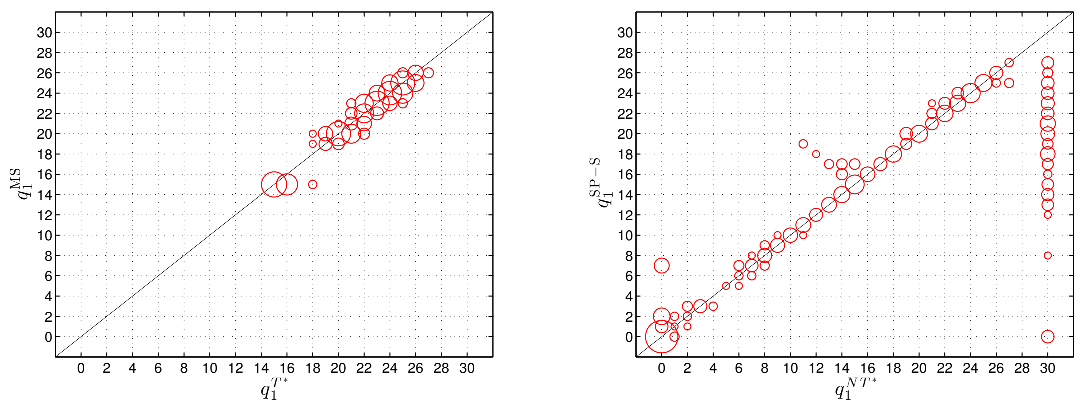
gap of a policy σ when timing information is observable is given by $(\Pi^{T^*} - \Pi^{\sigma})/\Pi^{T^*} \times 100\%$, where Π^{σ} is the ex-ante expected profit of policy σ.

3.6.1 Optimal Test Inventory Allocations

For all the two-store instances, we compute through an exhaustive enumeration the optimal quantity of test inventory allocated to store 1 when timing information is observable, $q_1^{T^*}$, and that when timing information is unobservable, $q_1^{NT^*}$. (Recall from Lemma 3.1 that when $N = 2$, the optimal allocation quantities to store 2 are just $Q - q_1^{T^*}$ and $Q - q_1^{NT^*}$, respectively.) Figure 3.3 shows a bubble plot of $(q_1^{T^*}, q_1^{NT^*})$ pairs for all instances. A bubble at $(q_1^{T^*}, q_1^{NT^*})$ means that there is at least one instance for which the optimal allocation quantity for store 1 is q_1^{T*} if the retailer observes timing information and is q_1^{NT*} if the retailer does not. The size of each bubble indicates the total number of such instances out of the 1,200 total instances explored.

Figure 3.3 demonstrates the pronounced difference between the behavior of the optimal allocation policy when timing information is observable and that when timing information is unobservable. Also, we observe that $q_1^{T*} \geq 15$ across all instances, which is consistent with Proposition 3.2 given that our instances have $\gamma_1 \geq \gamma_2$ and $Q = 30$.

Figure 3.4: Comparison of the heuristic and the optimal allocation policies ($N = 2$).



(a) Timing Information Observable

(b) Timing Information Unobservable

Note. The area of a bubble is linear in the number of instances at the bubble's coordinate.

3.6.2 Performance of the Max-Sales Heuristic

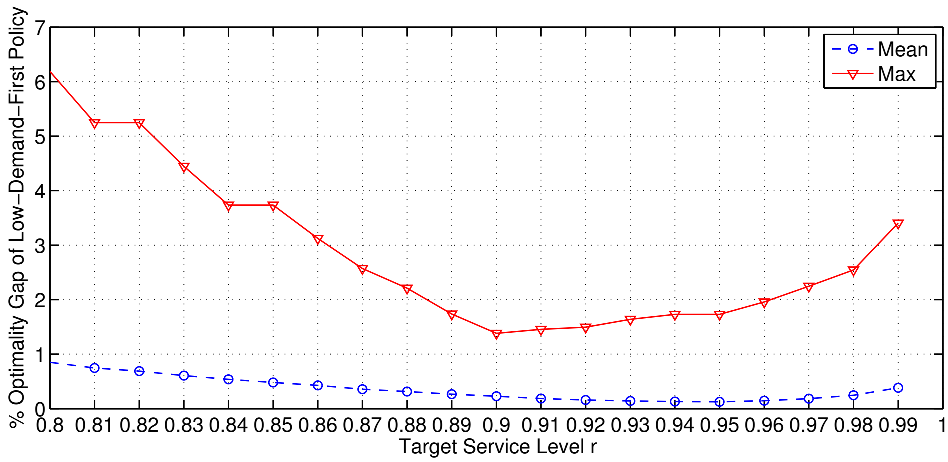
We test the performance of the Max-Sales heuristic proposed in §3.4.3 against the optimal allocation policy in the case where timing information is observable. Let "MS" denote the Max-Sales policy. Figure 3.4(a) shows a bubble plot of $(q_1^{T^*}, q_1^{\mathrm{MS}})$ pairs for all 1,200 instances. We observe that the bubbles are all located close to the diagonal line, implying that the Max-Sales policy closely follows the optimal allocation policy. The maximum difference between q_1^{T*} and q_1^{MS} for an instance is 3 units—at ($q_1^{T^*} = 18$, $q_1^{\mathrm{MS}} = 15$). In addition, We compute the optimality gap of the Max-Sales policy for each instance. The average gap is 0.0007% and the maximum is only 0.01% across all 1,200 instances[1]. The evidence lends strong support to the near-optimality of the Max-Sales policy when timing information is observable.

3.6.3 Performance of the Service-Priority Heuristic

We investigate the performance of the Service-Priority heuristic proposed in §3.5.3 against the optimal allocation policy in the case where timing information is unobservable.

[1]We compute all the ex-ante expected profits using Monte Carlo simulation with 1,000,000 trials. Therefore, the extremely small optimality gaps raise a natural question whether the Max-Sales heuristic is indeed optimal under Poisson demand with a gamma prior. We do not seem to have a proof (or an exact counterexample) for this claim and view it as an open question.

Figure 3.5: Optimality gaps of the Service-Priority policy with various values of target service level r when timing information is unobservable ($N = 2$).



We first consider the naïve approach in which the retailer arbitrarily chooses a relatively high target service level r and applies it uniformly across all instances. Figure 3.5 shows the summary statistics of the optimality gaps of Service-Priority policies under various specifications for the target service level r when timing information is unobservable. We observe that as r varies from 0.80 to 0.99 in increments of 0.01, the lowest mean optimality gap is 0.13% (achieved at $r = 0.95$) and the lowest maximum gap is 1.38% (achieved at $r = 0.90$). These results imply that the naïve variant of the Service-Priority policy can be reasonably satisfactory as long as the retailer chooses a relatively but not extremely high r.

We then examine an alternative approach in which a search over a set of r values is performed for each instance to find the r value that maximizes the ex-ante expected profit. We numerically test the performance of this variant of the Service-Priority policy, abbreviated to SP-S, with a search over $r \in \{0.50, 0.51, \ldots, 0.99\}$ for each of the 1,200 instances. Figure 3.4(b) shows a bubble plot of ($q_1^{S^*}$,$q_1^{\text{SP-S}}$) pairs for all the instances. We observe that in most of the cases the SP-S policy closely follows the optimal policy, with a few exceptions in which the optimal policy allocates zero units of test inventory to the low demand store 2. In addition, we compute the optimality gap for each instance. The average gap of the SP-S policy is only 0.05% across all 1,200 instances and the maximum gap is 1.24%. As expected, the performance of the Service-Priority policy further improves after we include a search for a better target service level r for each instance.

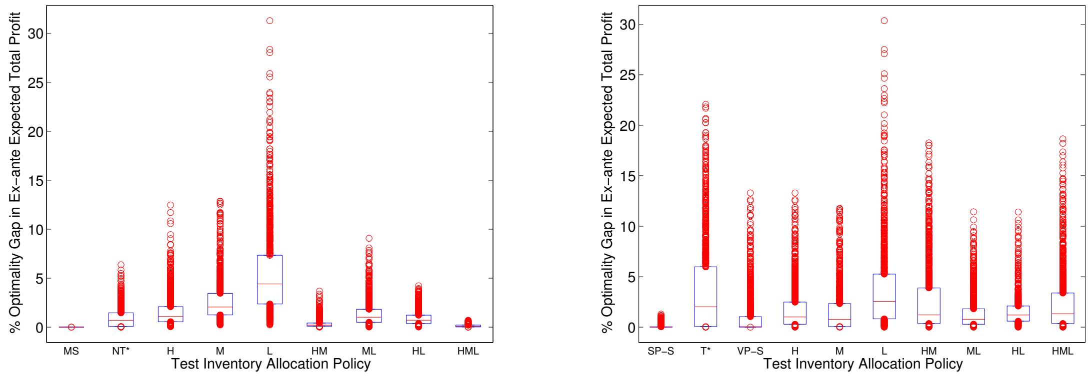
3.6.4 Three-Store Instances

We report in this section the results for another set of instances in which $N = 3$. The parameter setting is the same as in the $N = 2$ instances except for the relative demand coefficients. For ease of exposition, we label the three stores H, M, and L, respectively, which stand for relatively high, medium, and low demand. We vary their relative demand coefficients such that $\gamma_H/\gamma_L \in \{1, 2, 3, 4, 5\}$. The medium demand store's coefficient is set at $\gamma_M = (\gamma_H + \gamma_L)/2$. As in the two-store instances, we also normalize the relative demand coefficients such that $\gamma_H + \gamma_M + \gamma_L = 1$.

We compare the performance of the Max-Sales policy and the Service-Priority policy with a number of benchmark heuristics. We include a set of simple heuristics, each named with a set of the store labels, that always allocate the test inventory evenly to the stores in its name. For example, an H policy allocates $Q = 30$ units of test inventory to store H, an ML policy allocates 15 units each to store M and L, and so forth. We also include a "Volume-Priority" policy (VP-S), a modification of the SP-S policy that gives priority to stores with higher, instead of lower, demand for test inventory allocation. In addition, we include the optimal policy without timing information (NT^*) as a benchmark for the case in which timing information is observable, and also the optimal policy with timing information (T^*) as a benchmark for the case in which timing information is unobservable.

The optimality gaps are plotted in Figure 3.6. When timing information is observable, we observe that the Max-Sales performs extremely well with a mean gap of only 0.0008% and a maximum gap of 0.01%. We also see that the H, HM, and HML policies dominate other simple heuristics, consistent with our Proposition 3.3. In particular, we find that the HML policy, which is the even-split policy in this $N = 3$ case, performs reasonably well with a mean gap of 0.14% and a maximum gap of 0.71%. When timing information is unobservable, the SP-S policy significantly outperforms other heuristics with a mean gap of only 0.08% and a maximum gap of 1.30%. In addition, Figure 3.6 emphasizes again the pivotal impact of timing information on the inventory allocation decisions for a merchandise test. The optimal allocation policies can result in a significant loss in profit if employed in a wrong situation. The optimality gap could be as large as 6.38% if the optimal policy

Figure 3.6: Optimality gaps under heuristic test inventory allocation policies ($N = 3$).



(a) Timing Information Observable (b) Timing Information Unobservable

Note. For each policy, the box shows the median and the first and the third quartiles; all the instances with an optimality gap below the first quartile or above the third quartile are plotted as an circle outside the box.

without timing information were used when timing information is observable, and could be as large as 22.06% if the optimal policy with timing information were used when timing information is unobservable.

3.7 Concluding Remarks

This chapter uncovers new insights on the value of inventory for demand learning, in particular, on how demand censoring, demand learning across multiple locations, and the level of visibility into demand processes collectively impact inventory allocation decisions in merchandise testing.

There are fundamentally two ways to improve demand estimation given a fixed time frame to collect demand information: increasing the number, or the quantity, of the demand observations, and improving the quality of each observation. In the case where the retailer has a relatively coarse visibility into demand, i.e, the demand data contains only the sales quantities and stockout statuses, a single demand observation is made at each location if there is inventory for sale, and the quality of the observation is negatively associated with the probability of obtaining an imperfect demand observation due to stockout. One can increase the observation quantity by stocking inventory at more locations, and can improve the observation quality at a location by raising the inventory level thereat. Our results suggest that improving the quality of each demand observation is a higher priority than seeking a large observation quantity, especially when the total inventory is tightly constrainted. On the contrary, if sales timing data is observable for a reconstruction of the entire demand process, each sale transaction can be viewed as an exact demand observation. As a result, the value of inventory in demand learning is approximately maximized by simply selling as much inventory as possible so as to maximize the observation quantity. This generally involves spreading inventory among more locations.

These findings have two important managerial implications for retailers that have access to increasingly larger and richer demand data sets. First, when collecting and combining data from multiple locations for demand estimation, inventory allocation may have a significant impact on the outcome of demand learning. With the same amount of inventory, an inefficient allocation can lead to a significant loss in demand information and profit. Second, how to allocate inventory to maximizes its value in demand learning depends on the level of visibility into the demand. In particular, the use of sales timing information considerably reduces the need to maintain a high service level for demand learning as seen in reported practice.

Our work suggests several avenues for future research. In our model, demand at stores shares a common unknown parameter. Natural extensions would be to consider a hierarchical parameter structure under which each store has an unique unknown parameter in addition to the common parameter shared across stores. For example, Fisher and Rajaram (2000) cluster stores using sales histories. One could view our heuristics as solutions to the problem of inventory allocation within a store cluster; the question remains how to allocate a fixed quantity of inventory across clusters, which can be modeled by the hierarchical parameter structure described above. Another direction would be to consider a multi-product setting in which the retailer learns customer preferences in addition to demand volume through assortment experimentation. The retailer might choose to offer a full assortment with a low service level at each store, or to offer partial assortments at distinct stores with high service levels. It would be interesting to examine what would be the best inventory strategy in this setting.

CHAPTER 4

Optimal Personalized Offering when Customer Reviews Influence Demand

4.1 Introduction

There has been considerable growth in online retailing in recent years. The 2014 annual report of the National Retail Federation estimated a 9-12% growth in online sales, as opposed to a 3.6% growth in total retail sales. One of online retailers' strengths over their brick-and-mortar counterparts is their capability to collect consumer characteristics and preference data. Many online retailers, e.g., *Amazon.com*, require customers to log in their accounts to make purchases so as to keep track of their purchasing history. Typical flash deal websites such as *Gilt.com* even require a customer to log in in order to simply browse their sales listings. In addition to buying records, online retailers also mine real-time click-stream data and track customers' locations in hopes of obtaining a better understanding of their browsing behaviors and preferences. Other informative data include consumers' wish lists and abandoned shopping carts. On top of these, even more personal information may become available by encouraging individual customers to link their social network accounts. The variety and amount of data that online retailers collect has only been growing more rapidly with online retailers' increased interests in embracing big data and analytics.

Another edge of online retailers over traditional retailers is the extremely low cost and high flexibility in product listing and showcasing. Changing product assortments in a physical retail store can be labor-intensive and costly, and customizing product offers for individual customers is typically impossible. The story is quite different for an online retailer. Many online retailers have invested in sophisticated analytics and personalization tools, and customization has become increasingly common. By personalizing their front pages, product listings, search results, and product recommendations based on aforementioned rich consumer preference data, online retailers essentially have the opportunity to choose whether to offer a product to a customer. An even simpler way is to switch between the "in-stock" and "out-of-stock" tags (Bernstein et al., 2015), despite potential controversy over honesty and fairness.

The remote and virtual nature of online retailing also introduces significant uncertainty about product quality to consumers. Unlike shopping in a brick-and-mortar store where customers are able to gain hands-on experience by directly interacting with products, customers shopping online can receive product information via only limited channels, e.g., texts, pictures, and occasionally videos. To reduce such uncertainty, uninformed consumers may gather quality information from reviews generated by other peer consumers who have purchased and experienced the product. A third distinguishing feature of online retailing compared with brick-and-mortar retailing is the availability of these customer product reviews. The Internet has considerably facilitated the availability of consumer-generated reviews. Nowadays, most online retailers provide functionalities on their websites which allow consumers to post their reviews for a purchased product. The content of a review may range from a simple rating on a five-star scale, to as rich as a 500-word essay detailing every aspect regarding a product. The proliferation of social media sites such as Twitter and Facebook has opened up even more channels for consumers to share their likes and dislikes of a product.

Reviews generated by other consumers may substantially impact consumers' purchasing decisions. Academic research has empirically established the link between positive reviews and increased sales (Chevalier and Mayzlin, 2006). Product reviews can also be a valuable input to retailers' decision-making. Firms can use such review information to gauge consumers attitude toward their product and adapt their management decisions. For example, online reviews have been shown to significantly increase forecasting accuracy for motion picture revenues (Dellarocas et al., 2007).

Motivated by both online retailers' advantages in collecting consumer preference data and in personalizing product offerings, and their disadvantages in increased product quality uncertainties, this chapter incorporates two central elements: the consumers' ability to collectively learn the product quality through reviews generated by their peers, and the firm's ability to personalize offerings based on its knowledge about individual customers' preferences. We are interested in how an online retailer may wish to shape future demand by expanding or restricting access of current customers to an experiential product, and thereby influencing the trajectory of customer reviews.

Although a few papers have studied how to modulate social learning through operational levers, including inventory and pricing decisions, little work has been done in understanding product offering when the demand process is affected by social learning. We consider the problem of personalized offering when consumers generate and learn product quality information from public product reviews and the retailer monitors consumer reviews to learn customers' perception of the product quality. In particular, we consider a firm that sells an experiential product at an exogenous, constant price over a finite selling season. For each customer, the gross utility from consuming the product comprises two parts—an *ex ante* observable part that we refer to as customer preference and an *ex post* observable part that we refer to as product quality. The quality of the product is known to the firm but is unknown to and learned by customers. The customer base is heterogeneous and customers preferences for the product follow a random distribution. We assume the firm may be able to identify the preference of an arriving customer (by analyzing the customer's past purchasing and online behaviors) and choose whether to offer the product to that particular customer without incurring additional costs. Once offered, the customer purchases a unit if her *ex ante* expected net utility is positive.

We model consumers' review generation process by a stylized quasi-Bayesian social learning process. Consumers form a belief on the unknown quality of the product and update it as they observe reviews posted by previous buyers. By quasi-Bayesian, we mean that customers update their belief in a Bayesian fashion except that they ignore the potential selection biases and treat reviews *as if* they are randomly sampled from the entire population, instead of from those customers who have been offered and have purchased the product. We base our formulation upon empirical evidence on online product reviews (e.g., Li and Hitt (2008)). Each arriving customer bases her purchasing decision on the *ex ante* expected net utility. Once they purchase, customers generate reviews based on their *ex post* net utilities, namely, utilities received after they have purchased and experienced the product.

We formulate the firm's product offering problem as a finite-horizon dynamic program. We show that the optimal product offering policy is of a threshold type—the firm should only offer the product for sale to customers with a higher-than-threshold preference. We demonstrate that it can be optimal for the firm to forgo an arriving customer with a low preference, even when it is certain that the customer will buy the product if offered, in order to avoid a review that will negatively impact future sales. While our base model assumes no capacity or inventory constraints, we extend our analysis to the setting in which the firm has a limited inventory upfront.

In an numerical analysis to follow, we investigate the impact of the product price and consumers' mean belief and uncertainties about product quality on the optimal product offering decisions and on the potential value of personalized offerings. We find that personalized offering may significantly improve profit, especially in settings in which the product price is high and customers are optimistic but uncertain about product quality.

The remainder of this chapter is organized as follows. In §4.2 we review the relevant literature. In §4.3, we describe our model framework. §4.4 presents analysis on the social learning process and optimal inventory policies in a setting with ample supply, whereas §4.5 extends the analysis to the problem with a limited inventory. We conduct a numerical analysis in §4.6 to understand the value of personalized offering. In §4.7, we conclude the chapter with discussions and future research directions.

4.2 Literature Review

Product reviews have long received attention in the marketing literature. Most of this line of work focuses on studying the effect of product reviews on sales of experiential products. Findings are mixed on the association between product reviews and sales. There are empirical studies showing that positive reviews are associated with higher sales, while negative reviews may hurt sales of experiential goods (Chevalier and Mayzlin, 2006; Dellarocas et al., 2007). Some other papers do not find statistically significant relationships (Duan et al., 2008; Liu, 2006). This stream of literature centers on how online product reviews drive consumer choice behaviors, whereas our focus is on how firms should adaptively manage inventory in the presence of such a review-driven demand process.

Our work also connects to the social learning literature as late consumers learn quality information from early consumers through public product reviews. Our social learning process occurs when buyers report their ex post utilities, which is in line with the empirical work of Godes and Silva (2012), and the theoretical framework in Papanastasiou et al. (2014), Besbes and Scarsini (2014), and Ifrach et al. (2015). In these papers as in this chapter, consumers' purchasing decisions are non-informative of the product quality. This contrasts with the settings in Banerjee (1992) and Bikhchandani et al. (1992), in which social learning occurs when agents observe their predecessors' *ex ante* actions.

A growing stream of literature studies how firms can modulate social learning through their operational decisions. The majority of this literature focuses on pricing decisions when customer behavior is driven by social learning processes. Ifrach et al. (2012) consider monopoly pricing when buyers report whether their experience is positive or negative, and subsequent customers learn from these reports according to an intuitive non-Bayesian rule. Jing (2011), Papanastasiou and Savva (2014), and Yu et al. (2013) analyze dynamic pricing policies when forward-looking consumers may strategically delay their purchases in anticipation of product reviews. The literature is rather limited on inventory control in the presence of social learning. Papanastasiou et al. (2014) provides an explanation for early-supply shortage strategies in the presence of quasi-Bayesian social learning.

This chapter is also related to the assortment planning literature. For a thorough review, readers are referred to Kök et al. (2015). This literature generally considers a firm's optimal choice of a subset of multiple products to offer. In this regard, our problem can be seen as a special single-product assortment planning problem with only two assortment options, offering or not offering the product. Two papers on dynamic assortment customization are specifically relevant. Bernstein et al. (2015) consider a retailer selling identically-priced, substitutable products to a heterogeneous customer base. The firm is able to identify arriving customers' types and to customize the assortment seen by each arriving customer. They find it may be optimal for the retailer facing low inventory levels to reserve product for customers with stronger preferences. Golrezaei et al. (2014) propose algorithms for optimally personalizing assortment for each arriving customer with the availability of real-time consumer characteristics data. Both papers find that personalization leads to significant improvements in revenue, but none considers a demand process governed by a social learning process.

4.3 Model

In this section, we describe our model framework for the firm's optimal offering problem. Consider a firm selling an experiential product over T periods, indexed by T, $T-1$, ..., 1. We assume that the product's unit price $p > 0$ is exogenously given and constant throughout T periods. For simplicity, we assume a constant stream of customer arrivals; one customer arrives each period, and each purchases at most one unit of the product.

4.3.1 Consumers

A representative customer i's gross utility from consuming the product consists of two components, θ_i and q. The term θ_i represents a customer's idiosyncratic preference—utility derived from observable product features before purchase (e.g., product brand) and is known to the consumer *ex ante*. Customers are heterogenous in their preferences. We assume that θ_i's in the population follow a Normal distribution with $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$. The density function is denoted by $f_\Theta(\cdot)$, the distribution function by $F_\Theta(\cdot)$ and the complementary distribution function by $\overline{F}_\Theta(\cdot) := 1 - F_\Theta(\cdot)$. The distribution $F_\Theta(\cdot)$ is assumed to be known to both the firm and customers.

The term q represents the product's quality—utility derived from attributes which are unobservable before purchase (e.g., product usability) and is referred to generically as the product's quality for customer i; q is known by the firm but is *ex ante* unknown and is learned by the consumer only after purchasing and experiencing the product. For ease of exposition, we assume away randomness in consumers' *ex post* quality perceptions; namely, all customers perceive exactly q as the product's underlying quality after consuming the product.

Consumers hold a common prior belief over q, which summarizes their initial perception on the product quality. We assume this belief to be a Normal random variable $\hat{Q}_T$ where $\hat{Q}_T \sim \mathcal{N}(\hat{q}_T, \sigma_T^2)$. Customers update their belief over q as customer reviews accumulate. We will elaborate on this belief updating process in the sections to follow.

We assume customers to be risk-neutral utility-maximizers. Customer i with a non-negative *ex ante* expected utility, i.e., $\hat{q} + \theta_i - p \geq 0$, is willing to buy a unit of the product. Once she purchases, customer i derives a net utility of $\theta_i + q - p$ from purchasing and experiencing the product. We do not consider in our model strategic customers who delay purchasing to wait for more information about the product quality; each arriving customer leaves permanently regardless of her purchasing decision.

Review Generation. Each buyer i generates a review of the product that is viewable by the firm and all customers yet to arrive. Empirical evidence has shown that product reviews may be systematically biased by customers' idiosyncratic preferences. To capture this effect, we assume that a buyer i simply (and truthfully) reports her net utility $q + \theta_i - p$, or equivalently, generates a review $r_i = q + \theta_i - \mu_\theta$, as both p and μ_θ are assumed to be common knowledge. We note that reviews are sampled only from customers who are offered the product and who purchase, not from the entire population. As a result, reviews do not necessarily follow a $\mathcal{N}(q, \sigma_\theta^2)$ distribution.

Social Learning of q. Consumers collectively learn the underlying product quality q by monitoring reviews generated by previous buyers. A rational customer would update her belief using Bayes rule. However, a full Bayesian updating would require customers to keep track of the entire review history, to process it, and to anticipate the firm's offering policy, which is numerically challenging even for computers, and thus arguably impractical for customers to perform, considering the large amount of cognitive processing power it demands (see Appendix C.7 for a derivation of Bayesian updating that accounts for the selection bias induced by customer purchasing). Indeed, there is empirical support that online product reviews are subject to self-selection biases (Li and Hitt, 2008).

For the above reasons, in our model we take a quasi-Bayesian approach to model customers' belief updating process. More specifically, we assume that customers ignore the selection bias and update their belief *as if* reviews are i.i.d. sampled from the entire population, i.e., reviews follow a $\mathcal{N}(q, \sigma_\theta^2)$ distribution. Under this assumption, the belief updating could be carried out according to the usual parametric Bayesian paradigm. Let $\hat{Q}_t$ denote consumers' common belief at the beginning of period t. Following Bayes rule, $\hat{Q}_t$ is Normal; in particular, $\hat{Q}_t \sim \mathcal{N}(\hat{q}_t, \sigma_t^2)$, with

$$\hat{q}_t = \frac{\sigma_\theta^2 \hat{q}_T + \sigma_T^2 R_t}{\sigma_\theta^2 + s_t \sigma_T^2} \quad \text{and} \quad \sigma_t^2 = \frac{\sigma_\theta^2 \sigma_T^2}{\sigma_\theta^2 + s_t \sigma_T^2},$$

where R_t is the sum of reviews generated prior to period t and s_t the cumulative sales.

It is convenient in our problem setting to use mean belief $\hat{q}_t$ and cumulative sales s_t as sufficient statistics for belief updating. Suppose that a customer with preference θ_i buys a unit of the product in period t and generates a review $r_i = q + \theta_i - \mu_\theta$. Then, the updated $\hat{q}_{t-1}$ and s_{t-1} for period $t-1$ are given by

$$\begin{aligned} \hat{q}_{t-1} &= u(\hat{q}_t, s_t, \theta) := \frac{\hat{q}_t(\sigma_\theta^2 + s_t \sigma_T^2) + (q + \theta - \mu_\theta)\sigma_T^2}{\sigma_\theta^2 + (s_t + 1)\sigma_T^2}, \\ s_{t-1} &= s_t + 1. \end{aligned} \tag{4.1}$$

We define the mean belief updating function $u(\hat{q}, s, \theta)$ as a function of θ from the firm's perspective, as θ is essentially revealed to the firm after a customer posts a review $r = q + \theta$, since we assume that the firm knows the value of q. This parameterization also proves useful in our analysis of the problem to come. From the customers' perspective, the updating is also valid by treating $q + \theta - \mu_\theta$ as a single term. In addition, we remark that customers' belief remains unchanged, i.e.,

$$\hat{q}_{t-1} = \hat{q}_t \quad \text{and} \quad s_{t-1} = s_t.$$

if a customer does not make a purchases. In other words, a non-purchase is non-informative for customers to learn the true value of q. Intuitively, this is because when a customer makes a purchasing decision, she has no additional knowledge of q beyond the common belief.

We summarize in Lemma 4.1 some basic properties of the mean belief updating function $u(\hat{q}, s, \theta)$.

Lemma 4.1. *The following hold for all $s = 0, 1 \ldots,$:*

(a) $u(\hat{q}, s, \theta)$ is increasing in $\hat{q}$;

(b) $u(\hat{q}, s, \theta)$ is increasing in θ;

(c) $u(\hat{q}, s, \mu_\theta + \hat{q} - q) = \hat{q}$.

All the proofs can be found in Appendix C. Parts (a) and (b) of Lemma 4.1 show that consumers' posterior mean belief increases with their prior mean belief on q as well as the reviewer's preference. Part (c) provides a threshold value on consumers' preferences. An immediate corollary of part (b) and (c) is that a review generated by a customer with preference higher than $\mu_\theta + \hat{q} - q$ will raise consumers' posterior mean belief; otherwise, customers' mean belief decreases.

4.3.2 The Firm

A unique feature of our model is that the firm is allowed to personalize whether to offer the product to each individual customer. As is discussed in the introduction, this is rarely a viable option for traditional brick-and-mortar retailers but is increasingly adopted by online retailers through customized product listings, search results, and other methods. The firm seeks to maximize its expected total profit over the entire selling season by choosing whether to offer the product for sale to each arriving customer. As mentioned previously, we assume the firm knows the (exogenously determined) true value of q. We denote by $o_t \in \{0, 1\}$ the firm's offering decision for period t, where $o_t = 1$ indicates that the firm chooses to offer the product for sale to the arriving customer. In our base model, we assume that the firm has ample supply of the product. We will discuss in §4.5 the case in which the firm has only a limited inventory to sell.

An implicit assumption we make is that customers also ignore the potential bias in reviews induced by the firm's offering policy. That is, customers treat reviews as if they are randomly sampled from the entire population, instead of customers who are willing to purchase and are offered the product by the firm.

4.4 Analysis

4.4.1 Unobservable Customer Preference Information – A Benchmark

As the first step of our analysis, we consider a benchmark case in which individual customers' preferences are not revealed to the firm at their time of arrival. In other words, the firm only knows that an arriving customer's preference θ_i is a random draw from distribution $F_\Theta(\cdot)$. Practically, this corresponds to traditional situations in which the firm may lack or otherwise not use customers' personal information.

We formulate the firm's optimal offering problem as a finite-horizon dynamic program (DP). Let $V_t(\hat{q}, s)$ be the optimal expected profit with t periods left, consumers' mean belief $\hat{q}$ on q, and cumulative sales s. Then, the Bellman equations are given by

$$V_t(\hat{q}, s) = \max_{o_t \in \{0,1\}} o_t \cdot \left\{ F_\Theta(p - \hat{q}) V_{t-1}(\hat{q}, s) + \int_{p-\hat{q}}^{\infty} [p + V_{t-1}(u(\hat{q}, s, \theta), s+1)] \mathrm{d}F_\Theta(\theta) \right\}$$
$$+ (1 - o_t) \cdot V_{t-1}(\hat{q}, s),$$

with terminal value functions $V_0(\hat{q}, s) = 0$.

Denote by $o_t^*(\hat{q}, s)$ the firm's optimal offering decision for period t given consumer mean belief $\hat{q}$ and cumulative sales s. We show in Lemma 4.2 that it is optimal for the firm to offer the product to every arriving customer. This is not surprising given that the firm is not able to differentiate one customer from another when making offering decisions. We call this an "offer-to-all" policy and will use it as a benchmark in our numerical analysis in §4.6 when the firm has access to individual customers' preferences.

For the case in which the firm does not observe individual customer preferences,

Lemma 4.2. $o_t^*(\hat{q}, s) = 1$ *for all* t, $\hat{q}$, *and* s.

4.4.2 Observable Customer Preference Information

In this subsection, we consider the case in which the firm has knowledge of each arriving customer's preference. Specifically, we assume that the firm knows the exact value of each arriving customer's idiosyncratic preference θ_i. The implication of this assumption, given the review-generating process in our model, is that the firm knows exactly whether an arriving customer i with common mean belief $\hat{q}$ will make a purchase ($\hat{q} + \theta_i - p \geq 0$), and the review she would generate ($r_i = q + \theta_i - \mu_\theta$) if she buys. Although highly stylized, this assumption simplifies our exposition while capturing a key feature in our problem setting that the firm possesses information for identifying individual customers' preferences. In addition, our main results can be generalized to the more general case in which the firm observes θ plus a zero-mean Gaussian random noise. As the insights generated are qualitatively similar, we do not present this general case for expositional simplicity.

Dynamic Program. Similar to the benchmark problem in §4.4.1, we formulate the firm's optimal product offering problem as a finite-horizon DP. The firm's objective is to maximize the expected total profit over T periods. With a slight abuse of notation, let $V_t(\hat{q}, s, \theta)$ be the optimal expected total profit with t periods left, consumers' mean belief $\hat{q}$, cumulative sales s, and the arriving customer's preference θ. If the arriving customer with preference θ has a negative expected net utility, i.e., $\theta + \hat{q} - p < 0$, she does not buy the product. The firm collects no revenue and customers do not update their common belief on q. The next arriving customer's preference will be a random variable Θ following the preference distribution $F_\Theta(\cdot)$. If the arriving customer's expected net utility is positive, she will make a purchase as long as the firm offers the product for sale. In that case, the firm collects revenue p, and consumers update their common belief according to (4.1), in response to the review generated by the buying customer. In summary, the value functions $V_t(\hat{q}, s, \theta)$ are given by

$$V_t(\hat{q}, s, \theta) = \begin{cases} \mathsf{E}_\Theta V_{t-1}(\hat{q}, s, \Theta), & \theta + \hat{q} - p < 0, \\ \max\{p + \mathsf{E}_\Theta V_{t-1}(u(\hat{q}, s, \theta), s+1, \Theta), \mathsf{E}_\Theta V_{t-1}(\hat{q}, s, \Theta)\}, & \theta + \hat{q} - p \geq 0. \end{cases} \tag{4.2}$$

for $t = T, T-1, \ldots, 1$, with $V_0(\hat{q}, s, \theta) = 0$ for all $\hat{q}$, s, and θ.

Optimal Offering Policy. Let $o_t^*(\hat{q}, s, \theta) \in \{0, 1\}$ be the optimal offering decision with t periods left, customers' mean belief $\hat{q}$, and arriving customer's preference θ, where $o_t^*(\hat{q}, s, \theta) = 1$ indicates that retailer chooses to offer the product and $o_t^*(\hat{q}, s, \theta) = 1$ otherwise. We use the convention that $o_t^*(\hat{q}, s, \theta) \equiv 0$ whenever $\theta + \hat{q} - p < 0$. In the following proposition, we show monotonicity properties of the value functions, which we subsequently use to characterize the form of the optimal offering policy.

Proposition 4.1. *The following hold:*

(a) $V_t(\hat{q}_1, s, \theta) \leq V_t(\hat{q}_2, s, \theta)$ *for any* $\hat{q}_1 < \hat{q}_2$, $t = 0, \ldots, T$, $s = 0, \ldots, T - t$, *and all* θ.

(b) $V_t(\hat{q}, s, \theta_1) \leq V_t(\hat{q}, s, \theta_2)$ *for any* $\theta_1 < \theta_2$, $t = 0, \ldots, T$, $s = 0, \ldots, T - t$, *and* $\hat{q} \in \mathbb{R}$.

(c) There exists a series of functions, $\{\theta_t^*(\hat{q}, s), t = 1, \ldots, T\}$, *on* $\mathbb{R}$ *such that* $o_t^*(\hat{q}, s, \theta) = 1$ *if* $\theta \geq \theta_t^*(\hat{q}, s)$, *and* $o_t^*(\hat{q}, s, \theta) = 0$ *otherwise.*

Proposition 4.1(a) and (b) show that the optimal expected total profit increases with consumers' mean belief and the arriving customer's preference. These monotone properties are consistent with the intuition that higher mean belief/preference leads to not only a higher revenue from the arriving customers but also a higher mean belief for future customers.

Proposition 4.1(c) characterizes the firm's optimal offering policy to be of a threshold type. The threshold $\theta_t^*(\hat{q}, s)$ separates the potential arriving customers in period t into two segments; to maximize profit, the firm should only offer the product for sale to customers who have a preference higher than $\theta_t^*(\hat{q}, s)$.

We remark that the optimal offering thresholds may be non-trivial. Recall that a customer with preference θ and common mean belief $\hat{q}$ will buy if $\hat{q} + \theta - p \geq 0$, or equivalently, if $\theta \geq p - \hat{q}$. Therefore, a constant threshold $\theta_t^*(\hat{q}, s) \equiv p - \hat{q}$ would simply imply an offer-to-all policy. However, as will be evident in our numerical illustrations, $\theta_t^*(\hat{q}, s)$ can be greater than $p - \hat{q}$. Such a threshold value would mean that it can be optimal for the firm to "conceal" the product from an arriving customer, even if the firm is aware that the particular customer *will* purchase the product *if offered*. The underlying reason for the firm to forgo the immediate, guaranteed revenue is to avoid the negative externality of offering the product to a low-preference customer. Accompanied with a revenue of p, a low-preference customer also generates a low product review, which may decrease consumers' mean belief and thus decrease the expected revenue from future customers.

Moreoever, with Proposition 4.1(c), we are able to simplify the firm's problem by dropping θ as a state variable. Define $G_t(\hat{q}, s) := \mathsf{E}_\Theta V_t(\hat{q}, s, \Theta)$, where $G_t(\hat{q}, s)$ is the firm's expected optimal profit in anticipation of a customer arriving in period t, with consumers' mean belief being $\hat{q}$ and the cumulative sale s. We can rewrite the DP in (4.2) as

$$\begin{aligned} G_t(\hat{q}, s) &= \max_{\theta_t \geq p - \hat{q}} F_\Theta(\theta_t) G_{t-1}(\hat{q}, s) + \int_{\theta_t}^{\infty} [p + G_{t-1}(u(\hat{q}, s, \theta), s+1)] \mathrm{d}F_\Theta(\theta) \\ &= \max_{\theta_t \geq p - \hat{q}} F_\Theta(\theta_t) G_{t-1}(\hat{q}, s) + p\overline{F}_\Theta(\theta_t) + \int_{\theta_t}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1) \mathrm{d}F_\Theta(\theta) \\ &= F_\Theta(\theta_t^*(\hat{q}, s)) G_{t-1}(\hat{q}, s) + p\overline{F}_\Theta(\theta_t^*(\hat{q}, s)) + \int_{\theta_t^*(\hat{q}, s)}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1) \mathrm{d}F_\Theta(\theta) \end{aligned}$$

with terminal value functions $G_0(\hat{q}, s) = 0$ for all $\hat{q}$ and s.

With the above simplified notation, we further characterize the optimal offering threshold $\theta_t^*(\hat{q}, s)$ in Proposition 4.2.

Proposition 4.2. *Let $\theta_t^0(\hat{q}, s)$ be the unique solution to the equation*

$$G_{t-1}(\hat{q}, s) - G_{t-1}(u(\hat{q}, s, \theta), s+1) = p, \tag{4.3}$$

if it exists. Then, $\theta_t^*(\hat{q}, s) = \max\{p - \hat{q}, \theta_t^0(\hat{q}, s)\}$ if $\theta_t^0(\hat{q}, s)$ exists. Otherwise, $\theta_t^*(\hat{q}, s) = p - \hat{q}$.

Proposition 4.2 reveals the key tradeoff the firm faces when making its offering decision for each arriving customer. The solution to Equation (4.3), $\theta_t^0(\hat{q}, s)$, if it exists, strikes a balance between the firm's immediate return, the selling price p, and the expected marginal return from skipping a low-value customer and maintaining a high mean belief for future customers. We note that $\theta_t^0(\hat{q}, s)$ may not exist at all: a straightforward example is when $t = 1$, as $G_0(\hat{q}, s) = 0$ for all $\hat{q}$ and s. Indeed, the firm should offer the product to the last customer whenever she is willing to buy. When $\theta_t^0(\hat{q}, s)$ does not exist, it means that the future benefit of not offering the product to a potential buying customer never outweighs the immediate revenue; as a result, the optimal decision is to offer the product to all customers. Similarly, if $\theta_t^0(\hat{q}, s)$ does exists but is below the buying threshold $p - \hat{q}$, there are customers with low enough preferences to justify a no-offering decision as a result of their negative externality on future sales through social learning; however, such customers will not purchase due to their low preferences.

The following proposition indicates that the firm's optimal offering region shrinks as t increases.

Proposition 4.3. *$\theta_t^*(\hat{q}, s) \leq \theta_{t+1}^*(\hat{q}, s)$ for $t = 0, \ldots, T-1$, $\hat{q} \in \mathbb{R}$ and $s = 0, \ldots, T-t$.*

The intuition behind Proposition 4.3 is as follows. As the selling horizon lengthens, so does the marginal benefit of personalized offering, since the number of affected future customers increases. To illustrate, with $t+1$ periods left, if the firm chooses the optimal threshold for a t-period problem, $\theta_t^*(\hat{q}, s)$, the firm's marginal immediate profit loss remains unchanged as in a t-period problem, whereas the marginal future profit gain increases due to the additional period at the end to sell the product. Therefore, by (weakly) increasing the offering threshold from $\theta_t^*(\hat{q}, s)$ and thus offering to a smaller segment of customers in the current period, the firm re-strikes a balance between an increased its marginal profit loss and the marginal future profit gain.

4.5 Limited Inventory

In this section, we extend our analysis to the case in which the firm has a fixed quantity of inventory, $I \in \mathbb{Z}_+$, at the beginning of the selling season. The firm has no additional replenishment opportunity throughout the T-period horizon. We note that this problem reduces to an unlimited inventory problem when $I \geq T$, and our results in §4.4.2 immediately follow.

We incorporate the inventory constraint into the DP in Equation (4.2) by including inventory level x as an additional state variable. In particular, let $V_t(\hat{q}, s, x, \theta)$ be the optimal expected total profit with t periods left, consumers's mean belief $\hat{q}$, cumulative sales s, inventory level x, and the arriving customer's preference θ. The Bellman equations are given by

$$V_t(\hat{q}, s, x, \theta) = \begin{cases} \mathsf{E}_\Theta V_{t-1}(\hat{q}, s, x, \Theta), & \theta + \hat{q} - p < 0 \text{ or } x = 0, \\ \max\{p + \mathsf{E}_\Theta V_{t-1}(u(\hat{q}, s, \theta), s+1, x-1, \Theta), & \\ \qquad \mathsf{E}_\Theta V_{t-1}(\hat{q}, s, x, \Theta)\}, & \theta + \hat{q} - p \geq 0 \text{ and } x > 0. \end{cases} \tag{4.4}$$

for $t = T, T-1, \ldots, 1$, with $V_0(\hat{q}, s, x, \theta) = 0$ for all $\hat{q}$, s, x, and θ.

In a similar fashion to Proposition 4.1, we derive monotone properties of function $V_t(\hat{q}, s, x, \theta)$ and subsequently the form of the optimal offering policy.

Proposition 4.4. *The following hold:*

(a) $V_t(\hat{q}_1, s, x, \theta) \leq V_t(\hat{q}_2, s, x, \theta)$ *for any* $\hat{q}_1 < \hat{q}_2$, $t = 0, \ldots, T$, $s = 0, \ldots, T-t$, *and all* θ.

(b) $V_t(\hat{q}, s, x, \theta) \leq V_t(\hat{q}, s, x+1, \theta)$ *for* $x = 0, \ldots,$ *all* $\hat{q}$, s, *and* θ*;*

(c) $V_t(\hat{q}, s, x, \theta_1) \leq V_t(\hat{q}, s, x, \theta_2)$ *for any* $\theta_1 < \theta_2$, $t = 0, \ldots, T$, $s = 0, \ldots, T-t$, *and* $\hat{q} \in \mathbb{R}$.

(d) There exists a series of functions, $\{\theta_t^*(\hat{q}, s, x), t = 1, \ldots, T\}$, *such that* $o_t^*(\hat{q}, s, x, \theta) = 1$ *if* $\theta \geq \theta_t^*(\hat{q}, s, x)$, *and that* $o_t^*(\hat{q}, s, x, \theta) = 0$ *otherwise.*

While Proposition 4.4(a), (c) and (d) do not change qualitatively from their counterparts in Proposition 4.1, Proposition 4.4(b) suggests that the optimal expected profit is also monotonically increasing in the inventory level x.

Again, we see that the optimal offering policy is of the threshold type in the presence of an inventory constraint. The optimal offering thresholds are functions that depend on inventory level x, in addition to t, $\hat{q}$, and s.

Similarly, we are able to simplify the firm's problem by dropping θ as a state variable. Define $G_t(\hat{q}, s, x) := \mathsf{E}_\Theta V_t(\hat{q}, s, x, \Theta)$, where $G_t(\hat{q}, s, x)$ is the firm's expected optimal profit in anticipation of a customer arriving in period t, with consumers' mean belief $\hat{q}$, cumulative sales s, and inventory level x. We rewrite the DP in (4.4) as

$$G_t(\hat{q}, s, x) = \max_{\theta_t \geq p - \hat{q}} F_\Theta(\theta_t) G_{t-1}(\hat{q}, s, x) + \int_{\theta_t}^{\infty} [p + G_{t-1}(u(\hat{q}, s, \theta), s+1, x-1)] \mathrm{d}F_\Theta(\theta)$$

$$
\begin{aligned}
&= \max_{\theta_t \geq p-\hat{q}} F_\Theta(\theta_t)G_{t-1}(\hat{q}, s, x) + p\overline{F}_\Theta(\theta_t) + \int_{\theta_t}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1, x-1)\mathrm{d}F_\Theta(\theta) \\
&= F_\Theta(\theta_t^*(\hat{q}, s, x))G_{t-1}(\hat{q}, s, x) + p\overline{F}_\Theta(\theta_t^*(\hat{q}, s, x)) \\
&\quad + \int_{\theta_t^*(\hat{q},s,x)}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1, x-1)\mathrm{d}F_\Theta(\theta)
\end{aligned}
$$

for $x > 0$, with terminal value functions $G_0(\hat{q}, s, x) = 0$ for all $\hat{q}$ and s, and $G_t(\hat{q}, s, 0) = 0$ for all t, $\hat{q}$, and s by noticing that the expected profit is zero with no inventory to sell.

With the above simplified notation, we characterize the optimal offering threshold $\theta_t^*(\hat{q}, s, x)$ in Proposition 4.5, which extends Proposition 4.2 to a limited inventory setting.

Proposition 4.5. *Let $\theta_t^0(\hat{q}, s, x)$ be the unique solution to the equation*

$$
G_{t-1}(\hat{q}, s, x) - G_{t-1}(u(\hat{q}, s, \theta), s+1, x-1) = p, \tag{4.5}
$$

if it exists. Then, $\theta_t^*(\hat{q}, s, x) = \max\{p - \hat{q}, \theta_t^0(\hat{q}, s, x)\}$ if $\theta_t^0(\hat{q}, s, x)$ exists. Otherwise, $\theta_t^*(\hat{q}, s, x) = p - \hat{q}$.

In the following, we analyze the optimal offering threshold $\theta_t^*(\hat{q}, s, x)$ as a function of the selling horizon t as well as the inventory level x. We show in Proposition 4.6 that the optimal offering region shrinks as t or x increases.

Proposition 4.6. *The following hold:*

(a) $\theta_t^*(\hat{q}, s, x) \leq \theta_{t+1}^*(\hat{q}, s, x)$ for $x = 0, 1, \ldots,$ all $\hat{q}$ and s;

(b) $\theta_t^*(\hat{q}, s, x) \leq \theta_t^*(\hat{q}, s, x+1)$ for $x = 0, 1, \ldots,$ all t and s;

Proposition 4.6(a) shows that with more time left to sell the product, the firm should be more selective in product offering; that is, the firm should offer the product to a smaller segment of high-preference customers with the same consumer belief and leftover inventory. This confirms that our finding in Proposition 4.3 continues to hold with a limited inventory.

Proposition 4.6(b) suggests that the firm should be more selective in product offering with more leftover inventory, holding the consumer belief and selling horizon constant. This conforms with our intuition that more inventory corresponds to more selling opportunities, hence a stronger incentive to personalize offerings. We remark that this result is contrary to findings in the classic revenue management literature that the firm should be less selective (i.e., reduce price in a dynamic pricing context) with more leftover inventory (Elmaghraby and Keskinocak, 2003).

We also note that Proposition 4.6 links the optimal threshold for the limited inventory problem to that for the unlimited inventory problem by providing a lower and an upper bound for the former with the latter. Particularly, it follows straightforwardly that $\theta_x^*(\hat{q}, s) = \theta_x^*(\hat{q}, s, x) \le \theta_t^*(\hat{q}, s, x) \le \theta_t^*(\hat{q}, s, t) = \theta_t^*(\hat{q}, s)$, where $\theta_x^*(\hat{q}, s)$ and $\theta_t^*(\hat{q}, s)$ are the optimal offering thresholds for an x- and t-period unlimited inventory problem, respectively (recall that $x \le t$). That is, for a t-period problem with x units of inventory, solving a t-period problem with unlimited inventory gives an upper bound on the optimal offering threshold, whereas solving an x-period problem with unlimited inventory yields a lower bound.

4.6 Numerical Analysis

In this section, we conduct a numerical analysis in order to obtain a deeper understanding of the optimal product offering problem with customer preference and review information. We first describe in §4.6.1 how we parameterize the problem and vary the model parameters. Then in §4.6.2, we examine the value of personalized offering and demonstrate how the optimal offering policy behaves under different parameter settings.

4.6.1 Parameter Settings

We simplify our model parameterization by focusing on three derived parameters—initial belief mean $\hat{q}_T$, *nominal purchasing probability* $\beta := \mathsf{P}(\Theta + q - p \ge 0) = \overline{F}_\Theta(p - q)$, and *uncertainty ratio* $\gamma := \sigma_T / \sigma_\theta$. We normalize $q = 0$ throughout our numerical analysis without loss of generality. As a result, $\hat{q}_T$ represents consumers' relative expectation on product quality: a positive (negative) $\hat{q}_T$ value indicates that customers have an expectation on quality that is higher (lower) than the actual quality of the product.

We define the nominal purchasing probability, β, to be the probability that a random customer is willing to buy the product if the consumer mean belief $\hat{q} = q = 0$. One can interpret β as a normalized proxy for price p. We let β take values in $\{0.9,\ 0.5,\ 0.1\}$, corresponding to a relatively low, medium, and high price, respectively.

The uncertainty ratio γ reflects consumers' level of uncertainty about the unknown product quality. The higher the value of γ, the more susceptible is the consumers' mean belief to be affected by reviews. We choose the value of γ from $\{0.5,\ 1,\ 2\}$. In addition, we fix $\mu_\theta = 3$ and $\sigma_\theta^2 = 1$ to allow for low nominal purchasing probabilities β's with price still being positive.

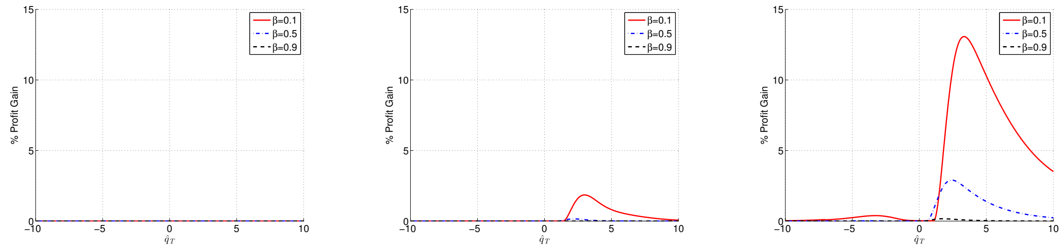
We perform the analysis for the unlimited inventory problem with $T = 10, 20, 50$ and the limited inventory problem with $T = 50$ and $I = 10, 20, 50$. We do not observe qualitative difference among these parameter settings. Therefore, in the following we only report our results for the unlimited inventory problem with $T = 10$.

4.6.2 The Value of Personalized Offering

We have discussed in Section 4.4.1 the "offer-to-all" policy, a benchmark product offering policy that simply offers the product to all arriving customers. We have also demonstrated its optimality in settings in which the firm does not observe individual customers' preferences. In this subsection, we investigate the value of personalized offering with customer review and preference information by comparing the performance of the optimal offering policy with that of the "offer-to-all" benchmark. Recall that $G_T(\hat{q}, 0)$ is the optimal expected total profit of a T-period problem with consumers' prior mean belief $\hat{q}$. Let $G_T^o(\hat{q}, 0)$ denote the expected total profit of the same problem under the offer-to-all benchmark policy. To evaluate the value of personalized offering over the traditional practice of offering to the entire customer base, we compute the percentage profit gain by using personalized offering across a set of numerical examples. In particular, the metric that we use is given by $\left(1 - \frac{G_T^o(\hat{q},0)}{G_T(\hat{q},0)}\right) \times 100\%$.

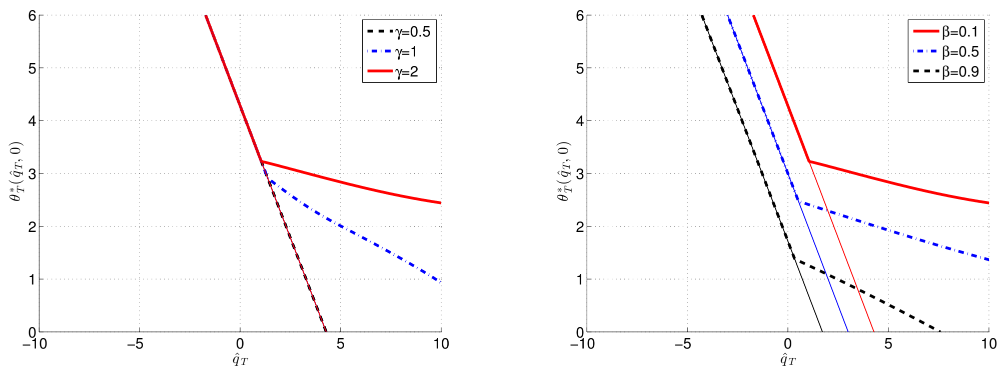
We plot in Figure 4.1 the percentage profit gain from personalized offering as a function of $\hat{q}$ under various β and γ values. We also show in Figure 4.2 the optimal offering threshold for the first arriving customer in a T-period problem, $\theta_T^*(\hat{q}_T, 0)$, as a function of consumers initial mean belief $\hat{q}_T$. The thin lines represent customers purchasing thresholds $\theta = p - \hat{q}_T$, which correspond to offering threshold for the offer-to-all policies. We observe that in general

Figure 4.1: Percentage profit gain from personalized offering.



(a) $\gamma = 0.5$ (b) $\gamma = 1$ (c) $\gamma = 2$

Figure 4.2: Optimal offering threshold $\theta_T^*(\hat{q}, 0)$.



(a) $\beta = 0.1$ (b) $\gamma = 2$

the firm optimally chooses not to offer the product when consumers' mean belief is high but the arriving customer's preference is low.

The Impact of $\hat{q}$. We observe that personalized offering seems to be most valuable for moderately positive $\hat{q}$ values, that is, when customers hold an initial expectation that is slightly higher than the product's true quality. Figure 4.1 shows that the potential profit gain from personalized offering can be more than 10% over an offer-to-all policy.

The value of personalized offering diminishes as $\hat{q}_T$ becomes either negative or very large. When $\hat{q}_T$ is low, customers have a low expectation on product quality. As a result, only customers with a high preference will purchase, and the generated reviews will only boost up consumers' mean belief on quality. In this case, the firm tends to adopt an offer-to-all policy as the immediate revenue and future expected sales are aligned rather than conflicted. This is reflected in Figure 4.2(b). When $\hat{q}_T$ is high, customers' expectation on product quality becomes so high that even offering the product to a low-preference customer will not bring down customers' expectation enough to negate future purchases. For example, Figure 4.2(b) shows that as $\hat{q}_T$ increases, the optimal offering threshold falls below $\mu_\theta - 3\sigma_\theta = 0$ for $\beta = 0.9$ and $\gamma = 2$; the firm is offering the product to essentially all customers.

The Impact of β. The value of personalized offering seems to decrease with β, or equivalently, to increase with price. As shown in Figure 4.1, its value almost vanishes at $\beta = 0.9$, which translates into a low price tag with a 90% purchasing probability when customers know the true quality q. The value is most significant at $\beta = 0.1$, which corresponds to a high price with only a 10% purchasing probability.

The intuition is as follows. When the price is low, most customers will buy regardless of a low expectation on product quality. Therefore, the benefit of personalized offering is small. On the contrary, when the price is high, an increase in customers' quality expectation will significantly increase the purchasing probability of an arriving customer. Hence, the value of personalized offering increases. We observe in Figure 4.2(b) that the firm tends to be more selective in its offerings as price increases, or, as β decreases.

The Impact of γ. Figure 4.1 shows that personalized offering is most valuable when γ is large; that is, when consumers are very uncertain about their belief on the product quality. A large γ implies that customers' mean belief is easily changed in response to reviews. Therefore, personalized offering makes a big difference. However, when γ is small, reviews hardly affect customers' belief about quality. As a result, the benefit of withholding product from customers goes down.

Figure 4.2(a) displays how the value of γ affects the optimal offering threshold with a fixed $\beta = 0.1$. For a fixed price, the optimal offering region shrinks as γ increases. Note that the optimal offering policy approaches the offer-to-all benchmark policy as γ decreases from 2 to 0.5.

To summarize, our numerical analysis shows that personalized offering may be most prominent and also may provide its highest benefit when the price is high and when consumers have a moderately optimistic but highly uncertain expectation on the underlying quality of the product.

4.7 Concluding Remarks

In this chapter, we have considered a firm that maximizes its profit by personalizing product offering decisions for a sequence of arriving customers based on their observable and heterogeneous preferences. Customers collectively learn product quality by updating their common belief based on product reviews generated by previous buyers. We characterize that the optimal offering policy is of threshold type. Our most important finding is that the optimal offering threshold can be higher than consumers' purchasing threshold; that is, it can be optimal for the firm not to offer the product for sale, even knowing the arriving customer is willing to buy. The decision to offer the product involves a key tradeoff between an immediate profit loss versus the marginal expected future sales from maintaining a high mean belief. We remark that both features in our model—that the firm can observe individual customer preferences and that customers learn product quality through reviews—are essential for our main results to hold. Removing either will immediately trivialize the problem, making an offer-to-all policy optimal.

Our work suggests several directions for future research. It would be interesting to endogenize the initial inventory level I and/or the selling price p as decision variables and to analyze the joint optimal pricing, inventory ordering, and product offering policy. Another natural avenue to pursue is to extend the problem to a multi-product setting. In this case, the problem closely resembles classic dynamic assortment planning problems, but with a demand process that can be affected by customer reviews. It would be particularly valuable to explore whether and how the firm would benefit from personalizing assortment for individual customers in the presence of substitution and complementarity among products.

Our results rely on the assumption that customers are not fully rational in updating their belief over product quality. In interpreting previous customer reviews, they do not adjust for the selection bias due to customer purchasing decisions nor for the selection bias due to the firm's product offering policy. Our numerical analysis has been able to isolate and showcase the effect and the value of personalized offering. However, it remains an intriguing question whether our results extend to other customer learning schemes that may be able to correct for selection biases due to purchasing (e.g., Besbes and Scarsini, 2014), or even those due to the adoption of personalized offerings.

We have considered a simplified model for consumer behaviors in which all customers are homogeneous in terms of their beliefs and behaviors except for their individual preferences. In reality, customers may be more diverse in their ways of generating and interpreting product reviews. For instance, customers may subject to “under-reporting” biases in reviews as people tend to write reviews only when they are either extremely satisfied or extremely unsatisfied (Hu et al., 2009); previously posted reviews may affect customers’ decisions on whether and what to contribute (Moe and Schweidel, 2012). As a result, it is worthwhile to check the validity of our results under richer consumer behavior models that are able to capture such effects.

APPENDIX A

Proofs of Results in Chapter 2

A.1 Proof of Proposition 2.2

Proof. $y_t^*(\pi_t)$ is the solution to

$$H_t(x|\pi_t) = -p + (h+p)\Phi(x|\pi_t) + \alpha \mathsf{E}_{D_t|\pi_t}\left[\frac{\partial C_{t+1}}{\partial x}(x - D_t|\pi_t \circ D_t)\right] = -c.$$

For $t = T$, $\frac{\partial C_{T+1}}{\partial x}(\cdot|\cdot) = 0$, thus $H_T(y_T^*(\pi_T)|\pi_T) = -p + (h+p)\Phi(y_T^*(\pi_T)|\pi_T) = -c$, or $\Phi(y_T^*(\pi_T)|\pi_T) = \frac{p-c}{p+h} = \Phi(y_T^M|\pi_T)$, namely, $y_T^M(\pi_T) = y_T^*(\pi_T)$. One can show that due to the convexity of $C_{t+1}(\cdot|\cdot)$,

$$\mathsf{E}_{D_t|\pi_t}\left[\frac{\partial C_{t+1}}{\partial x}(x - D_t|\pi_t \circ D_t)\right] \geq -c,$$

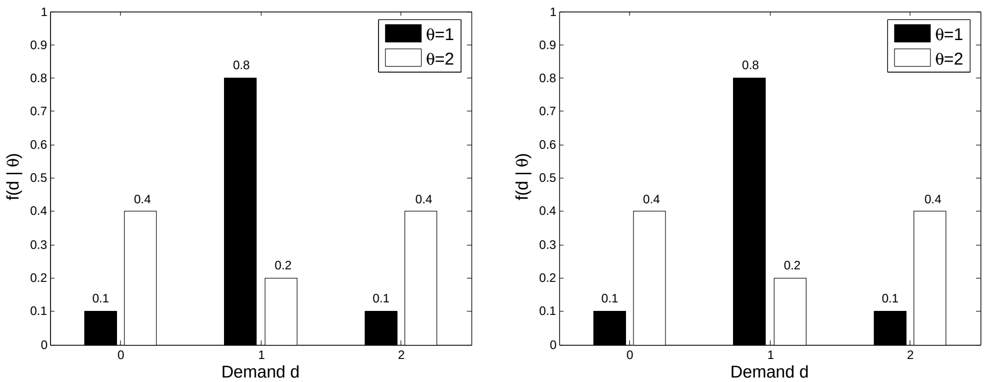
therefore for $t = 1, \ldots, T-1$,

$$\begin{aligned} H_t(y_t^*(\pi_t)|\pi_t) &= -p + (h+p)\Phi(y_t^*(\pi_t)|\pi_t) + \alpha \mathsf{E}_{D_t|\pi_t}\left[\frac{\partial C_{t+1}}{\partial x}(x - D_t|\pi_t \circ D_t)\right] \\ &\geq -p + (h+p)\Phi(y_t^*(\pi_t)|\pi_t) - \alpha c. \end{aligned}$$

Since $y_t^*(\pi_t)$ satisfies $H_t(y_t^*(\pi_t)|\pi_t) = -c$, we have

$$\Phi(y_t^*(\pi_t)|\pi_t) \leq \frac{p - (1-\alpha)c}{p+h} = \Phi(y_t^M(\pi_t)|\pi_t),$$

namely, $y_t^*(\pi_t) \leq y_t^M(\pi_t)$. $\square$



(c) (d)

Figure A.1: An example showing that Proposition 2.3(b) may not hold if $f(\cdot|\theta)$ does not have the MLRP property.

A.2 Proof of Proposition 2.3

Proof. We postpone the proof of part (a) until Appendix A.4, where Lemma A.3 includes this result as a special case. Alternatively, part (a) can be proved directly with an extension of Theorem 2 in Scarf (1959) to all demand distribution families that have MLRP.

To prove part (b), we note that for any $d_\tau < d'_\tau$, Lemma 2 of Chen (2010) establishes that $\pi_t \leq_{lr} \pi'_t$. The result then follows from (a). $\square$

We provide an example showing that it is necessary for $f(\cdot|\theta)$ to have the MLRP property for part (b) to hold. Let the demand parameter θ take values in the set $\{1, 2\}$. Demand in each period is 0, 1, or 2 units, and the demand probability mass function is shown in Figure A.1(c). Note that $f(\cdot|\theta)$ does not have the MLRP property; in particular, $\frac{f(0|\theta=2)}{f(0|\theta=1)} = \frac{f(2|\theta=2)}{f(2|\theta=1)} = 4 > 0.25 = \frac{f(1|\theta=2)}{f(1|\theta=1)}$. Now consider a two-period inventory problem. Figure A.1(d) shows the predictive cumulative demand distribution given a uniform initial prior $\pi_1(\theta = 1) = \pi_1(\theta = 2) = 0.5$. Suppose we choose cost parameters such that the newsvendor critical ratio determining the period 2 base-stock level is 0.70. Then the optimal base-stock level in period 2 is two units if $d_1 = 0$ or $d_1 = 2$ but is one unit if $d_1 = 1$. Therefore, the optimal base-stock level is not increasing in d_1.

A.3 Proof of Lemma 2.1

Proof. The proof is by induction. The lemma is true for $t = 1$. Suppose it is true for some $t \geq 1$; that is

$$\pi_t(\cdot|\mathbf{d}_{t-1}) = (1 - \gamma_t(\mathbf{d}_{t-1}))\pi_t^h(\cdot|\mathbf{d}_{t-1}) + \gamma_t(\mathbf{d}_{t-1})\pi_t^c(\cdot|\mathbf{d}_{t-1}).$$

Using Bayes rule, for $i \in \{h, c\}$, we have

$$\pi_{t+1}^i(\theta|\mathbf{d}_t) = \frac{\pi_t^i(\theta|\mathbf{d}_{t-1})f(d_t|\theta)}{\int_\Theta \pi_t^i(\omega|\mathbf{d}_{t-1})f(d_t|\omega)\mathrm{d}\omega},$$

and

$$\pi_{t+1}(\theta|\mathbf{d}_t) = \frac{[(1 - \gamma_t(\mathbf{d}_{t-1}))\pi_t^h(\theta|\mathbf{d}_{t-1}) + \gamma_t(\mathbf{d}_{t-1})\pi_t^c(\theta|\mathbf{d}_{t-1})]f(d_t|\theta)}{\int_\Theta[(1 - \gamma_t(\mathbf{d}_{t-1}))\pi_t^h(\omega|\mathbf{d}_{t-1}) + \gamma_t(\mathbf{d}_{t-1})\pi_t^c(\omega|\mathbf{d}_{t-1})]f(d_t|\theta)\mathrm{d}\omega}.$$

Write $I^i = \int_\Theta \pi_t^i(\omega|\mathbf{d}_{t-1})f(d_t|\omega)\mathrm{d}\omega$ for $i \in \{h, c\}$, then

$$\begin{aligned}
\pi_{t+1}(\theta|\mathbf{d}_t) &= \frac{(1 - \gamma_t(\mathbf{d}_{t-1}))\pi_t^h(\theta|\mathbf{d}_{t-1})f(d_t|\theta)}{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h + \gamma_t(\mathbf{d}_{t-1})I^c} + \frac{\gamma_t(\mathbf{d}_{t-1})\pi_t^c(\theta|\mathbf{d}_{t-1})f(d_t|\theta)}{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h + \gamma_t(\mathbf{d}_{t-1})I^c} \\
&= \frac{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h}{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h + \gamma_t(\mathbf{d}_{t-1})I^c}\pi_{t+1}^h(\theta|\mathbf{d}_t) \\
&\quad + \frac{\gamma_t(\mathbf{d}_{t-1})I^c}{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h + \gamma_t(\mathbf{d}_{t-1})I^c}\pi_{t+1}^c(\theta|\mathbf{d}_t).
\end{aligned}$$

By defining

$$\gamma_{t+1}(\mathbf{d}_t) = \frac{\gamma_t(\mathbf{d}_{t-1})I^c}{(1 - \gamma_t(\mathbf{d}_{t-1}))I^h + \gamma_t(\mathbf{d}_{t-1})I^c},$$

the lemma is true for $t + 1$, which completes the induction. $\square$

A.4 Proof of Proposition 2.6

For the purposes of this section, we consider an T-period generalized Bayesian inventory problem as described below. Let $\hat{C}_t(x|\pi)$ be the optimal expected cost for periods $t, \ldots, T$

given initial inventory level x and prior distribution π, where

$$\hat{C}_t(x|\pi) = \min_{y \geq x} \left\{ c(y-x) + L(y|\pi) + \alpha \mathsf{E}_{\mathbf{D}_t = (\hat{D}_t, D_t)|\pi} \left[\hat{C}_{t+1}(y - \hat{D}_t | \pi \circ D_t) \right] \right\},$$

with terminal value $\hat{C}_{T+1}(\cdot|\cdot) = 0$. We assume that $\hat{D}_t$ and D_t have the same marginal distribution induced by the prior π but their dependence is induced by some copula. We denote the minimizer of this expression by $\hat{y}_t^*(\pi)$. Note that the original and the independentized problems are both special cases of this formulation. In the original problem, $\hat{D}_t = D_t$, whereas in the independentized problem, $\hat{D}_t$ and D_t are independent with the same distribution induced by π.

The proof of Proposition 2.6 requires a few lemmas:

Lemma A.1. *For all π and $t = 1, \ldots, T+1$:*

(i) $\hat{C}_t(x|\pi)$ has a continuous derivative with respect to x, and is convex with respect to x;

(ii) The optimal policies are defined by single critical numbers $\hat{y}_t^*(\pi) \geq 0$;

(iii) $\hat{C}_t(x|\pi)$ has a continuous second derivative with respect to x at all points except perhaps $x = \hat{y}_t^*(\pi)$, at which point both the left and right hand second derivatives exist.

We omit the proof, as it is a minor modification of the one for Proposition 2.1.

Lemma A.2. *Let $\mathbf{D}^i = (\hat{D}, D)|\pi^i$ be a random vector in which $\hat{D}$ and D have the same marginal predictive demand density $\phi(\cdot|\pi^i)$, for $i = 1, 2$, and suppose that $\mathbf{D}^1$ and $\mathbf{D}^2$ have a common copula. If $\pi^1 \leq_{lr} \pi^2$, then $\mathbf{D}^1 \leq_{st} \mathbf{D}^2$.*

Proof. Let $\hat{D}|\pi^i$ and $D|\pi^i$ denote random variables with density $\phi(\cdot|\pi^i)$ for $i = 1, 2$, then $\hat{D}|\pi^1 \leq_{st} \hat{D}|\pi^2$ and $D|\pi^1 \leq_{st} D|\pi^2$ (Lemma 2(d), Chen, 2010). The lemma follows from Theorem 6.B.14 in Shaked and Shanthikumar (2007). □

Lemma A.3. *If $\pi^1 \leq_{lr} \pi^2$, the following hold for all x, π and $t = 1, \ldots, T+1$:*

(i) $\frac{\partial \hat{C}_t}{\partial x}(x|\pi^1) \geq \frac{\partial \hat{C}_t}{\partial x}(x|\pi^2)$;

(ii) $\hat{y}_t^*(\pi^1) \leq \hat{y}_t^*(\pi^2)$.

Proof. The proof is by induction. The lemma clearly holds when $t = T+1$ because $\hat{C}_{T+1}(\cdot|\cdot) = 0$. Suppose it is true for $t+1$. One can show that

$$\frac{\partial \hat{C}_t}{\partial x}(x|\pi) = \begin{cases} -c & , x < \hat{y}_t^*(\pi), \\ \hat{H}_t(x|\pi) & , x \geq \hat{y}_t^*(\pi), \end{cases}$$

where function $\hat{H}_t(\cdot|\pi)$ is defined by

$$\hat{H}_t(x|\pi) = -p + (h+p)\Phi(x|\pi) + \alpha \mathsf{E}_{(\hat{D}_t, D_t)|\pi}\left[\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{D}_t|\pi \circ D_t)\right].$$

For fixed $(\hat{d}_t, d_t)$, by the induction assumption, we have

$$\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t|\pi^1 \circ d_t) \geq \frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t|\pi^2 \circ d_t), \tag{A.1}$$

since $\pi^1 \circ d_t \leq_{lr} \pi^2 \circ d_t$ (Lemma 2(c), Chen, 2010). In addition, for $\hat{d}_t^1 \leq \hat{d}_t^2$, $d_t^1 \leq d_t^2$, we have

$$\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t^1|\pi^2 \circ d_t^1) \geq \frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t^1|\pi^2 \circ d_t^2) \tag{A.2}$$

$$\geq \frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t^2|\pi^2 \circ d_t^2), \tag{A.3}$$

where (A.2) follows from the induction assumption and that $\pi^2 \circ d_t^1 \leq_{lr} \pi^2 \circ d_t^2$ (Lemma 2(a), Chen, 2010), and (A.3) from the convexity of $\hat{C}_{t+1}(\cdot|\pi^2 \circ d_t^2)$. Therefore, $\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t|\pi^2 \circ d_t)$ is decreasing in $(\hat{d}_t, d_t)$. In addition, $\pi^1 \leq_{lr} \pi^2$ together with Lemma A.2 imply that

$$(\hat{D}_t, D_t)|\pi^1 \leq_{st} (\hat{D}_t, D_t)|\pi^2. \tag{A.4}$$

We thus have

$$\mathsf{E}_{(\hat{D}_t, D_t)|\pi^1}\left[\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{D}_t|\pi^1 \circ D_t)\right] \geq \mathsf{E}_{(\hat{D}_t, D_t)|\pi^1}\left[\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{D}_t|\pi^2 \circ D_t)\right] \tag{A.5}$$

$$\geq \mathsf{E}_{(\hat{D}_t, D_t)|\pi^2}\left[\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{D}_t|\pi^2 \circ D_t)\right], \tag{A.6}$$

where (A.5) results from (A.1), and (A.6) from (A.4) and the fact that $\frac{\partial \hat{C}_{t+1}}{\partial x}(x - \hat{d}_t|\pi^2 \circ d_t)$ is decreasing in $(\hat{d}_t, d_t)$ (Section 6.B.1, Shaked and Shanthikumar, 2007). We conclude that $\hat{H}_t(x|\pi^1) \geq \hat{H}_t(x|\pi^2)$. Note that $\hat{y}_t^*(\pi)$ is the solution to the equation $H_t(x|\pi) = -c$. Also note that $H_t(x|\pi)$ is increasing in x. Hence,

$$\hat{H}_t(\hat{y}_t^*(\pi^1)|\pi^1) = -c = \hat{H}_t(\hat{y}_t^*(\pi^2)|\pi^2) \leq \hat{H}_t(\hat{y}_t^*(\pi^2)|\pi^1),$$

which indicates that $\hat{y}_t^*(\pi^1) \leq \hat{y}_t^*(\pi^2)$.

It remains to show that $\frac{\partial \hat{C}_t}{\partial x}(x|\pi^1) \geq \frac{\partial \hat{C}_t}{\partial x}(x|\pi^2)$. Consider three cases:

(i) $x < \hat{y}_t^*(\pi^1)$. In this case, $\frac{\partial \hat{C}_t}{\partial x}(x|\pi^1) = \frac{\partial \hat{C}_t}{\partial x}(x|\pi^2) = -c$;

(ii) $\hat{y}_t^*(\pi^1) \leq x < \hat{y}_t^*(\pi^2)$. In this case, $\frac{\partial \hat{C}_t}{\partial x}(x|\pi^1) = \hat{H}(x|\pi^1) \geq \hat{H}(\hat{y}_t^*(\pi^1)|\pi^1) = -c = \frac{\partial \hat{C}_t}{\partial x}(x|\pi^2)$;

(iii) $x > \hat{y}_t^*(\pi^2)$. In this case, $\frac{\partial \hat{C}_t}{\partial x}(x|\pi^1) = \hat{H}(x|\pi^1) \geq \hat{H}(x|\pi^2) = \frac{\partial \hat{C}_t}{\partial x}(x|\pi^2)$;

This completes the induction proof. □

With these lemmas established, we can now proceed to the proof of Proposition 2.6.

Proof of Proposition 2.6. The proof is by induction. The proposition is clearly true when $t = T + 1$. Suppose for period $t + 1$, $C_{t+1}^{\perp}(x|\pi) \leq C_{t+1}(x|\pi)$ for all x, π.

Fix y. Consider function $K(d_t^{\perp}, d_t) = C_{t+1}^{\perp}(y - d_t^{\perp}|\pi \circ d_t)$. Taking the derivative with respect to $d_t^{\perp}$, we obtain

$$\frac{\partial K}{\partial d_t^{\perp}}(d_t^{\perp}, d_t) = -\frac{\partial C_{t+1}^{\perp}}{\partial (y - d_t^{\perp})}(y - d_t^{\perp}|\pi \circ d_t).$$

For $d_t^1 \leq d_t^2$, Lemma 2 of Chen (2010) implies that $\pi \circ d_t^1 \leq_{lr} \pi \circ d_t^2$. Lemma A.3 therefore yields

$$\frac{\partial K}{\partial d_t^{\perp}}(d_t^{\perp}, d_t^1) \leq \frac{\partial K}{\partial d_t^{\perp}}(d_t^{\perp}, d_t^2).$$

In other words, $K(\cdot, \cdot)$ has increasing differences in $(d_t^{\perp}, d_t)$. Thus, $K(\cdot, \cdot)$ is supermodular in $(d_t^{\perp}, d_t)$.

Let $F_t^o(\hat{d}_t, d_t)$ and $F_t^{\perp}(\hat{d}_t, d_t)$ be the distribution functions of the random vectors $\mathbf{D}_t^o = (D_t, D_t)|\pi$ and $\mathbf{D}_t^{\perp} = (D_t^{\perp}, D_t)|\pi$, respectively. Then we have

$$\begin{aligned} F_t^{\perp}(\hat{d}_t, d_t) &= \mathsf{P}\left\{D_t^{\perp} \le \hat{d}_t, D_t \le d_t\right\} = \mathsf{P}\left\{D_t^{\perp} \le \hat{d}_t\right\} \mathsf{P}\left\{D_t \le d_t\right\} \\ &\le \min\left[\mathsf{P}\left\{D_t^{\perp} \le \hat{d}_t\right\}, \mathsf{P}\left\{D_t \le d_t\right\}\right] = \min\left[\mathsf{P}\left\{D_t \le \hat{d}_t\right\}, \mathsf{P}\left\{D_t \le d_t\right\}\right] \\ &= \mathsf{P}\left\{D_t \le \hat{d}, D_t \le d\right\} = F_o(\hat{d}, d). \end{aligned}$$

Therefore, by (9.A.3) in Shaked and Shanthikumar (2007), $\mathbf{D}_t^o$ and $\mathbf{D}_t^{\perp}$ are ranked in the positive quadrant dependent (PQD) order: $\mathbf{D}_t^{\perp} = (D_t^{\perp}, D_t)|\pi \le_{PQD} (D_t, D_t)|\pi = \mathbf{D}_t^o$. By (9.A.18) in Shaked and Shanthikumar (2007), $\mathbf{D}_t^o$ and $\mathbf{D}_t^{\perp}$ are thus ranked in the supermodular order as follows:

$$\mathbf{D}_t^{\perp} = (D_t^{\perp}, D_t)|\pi \le_{sm} (D_t, D_t)|\pi = \mathbf{D}_t^o. \tag{A.7}$$

We finally have

$$\begin{aligned} C_t^{\perp}(x|\pi) &= \min_{y \ge x}\left\{c(y-x) + L(y|\pi) + \alpha \mathsf{E}_{\mathbf{D}_t^{\perp}=(D_t^{\perp}, D_t)|\pi}\left[K(D_t^{\perp}, D_t)\right]\right\} \\ &\le \min_{y \ge x}\left\{c(y-x) + L(y|\pi) + \alpha \mathsf{E}_{\mathbf{D}_t^o=(D_t, D_t)|\pi}\left[K(D_t, D_t)\right]\right\} \\ &= \min_{y \ge x}\left\{c(y-x) + L(y|\pi) + \alpha \mathsf{E}_{\mathbf{D}_t^o=(D_t, D_t)|\pi}\left[C_{t+1}^{\perp}(y - D_t|\pi \circ D_t)\right]\right\} \\ &\le \min_{y \ge x}\left\{c(y-x) + L(y|\pi) + \alpha \mathsf{E}_{\mathbf{D}_t^o=(D_t, D_t)|\pi}\left[C_{t+1}(y - D_t|\pi \circ D_t)\right]\right\} \\ &= C_t(x|\pi), \end{aligned}$$

where the first inequality follows from (A.7) and the definition of supermodular ordering, and the second follows from the induction assumption. This completes the proof. □

A.5 Proof of Proposition 2.8

Proof. Let $\mathbf{d}_T = (d_1, \ldots, d_T)$ and $\mathbf{d}'_T = (d'_1, \ldots, d'_T)$ be two demand paths such that $d_t \le d'_t$ for all t. Let $\mathbf{d}_t = (d_1, \ldots, d_t)$ and $\mathbf{d}'_t = (d'_1, \ldots, d'_t)$ be the subsequences of $\mathbf{d}_T$ and $\mathbf{d}'_T$ until period t, respectively. Let $\mathbf{y}(\pi) = (y_t(\pi \circ \mathbf{d}_{t-1}))$ denote the myopic (also the optimal) policy. It follows that $\underline{\pi} \circ \mathbf{d}_{t-1} \leq_{lr} \pi \circ \mathbf{d}_{t-1}$ and hence

$$G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\underline{\pi} \circ \mathbf{d}_{t-1}) \leq G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\pi \circ \mathbf{d}_{t-1}) \tag{A.8}$$

for all $\pi \in \mathcal{P}$ because $G(y|\pi) \leq G(y|\pi')$ for all $\pi \leq_{lr} \pi'$. To see this, note that $D|\pi \leq_{st} D|\pi'$ for all $\pi \leq_{lr} \pi'$ and that $\min\{y, d\}$ is an increasing function in d.

We also have $\underline{\pi} \circ \mathbf{d}_{t-1} \leq_{lr} \underline{\pi} \circ \mathbf{d}'_{t-1} \leq_{lr} \pi \circ \mathbf{d}'_{t-1}$ for all $\pi \in \mathcal{P}$, which implies that $y_t(\underline{\pi} \circ \mathbf{d}_{t-1}) \leq y_t(\underline{\pi} \circ \mathbf{d}'_{t-1}) \leq y_t(\pi \circ \mathbf{d}'_{t-1})$. Therefore, we have

$$G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\pi \circ \mathbf{d}_{t-1}) \leq G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\pi \circ \mathbf{d}'_{t-1}) \leq G(y_t(\underline{\pi} \circ \mathbf{d}'_{t-1})|\pi \circ \mathbf{d}'_{t-1}), \tag{A.9}$$

where the first inequality follows from the fact that $\pi \circ \mathbf{d}_{t-1} \leq_{lr} \pi \circ \mathbf{d}'_{t-1}$ and the second from the fact that $G(y|\pi \circ \mathbf{d}'_{t-1})$ is increasing over $y_t(\underline{\pi} \circ \mathbf{d}_{t-1}) \leq y \leq y_t(\pi \circ \mathbf{d}'_{t-1})$. As a consequence,

$$\begin{aligned} E[G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\underline{\pi} \circ \mathbf{d}_{t-1})|\underline{\pi}] &\leq E[G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\pi \circ \mathbf{d}_{t-1})|\underline{\pi}] \\ &\leq E[G(y_t(\underline{\pi} \circ \mathbf{d}_{t-1})|\pi \circ \mathbf{d}_{t-1})|\pi] \end{aligned}$$

for all $\pi \in \mathcal{P}$. The expectations are with respect to the random variable $\mathbf{d}_{t-1}$ over $\underline{\pi}$ and π in the first two expressions and the third expression, respectively. The first inequality directly follows from (A.8) whereas the second is due to (A.8) and that $\underline{\pi} \leq_{lr} \pi$.

Denote by $\Pi_T(\mathbf{y}, \pi)$ the expected total profit when policy $\mathbf{y}$ is employed and π is used as the "true" prior for generating the Bayesian demand process. More specifically, $\Pi_T(\mathbf{y}, \pi) = \sum_{t=1}^{T} E[G(y_t|\pi \circ \mathbf{D}_{t-1})|\pi]$. With this notation we can write $R_T(\mathcal{P}) = \max_{\mathbf{y}} \min_{\pi \in \mathcal{P}} \Pi_T(\mathbf{y}, \pi)$. Let $\pi^*(\mathbf{y}) = \arg\min_{\pi \in \mathcal{P}} \Pi_T(\mathbf{y}, \pi)$ for any policy $\mathbf{y}$, thus $\Pi_T(\mathbf{y}, \pi^*(\mathbf{y})) \leq \Pi_T(\mathbf{y}, \underline{\pi})$ by definition. For policy $\mathbf{y}(\underline{\pi})$, it follows from the previous result that $\pi^*(\mathbf{y}(\underline{\pi})) = \underline{\pi}$, or $\Pi_T(\mathbf{y}(\underline{\pi}), \underline{\pi}) \leq \Pi_T(\mathbf{y}(\underline{\pi}), \pi)$ for all $\pi \in \mathcal{P}$. Together with the fact that, for any policy $\mathbf{y}$, $\Pi_T(\mathbf{y}, \underline{\pi}) \leq \Pi_T(\mathbf{y}(\underline{\pi}), \underline{\pi}) = V_T(\underline{\pi})$, we have $R_T(\mathcal{P}) = \max_{\mathbf{y}} \min_{\pi \in \mathcal{P}} \Pi_T(\mathbf{y}, \pi) = \max_{\mathbf{y}} \Pi_T(\mathbf{y}, \pi^*(\mathbf{y})) = \Pi_T(\mathbf{y}(\underline{\pi}), \underline{\pi}) = V_T(\underline{\pi})$. $\square$

APPENDIX B

Proofs of Results in Chapter 3

B.1 Proof of Lemma 3.1

The proof of Lemma 3.1 employs results from the statistical literature on comparison of experiments (Blackwell, 1951, 1953). In the following, we introduce the definitions of statistic experiment.

Definition B.1 (Blackwell, 1951, 1953; Ginebra, 2007). *A statistical experiment* $\mathcal{E} = \{(X, S_X); (P_\theta, \Theta)\}$ *($\mathcal{E}\{X; P_\theta\}$ for short) yields an observation on a random variable* X *defined on* S_X*, with an unknown probability distribution that is known to be in the family* $(P_\theta, \theta \in \Theta)$.

The two observation types introduced in §3.3.1 can be viewed as outcomes of the following statistical experiments.

(i) *Observations without timing information ($X^{NT}(\mathbf{q})$)*: Consider $\mathcal{E}^{NT}(\mathbf{q}) = \{X^{NT}(\mathbf{q}); P_\theta^{NT}\}$, where $X^{NT}(\mathbf{q}) = \{\mathbf{s}, \mathbf{e}\}$ as defined in §3.3.1. The joint distribution P_θ^{NT} of $X^{NT}(\mathbf{q})$ under $\theta \in \Theta$ is given by (3.2). $\mathcal{E}^{NT}(\mathbf{q})$ is by definition a statistical experiment which corresponds to using an allocation $\mathbf{q}$ when timing information is unobservable.

(ii) *Observations with timing information ($X^{T}(\mathbf{q})$)*: Consider $\mathcal{E}^{T}(\mathbf{q}) = \{X^{T}(\mathbf{q}); P_\theta^{T}\}$, where $X^{T}(\mathbf{q}) = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ as defined in §3.3.1. The joint distribution P_θ^{T} of $X^{T}(\mathbf{q})$ under $\theta \in \Theta$ is given by (3.3). $\mathcal{E}^{T}(\mathbf{q})$ is by definition a statistical experiment which corresponds to using an allocation $\mathbf{q}$ when timing information is observable.

Next, we introduce the sufficiency ordering between experiments.

Definition B.2 (Ginebra, 2007). *Experiment* $\mathcal{E} = \{X; P_\theta\}$ *is* sufficient for *experiment* $\mathcal{F} = \{Y; Q_\theta\}$ *if there is a stochastic transformation of* X *to a random variable* $W(X)$ *such that* $W(X)$ *and* Y *have identical distribution under each* $\theta \in \Theta$.

The following lemma is a restatement of Proposition 3.2 in Ginebra (2007) by **?**.

Lemma B.1 (Ginebra, 2007; Jain et al., 2015)**.** *Experiment $\mathcal{E}$ is sufficient for experiment $\mathcal{F}$ if and only if, for every decision problem involving θ, the Bayes risk for $\mathcal{E}$ does not exceed the Bayes risk for $\mathcal{F}$, i.e., $r_\pi(\mathcal{E}) \leq r_\pi(\mathcal{F})$ for all prior π on θ.*

Finally, we prove Lemma 3.1 by establishing sufficiency orderings between experiments with different test inventory allocations.

Proof of Lemma 3.1(a). Let $\mathcal{E}^T = \{X^T(\mathbf{q}+\delta_n); P_\theta\}$ and $\mathcal{F}^T = \{X^T(\mathbf{q}); Q_\theta\}$ be the experiments with timing information under allocation $\mathbf{q}+\delta_n$ and allocation $\mathbf{q}$, respectively. Consider the following transformation from $X^T(\mathbf{q}+\delta_n) = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ to $X' = \{\mathbf{s}', \mathbf{e}', \vec{\tau}'\}$:

(1) For all $m \neq n$: let $s'_m = s_m$, $e'_m = e_m$, ;

(2) if $s_n = q_n + 1$, $e_n = 1$: let $s'_n = q_n$, $e'_n = 1$, and $\vec{\tau}'_n = (\tau_n^1, \tau_n^2, \ldots, \tau_n^{q_n})$;

(3) if $s_n = q_n$, $e_n = 0$: let $s'_n = q_n$, $e'_n = 1$, and $\vec{\tau}'_n = (\tau_n^1, \tau_n^2, \ldots, \tau_n^{q_n})$;

(4) if $s_n < q_n$, $e_n = 0$: let $s'_n = s_n$, $e_n = 0$, $\vec{\tau}'_n = (\tau_n^1, \tau_n^2, \ldots, \tau_n^{s_n})$.

It can be verified that X' and $X^T(\mathbf{q})$ have identical distributions. Therefore, $\mathcal{E}^T$ is sufficient for $\mathcal{F}^T$ according to Definition B.2.

We can recast our merchandise testing problem as a statistical decision problem with period 2 ordering quantities $\mathbf{y}$ as the decision, and a loss function $L(\mathbf{y}, \theta) = \hat{\Pi}^*(\theta) - \hat{\Pi}(\mathbf{y}|\theta)$, where $\hat{\Pi}^*(\theta)$ is the optimal expected profit in period 2 under θ, and $\hat{\Pi}(\mathbf{y}|\theta)$ the expected profit if the retailer's ordering decision is $\mathbf{y}$. The Bayes risk is thus $r_\pi(\mathcal{E}^T) = E_\pi[E_{P_\theta}[L(\mathbf{y}(\pi \circ X^T(\mathbf{q}+\delta_n)), \theta)]] = E_\pi[E_{P_\theta}[\hat{\Pi}^*(\theta)]] - E_\pi[E_{P_\theta}[\hat{V}(\pi \circ X^T(\mathbf{q}+\delta_n))]] = E_\pi[E_{P_\theta}[\hat{\Pi}^*(\theta)]] - \Pi^T(\mathbf{q}+\delta_n|\pi)$ for $\mathcal{E}^T$, and is $r_\pi(\mathcal{F}^T) = E_\pi[E_{P_\theta}[L(\mathbf{y}(\pi \circ X^T(\mathbf{q})), \theta)]] = E_\pi[E_{P_\theta}[\hat{\Pi}^*(\theta)]] - E_\pi[E_{Q_\theta}[\hat{V}(\pi \circ X^T(\mathbf{q}))]] = E_\pi[E_{Q_\theta}[\hat{\Pi}^*(\theta)]] - \Pi^T(\mathbf{q}|\pi)$ for $\mathcal{F}^T$. Note that $E_\pi[E_{P_\theta}[\hat{\Pi}^*(\theta)]] = E_\pi[E_{Q_\theta}[\hat{\Pi}^*(\theta)]] = E_\pi[\hat{\Pi}^*(\theta)]$. Lemma 3.1(a) thus follows from $r_\pi(\mathcal{E}^{NT}) \leq r_\pi(\mathcal{F}^{NT})$ as a result of Lemma B.1. □

Proof of Lemma 3.1(b). Let $\mathcal{E}^{NT} = \{X^{NT}(\mathbf{q}+\delta_n); P_\theta\}$ and $\mathcal{F}^{NT} = \{X^{NT}(\mathbf{q}); Q_\theta\}$ be the experiments without timing information under allocation $\mathbf{q}+\delta_n$ and allocation $\mathbf{q}$, respectively. Consider the following transformation from $X^{NT}(\mathbf{q}+\delta_n) = \{\mathbf{s}, \mathbf{e}\}$ to $X' = \{\mathbf{s}', \mathbf{e}'\}$:

(1) For all $m \neq n$: let $s'_m = s_m$, $e'_m = e_m$;

(2) if $s_n = q_n + 1$, $e_n = 1$: let $s'_n = q_n$, $e'_n = 1$;

(3) if $s_n = q_n$, $e_n = 0$: let $s'_n = q_n$, $e'_n = 1$;

(4) if $s_n < q_n$, $e_n = 0$: let $s'_n = s_n$, $e_n = 0$.

It can be verified that X' and $X^{NT}(\mathbf{q})$ have identical distributions. Therefore, $\mathcal{E}^{NT}$ is sufficient for $\mathcal{F}^{NT}$ according to Definition B.2. We show $\Pi^{NT}(\mathbf{q}) \leq \Pi^{NT}(\mathbf{q} + \delta_n)$ by an argument similar to that in proving part (a). $\square$

Proof of Lemma 3.1(c). Suppose that allocation $\mathbf{q}^*$ is optimal with timing information, i.e., $\Pi^T(\mathbf{q}^*) \geq \Pi^T(\mathbf{q})$ for any allocation $\mathbf{q}$. If $\sum_{n=1}^N q^*_n < Q$, consider allocation $\mathbf{q}' = \mathbf{q}^* + (Q - \sum_{n=1}^N q^*_n)\delta_1$, which has $\sum_{n=1}^N q'_n = Q$. It follows from part (a) that $\Pi^T(\mathbf{q}') \geq \Pi^T(\mathbf{q}^*)$, which indicates that $\mathbf{q}'$ is also optimal. A similar argument applies to the case without timing information. $\square$

B.2 Proof of Proposition 3.1

Proof of Proposition 3.1(a). Let $\mathcal{E} = \{X^T(\mathbf{q} - \delta_i + \delta_j); P_\theta\}$ and $\mathcal{F} = \{X^T(\mathbf{q}); Q_\theta\}$ be the experiments with timing information under allocation $\mathbf{q} - \delta_i + \delta_j$ and allocation $\mathbf{q}$, respectively. Consider the following transformation from $X^T(\mathbf{q} - \delta_i + \delta_j) = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ to $X' = \{\mathbf{s}', \mathbf{e}', \vec{\tau}'\}$:

(1) For $m \neq i, j$, let $s'_m = s_m$, $\vec{\tau}'_m = \vec{\tau}_m$, and $e'_m = e_m$;

(2) Let $s'_j = 0$, $\vec{\tau}'_j = \varnothing$, $e'_j = 1$;

(3) if $s_i < q_i - 1$: let $s'_i = s_i, \vec{\tau}'_i = \vec{\tau}_i$, and $e'_i = 0$;

(4) if $s_i = q_i - 1$, $s_j = 0$: let $s'_i = q_i - 1$, $\vec{\tau}'_i = \vec{\tau}_i$, $e'_i = 0$;

(5) if $s_i = q_i - 1$, $s_j = 1$, and $\tau^1_j > T - \sum_{k=1}^{q_i - 1} \tau^k_i$: let $s'_i = q_i - 1$, $\vec{\tau}'_i = \vec{\tau}_i$, and $e'_i = 0$;

(6) if $s_i = q_i - 1$, $s_j = 1$, and $\tau^1_j \leq T - \sum_{k=1}^{q_i - 1} \tau^k_i$: let $s'_i = q_i$, $\vec{\tau}'_i = \{\tau^1_i, \ldots, \tau^{q_i - 1}_i, \tau^1_j\}$, $e'_i = 1$.

It can be verified that X' and $X^T(\mathbf{q})$ have identical distributions. Therefore, $\mathcal{E}$ is sufficient for $\mathcal{F}$. We show $\Pi^T(\mathbf{q}) \leq \Pi^T(\mathbf{q} - \delta_i + \delta_j)$ by an argument similar to that in the proof of Lemma 3.1. □

Proof of Proposition 3.1(b). Immediately follows from part (a). □

B.3 Proof of Proposition 3.2

Proof of Proposition 3.2(a). We show the proof only for $N = 2$ and let $i = 1$ and $j = 2$ without loss of generality, namely, $\Pi^T(q_1, q_2|\alpha, \beta) \leq \Pi^T(q_1 - 1, q_2 + 1|\alpha, \beta)$ for all $q_2 \geq 0$ and $q_1 \geq q_2 + 2$. The proof extends to the $N > 2$ cases by conditioning on the demand processes at stores other than i and j.

We prove by two inductions—an inner induction nested inside an outer induction. Proposition 3.1 guarantees that part (a) holds for $q_2 = 0$ and all $q_1 \geq 2$. Suppose that it holds for some $q_2 = q \geq 0$ and all $q_1 \geq q + 2$, that is,

Assumption B.1. $\Pi_T^T(q_1, q|\alpha, \beta) \geq \Pi_T^T(q_1 - 1, q + 1|\alpha, \beta)$ *for some* $q \geq 0$ *and all* $q_1 \geq q + 2$.

Assumption B.1 is the assumption for the outer induction. The subscript T in Π_T^T makes explicit the length T of period 1. We first show it holds for $q_2 = q + 1$ and $q_1 = q_2 + 2 = q + 3$. By conditioning on the time $s \geq 0$ until the first demand arrival at either store, which is exponential with rate 2λ, we have

$$\begin{aligned}
&\Pi_T^T(q+3, q+1|\alpha, \beta)\\
&\quad = E_{\lambda|\alpha,\beta} E_{s|\lambda}\Bigg[\bigg\{\frac{1}{2}\Pi_{T-s}^T(q+2, q+1|\alpha+1, \beta+2s)\\
&\qquad\quad + \frac{1}{2}\Pi_{T-s}^T(q+3, q|\alpha+1, \beta+2s)\bigg\}\mathbb{1}_{s\leq T} + \Pi_0^T(q+3, q+1|\alpha, \beta+2T)\mathbb{1}_{s>T}\Bigg],\\
&\Pi_T^T(q+2, q+2|\alpha, \beta)\\
&\quad = E_{\lambda|\alpha,\beta} E_{s|\lambda}\Bigg[\bigg\{\frac{1}{2}\Pi_{T-s}^T(q+1, q+2|\alpha+1, \beta+2s)\\
&\qquad\quad + \frac{1}{2}\Pi_{T-s}^T(q+2, q+1|\alpha+1, \beta+2s)\bigg\}\mathbb{1}_{s\leq T} + \Pi_0^T(q+2, q+2|\alpha, \beta+2T)\mathbb{1}_{s>T}\Bigg].
\end{aligned}$$

Note that

(i) $\Pi^T_{T-s}(q+2,q+1|\alpha+1,\beta+2s) = \Pi^T_{T-s}(q+1,q+2|\alpha+1,\beta+2s)$ because stores are identical;

(ii) $\Pi^T_{T-s}(q+3,q|\alpha+1,\beta+2s) \leq \Pi^T_{T-s}(q+2,q+1|\alpha+1,\beta+2s)$ by Assumption B.1;

(iii) $\Pi^T_0(q+3,q+1|\alpha,\beta+2T) = \Pi^T_0(q+2,q+2|\alpha,\beta+2T)$ since period 1 has zero length.

As a result, $\Pi^T_T(q+3,q+1|\alpha,\beta) \leq \Pi^T_T(q+2,q+2|\alpha,\beta)$; i.e., Assumption B.1 holds for $q+1$ and $q_1 = q+3$.

We still need to show that Assumption B.1 holds for $q+1$ and all $q_1 > q+3$ to complete the outer induction. We prove that by an inner induction on q_1 which makes the following induction assumption.

Assumption B.2. $\Pi^T_T(q+\Delta_q+3,q+1|\alpha,\beta) \geq \Pi^T_T(q+\Delta_q+2,q+2|\alpha,\beta)$ *for* q *and some* $\Delta_q \geq 0$.

We have shown that Assumption B.2 is true for q and $\Delta_q = 0$. To show it holds for $\Delta_q = 1$, we again condition on the time $s \geq 0$ until the first demand arrival at either store and get

$$
\begin{aligned}
&\Pi^T_T(q+\Delta_q+1+3,q+1|\alpha,\beta)\\
&\quad= E_{\lambda|\alpha,\beta}E_{s|\lambda}\Bigg[\Bigg\{\frac{1}{2}\Pi^T_{T-s}(q+\Delta_q+3,q+1|\alpha+1,\beta+2s)\\
&\qquad+\frac{1}{2}\Pi^T_{T-s}(q+\Delta_q+4,q|\alpha+1,\beta+2s)\Bigg\}\mathbb{1}_{s\leq T}\\
&\qquad+\Pi^T_0(q+\Delta_q+4,q+1|\alpha,\beta+2T)\mathbb{1}_{s>T}\Bigg],\\
&\Pi^T_T(q+\Delta_q+1+2,q+2|\alpha,\beta)\\
&\quad= E_{\lambda|\alpha,\beta}E_{s|\lambda}\Bigg[\Bigg\{\frac{1}{2}\Pi^T_{T-s}(q+\Delta_q+2,q+2|\alpha+1,\beta+2s)\\
&\qquad+\frac{1}{2}\Pi^T_{T-s}(q+\Delta_q+3,q+1|\alpha+1,\beta+2s)\Bigg\}\mathbb{1}_{s\leq T}\\
&\qquad+\Pi^T_0(q+\Delta_q+3,q+2|\alpha,\beta+2T)\mathbb{1}_{s>T}\Bigg].
\end{aligned}
$$

Note that

(i) $\Pi^T_{T-s}(q+\Delta_q+3, q+1|\alpha+1, \beta+2s) \leq \Pi^T_{T-s}(q+\Delta_q+2, q+2|\alpha+1, \beta+2s)$ by Assumption B.2;

(ii) $\Pi^T_{T-s}(q+\Delta_q+4, q|\alpha+1, \beta+2s) \leq \Pi^T_{T-s}(q+\Delta_q+3, q+1|\alpha+1, \beta+2s)$ by Assumption B.1;

(iii) $\Pi^T_0(q+\Delta_q+4, q+1|\alpha, \beta+2T) = \Pi^T_0(q+\Delta_q+3, q+2|\alpha, \beta+2T)$ since period 1 has zero length.

Consequently we have $\Pi^T_T(q+\Delta_q+1+3, q+1|\alpha, \beta) \leq \Pi^T_T(q+\Delta_q+1+2, q+2|\alpha, \beta)$ which completes the inner induction. This also completes the outer induction in showing that Assumption B.1 holds for $q+1$ and all $q_1 \geq (q+1)+2 = q+3$. □

Proof of Proposition 3.2(b). Immediately follows from part (a). □

B.4 Proof of Proposition 3.3

Proof of Proposition 3.3(a). Let $\mathcal{E} = \{X^T(\mathbf{q}+\delta_i-\delta_j); P_\theta\}$ and $\mathcal{F} = \{X^T(\mathbf{q}); Q_\theta\}$ be the experiments with timing information under allocation $\mathbf{q}+\delta_i-\delta_j$ and allocation $\mathbf{q}$, respectively. Consider the following transformation from $X^T(\mathbf{q}+\delta_i-\delta_j) = \{\mathbf{s}, \mathbf{e}, \vec{\tau}\}$ to $X' = \{\mathbf{s}', \mathbf{e}', \vec{\tau}'\}$ for every $\theta \in \Theta$:

(1) For $m \neq i, j$, let $s'_m = s_m$, $\vec{\tau}'_m = \vec{\tau}_m$, and $e'_m = e_m$;

(2) let $s'_i = 0$, $\vec{\tau}'_i = \varnothing$, $e'_i = 1$;

(3) if $s_j < q_j - 1$ ($q_j \geq 2$): let $s'_j = s_j, \vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(4) if $s_i = 0$, $s_j = q_j - 1$: let $s'_j = s_j = q_j - 1$, $\vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(5) if $s_i = 1$, $s_j = q_j - 1$, and $\Psi_i(\tau^1_i|\theta) > \Psi_j(T - \sum_{k=1}^{q_j-1} \tau^k_j|\theta)$: let $s'_j = q_j - 1$, $\vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(6) if $s_i = 1$, $s_j = q_j - 1$, and $\Psi_i(\tau^1_i|\theta) \leq \Psi_j(T - \sum_{k=1}^{q_j-1} \tau^k_j|\theta)$: let $s'_j = q_j$, $\vec{\tau}'_j = \{\tau^1_j, \ldots, \tau^{q_j-1}_j, \Psi^{-1}_j(\Psi_i(\tau^1_i|\theta)|\theta)\}$, $e'_j = 1$,

where $\Psi_j^{-1}(\cdot|\theta)$ is the inverse cdf of the corresponding inter-arrival time cdf $\Psi_j(\cdot|\theta)$. It can be verified that X' and $X^T(\mathbf{q})$ have identical distributions under every $\theta \in \Theta$. Therefore, $\mathcal{E}$ is sufficient for $\mathcal{F}$. We show $\Pi^T(\mathbf{q}) \leq \Pi^T(\mathbf{q} + \delta_i - \delta_j)$ by an argument similar to that in the proof of Lemma 3.1. □

Proof of Proposition 3.3(b). Immediately follows from part (a). □

B.5 Proof of Proposition 3.4

The proof involves an equivalent transformation from a problem in which stores have non-identical arrival rates and identical lengths of period 1, to one in which stores have identical arrival rates and non-identical lengths of period 1. We use $\Pi^T(\mathbf{q}|\gamma_1\lambda_0, \ldots, \gamma_N\lambda_0; T_1, \ldots, T_N; \alpha, \beta)$ to denote the ex-ante expected profit of an allocation $\mathbf{q}$ for a merchandise testing problem with observable timing information, unknown arrival rate parameter λ_0, relative demand coefficients $\gamma_1, \ldots, \gamma_N$, and lengths of period 1, $T_1, \ldots, T_N$. The retailer has a gamma prior with parameters (α, β) on λ_0. The following lemma shows that under gamma-Poisson demand, a store n with arrival rate $\gamma_n\lambda_n$ and length T_n in period 1 is equivalent in terms of ex-ante expected profit to one with arrival rate λ_n and length $\gamma_n T_n$.

Lemma B.2. $\Pi^T(\mathbf{q}|\gamma_1\lambda_0, \ldots, \gamma_n\lambda_0, \ldots, \gamma_N\lambda_0; T_1, \ldots, T_N; \alpha, \beta) = \Pi^T(\mathbf{q}|\gamma_1\lambda_0, \ldots, \lambda_0, \ldots, \gamma_N\lambda_0; T_1, \ldots, \gamma_n T_n, \ldots, T_N)$ *for all* $\mathbf{q}$*,* $\gamma_n > 0$*,* $\alpha > 0$*, and* $\beta > 0$*.*

Proof. We show the proof only for $N = 1$, i.e., $\Pi^T(q|\gamma\lambda_0; T_0; \alpha, \beta) = \Pi^T(q|\lambda_0; \gamma T_0; \alpha, \beta)$. The result extends to the $N > 1$ cases by conditioning on the demand processes at other stores.

Observations during period 1 can be summarized by a pair of sufficient statistics (s, t), where s is the sales quantity and t is the effective sales duration. Let $\hat{\alpha}$ and $\hat{\beta}$ denote the posterior parameters. Let $P(D(T|\lambda) = s)$ be the probability of total arrivals being s during time $[0, T]$ in a Poisson process with arrival rate λ, and $g(t|q, \lambda)$ be the probability density of the q-th arrival time. Then,

(i) if $\lambda = \gamma\lambda_0$, $T = T_0$: $\hat{\alpha} = \alpha + s$, $\hat{\beta} = \beta + \gamma t$. By conditioning on unknown parameter λ_0 and observation (s, t), one obtains

$$
\begin{aligned}
\Pi^T(q|\gamma\lambda_0; T_0; \alpha, \beta) &= E_{\lambda_0|\alpha,\beta} E_{(s,t)|\gamma\lambda_0}[\hat{V}(\alpha + s, \beta + \gamma t)] \\
&= E_{\lambda_0|\alpha,\beta}\Bigg[\sum_{s=0}^{q-1} \hat{V}(\alpha + s, \beta + \gamma T_0) \cdot P(D(T_0|\gamma\lambda_0) = s) \\
&\quad + \int_0^{T_0} \hat{V}(\alpha + q, \beta + \gamma t) \cdot g(t|q, \gamma\lambda_0)\mathrm{d}t\Bigg] \\
&= E_{\lambda_0|\alpha,\beta}\Bigg[\sum_{s=0}^{q-1} \hat{V}(\alpha + s, \beta + \gamma T_0) \cdot \frac{(\gamma\lambda_0 T_0)^s e^{-\gamma\lambda_0 T_0}}{s!} \\
&\quad + \int_0^{T_0} \hat{V}(\alpha + q, \beta + \gamma t) \cdot \frac{(\gamma\lambda_0)^q t^{q-1} e^{-q\gamma\lambda_0 t}}{\Gamma(q)}\mathrm{d}t\Bigg]; \qquad \text{(B.1)}
\end{aligned}
$$

(ii) if $\lambda = \lambda_0$, $T = \gamma T_0$: $\hat{\alpha} = \alpha + s$, $\hat{\beta} = \beta + t$. Similarly, by conditioning on λ_0 and observation (s', t') we have

$$
\begin{aligned}
\Pi^T(q|\lambda_0; \gamma T_0; \alpha, \beta) &= E_{\lambda_0|\alpha,\beta} E_{(s',t')|\lambda_0}[\hat{V}(\alpha + s', \beta + t')] \\
&= E_{\lambda_0|\alpha,\beta}\Bigg[\sum_{s'=0}^{q-1} \hat{V}(\alpha + s', \beta + \gamma T_0) \cdot P(D(\gamma T_0|\lambda_0) = s') \\
&\quad + \int_0^{\gamma T_0} \hat{V}(\alpha + q, \beta + t') \cdot g(t'|q, \lambda_0)\mathrm{d}t\Bigg] \\
&= E_{\lambda_0|\alpha,\beta}\Bigg[\sum_{s'=0}^{q-1} \hat{V}(\alpha + s', \beta + \gamma T_0) \cdot \frac{(\gamma\lambda_0 T_0)^{s'} e^{-\gamma\lambda_0 T_0}}{s'!} \\
&\quad + \int_0^{\gamma T_0} \hat{V}(\alpha + q, \beta + t') \cdot \frac{\lambda_0^q (t')^{q-1} e^{-q\lambda_0 t'}}{\Gamma(q)}\mathrm{d}t'\Bigg].
\end{aligned}
$$

Let $s' = s$, $t' = \gamma t$, then

$$
\begin{aligned}
\Pi^T(q|\lambda_0; \gamma T_0; \alpha, \beta) &= E_{\lambda_0|\alpha,\beta}\Bigg[\sum_{s=0}^{q-1} \hat{V}(\alpha + s, \beta + \gamma T_0) \cdot \frac{(\gamma\lambda_0 T_0)^s e^{-\gamma\lambda_0 T_0}}{s!} \\
&\quad + \int_0^{T_0} \hat{V}(\alpha + q, \beta + \gamma t) \cdot \frac{\lambda_0^q (\gamma t)^{q-1} e^{-q\lambda_0 \gamma t}}{\Gamma(q)}\mathrm{d}\gamma t\Bigg] \qquad \text{(B.2)} \\
&= \Pi^T(q|\gamma\lambda_0; T_0; \alpha, \beta).
\end{aligned}
$$

The last equality follows from a comparison between (B.1) and (B.2).

□

The following lemma is a counterpart to Proposition 3.3 in which stores have identical inter-arrival time distributions but have different lengths of period 1.

Lemma B.3. *Suppose that stores have identical inter-arrival time distributions $\Psi(\tau|\theta)$. Then, $\Pi^T(\mathbf{q}|\Psi,\ldots,\Psi;T_1,\ldots,T_N;\pi) \leq \Pi^T(\mathbf{q}+\delta_i-\delta_j|\Psi,\ldots,\Psi;T_1,\ldots,T_N;\pi)$ for all $\mathbf{q}$ that has $q_i = 0$ and $q_j > 0$ for some $i < j$, $T_1 \geq \cdots \geq T_N$, and all π.*

Proof. Let $\mathcal{E} = \{X^T(\mathbf{q}+\delta_i-\delta_j); P_\theta\}$ and $\mathcal{F} = \{X^T(\mathbf{q}); Q_\theta\}$ be the experiments with timing information under allocation $\mathbf{q}+\delta_i-\delta_j$ and allocation $\mathbf{q}$, respectively. Consider the following transformation from $X^T(\mathbf{q}+\delta_i-\delta_j) = \{\mathbf{s},\mathbf{e},\vec{\tau}\}$ to $X' = \{\mathbf{s}',\mathbf{e}',\vec{\tau}'\}$ for every $\theta \in \Theta$:

(1) For $m \neq i,j$, let $s'_m = s_m$, $\vec{\tau}'_m = \vec{\tau}_m$, and $e'_m = e_m$;

(2) let $s'_i = 0$, $\vec{\tau}'_i = \varnothing$, $e'_i = 1$;

(3) if $s_j < q_j - 1$: let $s'_j = s_j, \vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(4) if $s_i = 0$, $s_j = q_j - 1$: let $s'_j = q$, $\vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(5) if $s_i = 1$, $s_j = q_j - 1$, and $\tau_i^1 > T_j - \sum_{k=1}^{q_j-1} \tau_j^k$: let $s'_j = q_j - 1$, $\vec{\tau}'_j = \vec{\tau}_j$, $e'_j = 0$;

(6) if $s_i = 1$, $s_j = q_j - 1$, and $\tau_i^1 \leq T_j - \sum_{k=1}^{q_j-1} \tau_j^k$: let $s'_j = q_j$, $\vec{\tau}'_j = \{\tau_j^1,\ldots,\tau_j^{q_j-1},\tau_i^1\}$, $e'_j = 1$.

It can be verified that X' and $X^T(\mathbf{q})$ have identical distributions. (Note that $T_i \geq T_j$ guarantees that $\sum_{k=1}^{q_j} \tau_j'^k = \sum_{k=1}^{q_j-1} \tau_j^k + \tau_i^1$ in (6) covers the entire $[0,T_j]$ interval.) Therefore, $\mathcal{E}$ is sufficient for $\mathcal{F}$. The lemma follows from an argument similar to that in the proof of Lemma 3.1. □

The following corollary applies Lemma B.3 to gamma-Poisson demand.

Corollary B.1. *Suppose that demand is gamma-Poisson and that stores have identical arrival rates λ, lengths $T_1 \geq \cdots \geq T_N \geq 0$ of period 1. Then, $\Pi^T(\mathbf{q}|\lambda,\ldots,\lambda;T_1,\ldots,T_N;\pi) \leq \Pi(\mathbf{q}+\delta_i-\delta_j|\lambda,\ldots,\lambda;T_1,\ldots,T_N;\pi)$ for all $\mathbf{q}$ that has $q_i = 0$ and $q_j > 0$ for some $i < j$, $\alpha > 0$, and $\beta > 0$.*

The following corollary applies Proposition 3.3 to gamma-Poisson demand.

Corollary B.2. *Suppose that demand is gamma-Poisson and that stores have identical length T of period 1 and relative demand coefficients $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N$), respectively. Then,* $\Pi^T(\mathbf{q}|\gamma_1\lambda, \ldots, \gamma_N\lambda; T, \ldots, T; \pi) \leq \Pi^T(\mathbf{q} + \delta_i - \delta_j|\gamma_1\lambda, \ldots, \gamma_N\lambda; T, \ldots, T; \pi)$ *for all* $\mathbf{q}$ *that has* $q_i = 0$ *and* $q_j > 0$ *for some* $i < j$, $\alpha > 0$, *and* $\beta > 0$.

Proof. Follows from Lemma B.2 and Proposition 3.3. □

The following lemma shows that under gamma-Poisson demand, the retailer prefers allocation $\mathbf{q} + \delta_i - \delta_j$ to allocation $\mathbf{q}$ that has $q_j = q_i + 1$ for some $i > j$, if stores have identical arrival rates but store i has a longer length of period 1 than store j does.

Lemma B.4. *Suppose that demand is gamma-Poisson and that stores have identical arrival rates λ and lengths $T_1, \ldots, T_N$ of period 1, where $T_i \geq T_j$ for some $i \neq j$. Then,* $\Pi^T(\mathbf{q}|\lambda, \ldots, \lambda; T_1, \ldots, T_N; \alpha, \beta) \leq \Pi^T(\mathbf{q} + \delta_i - \delta_j|\lambda, \ldots, \lambda; T_1, \ldots, T_N; \alpha, \beta)$ *for all* $\mathbf{q}$ *that has* $q_j = q_i + 1$, $\alpha > 0$, *and* $\beta > 0$.

Proof. We show the proof only for $N = 2$ and let $i = 1$ and $j = 2$ without loss of generality, i.e., $\Pi^T(q, q+1|\lambda, \lambda; T_1, T_2; \alpha, \beta) \leq \Pi^T(q+1, q|\lambda, \lambda; T_1, T_2; \alpha, \beta)$ for all $q = 0, 1, \ldots$, α, and β. The proof extends to the $N > 2$ cases by conditioning on the demand processes at stores other than i and j.

We write $T = T_2 \geq 0$ and $\Delta T = T_1 - T_2 \geq 0$. Consider a modification of the problem where period 1 at each store always *ends*, instead of starts, at the same time. In this case, after the modification, period 1 at store 1 covers time interval $[0, T + \Delta T]$, whereas period 1 at store 2 covers time interval $[\Delta T, T + \Delta T]$. We use a tilde as an identifier for corresponding notation in the modified problem. Note that the lengths of period 1 for both stores remain the same: $\tilde{T_1} = T + \Delta T = T_1$, $\tilde{T_2} = T = T_2$. Since the stores are independent conditional on λ and period 1 is purely for information learning purpose, such a modification in the start time of testing at store 2 does not affect the ex-ante expected profit, i.e.,

$$\tilde{\Pi}^T(q, q+1|\lambda, \lambda; T_1, T_2; \alpha, \beta) = \Pi^T(q, q+1|\lambda, \lambda; T_1, T_2; \alpha, \beta),$$
$$\tilde{\Pi}^T(q+1, q|\lambda, \lambda; T_1, T_2; \alpha, \beta) = \Pi^T(q+1, q|\lambda, \lambda; T_1, T_2; \alpha, \beta).$$

Let $\overline{T} = \min\{t_1^q, \Delta T\}$, where t_1^q is the time of q-th arrival at store 1. By conditioning on λ, $\overline{T}$, and observation (s,t) (sales quantity and effective selling duration) during $[0,\overline{T}]$, we obtain

$$\begin{aligned}
&\tilde{\Pi}^T(q, q+1|\lambda,\lambda;T+\Delta T, T;\alpha,\beta)\\
&\quad = E_{\lambda|\alpha,\beta}\Bigg[\sum_{s=0}^{q-1}\Pi^T(q-s, q+1|\lambda,\lambda;T,T;\alpha+s,\beta+\Delta T)P(D(\Delta T|\lambda)=s)\\
&\qquad\quad + \int_0^{\Delta T}\tilde{\Pi}^T(0, q+1|\lambda,\lambda;T+\Delta T-t, T;\alpha+q,\beta+t)g(t|q,\lambda)\mathrm{d}t,\Bigg],\\
&\tilde{\Pi}^T(q+1, q|\lambda,\lambda;T+\Delta T, T;\alpha,\beta)\\
&\quad = E_{\lambda|\alpha,\beta}\Bigg[\sum_{s=0}^{q-1}\Pi^T(q+1-s, q|\lambda,\lambda;T,T;\alpha+s,\beta+\Delta T)P(D(\Delta T|\lambda)=s)\\
&\qquad\quad + \int_0^{\Delta T}\tilde{\Pi}^T(1, q|\lambda,\lambda;T+\Delta T-t, T;\alpha+q,\beta+t)g(t|q,\lambda)\mathrm{d}t\Bigg],
\end{aligned}$$

where $g(t|q,\lambda)$ is the pdf of t_1^q conditional on λ. Note that

(i) $\Pi^T(q-s, q+1|\lambda,\lambda;T,T;\alpha+s,\beta+\Delta T) \le \Pi^T(q+1-s, q|\lambda,\lambda;T,T;\alpha+s,\beta+\Delta T)$ for all $s = 0,\dots,q-1$, following from Proposition 3.2;

(ii) $\tilde{\Pi}^T(0,q+1|\lambda,\lambda;T+\Delta T-t,T;\alpha+q,\beta+t) = \Pi^T(0,q+1|\lambda,\lambda;T+\Delta T-t,T;\alpha+q,\beta+t) \le \Pi^T(1,q|\lambda,\lambda;T+\Delta T-t,T;\alpha+q,\beta+t) = \tilde{\Pi}^T(1,q|\lambda,\lambda;T+\Delta T-t,T;\alpha+q,\beta+t)$, where the inequality follows from Corollary B.1.

As a result, $\tilde{\Pi}^T(q,q+1|\lambda,\lambda;T+\Delta T,T;\alpha,\beta) \le \tilde{\Pi}^T(q+1,q|\lambda,\lambda;T+\Delta T,T;\alpha,\beta)$, or, $\Pi^T(q,q+1|\lambda,\lambda;T_1,T_2;\alpha,\beta) \le \Pi^T(q+1,q|\lambda,\lambda;T_1,T_2;\alpha,\beta)$. □

The following corollary is a counterpart to Lemma B.4 in which stores have identical lengths of period 1 but different arrival rates.

Corollary B.3. *Suppose that demand is gamma-Poisson and that stores have identical lengths T of period 1, and relative demand coefficients $\gamma_1 \ge \gamma_2 \ge \cdots \ge \gamma_N$. Then, $\Pi^T(\mathbf{q}|\gamma_1\lambda,\dots,\gamma_N\lambda;T,\dots,T;\alpha,\beta) \le \Pi^T(\mathbf{q}+\delta_i-\delta_j|\gamma_1\lambda,\dots,\gamma_N\lambda;T,\dots,T;\alpha,\beta)$ for all $\mathbf{q}$ that has $q_j = q_i+1$ for some $i<j$, $\alpha>0$, and $\beta>0$.*

Proof. Follows from Lemma B.2 and Lemma B.4. $\square$

Proof of Proposition 3.4(a). We show the proof only for $N = 2$ and let $i = 1$ and $j = 2$ without loss of generality, i.e., $\Pi^T(q, q+n|\gamma_1\lambda, \gamma_2\lambda; T, T; \alpha, \beta) \leq \Pi^T(q+1, q+n-1|\gamma_1\lambda, \gamma_2\lambda; T, T; \alpha, \beta)$ for $q = 0, 1, \ldots$, $n = 1, 2, \ldots$, $\gamma_1 \geq \gamma_2$, $T > 0$, $\alpha > 0$, and $\beta > 0$. The proof extends to the $N > 2$ cases by conditioning on the demand processes at stores other than i and j.

The proof is by induction. The proposition holds for $q = 0$ and all $n \geq 1$ according to Corollary B.2. It also holds for all $q \geq 0$ and $n = 1$ according to Corollary B.3.

Assume that the proposition holds for some q and all $n \geq 1$. In addition, assume that it holds for $q+1$ and some $n \geq 1$. We show that it continues to hold for $q+1$ and $n+1$ by conditioning on the time $t \geq 0$ until the next arrival at either store, which is exponential with rate $(\gamma_1+\gamma_2)\lambda$. Once arrives, the next arrival occurs at store 1 with probability $\frac{\gamma_1}{\gamma_1+\gamma_2}$ and at store 2 with probability $\frac{\gamma_2}{\gamma_1+\gamma_2}$. We have

$$
\begin{aligned}
&\Pi^T(q+1, q+1+n+1|\gamma_1\lambda, \gamma_2\lambda; T, T; \alpha, \beta) \\
&\quad = E_{\lambda|\alpha,\beta} E_{t|\lambda}\Bigg[\Bigg\{\frac{\gamma_1}{\gamma_1+\gamma_2}\Pi^T(q, q+1+n+1|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t) \\
&\qquad + \frac{\gamma_2}{\gamma_1+\gamma_2}\Pi^T(q+1, q+1+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t)\Bigg\}\mathbb{1}\{t \leq T\} \\
&\qquad + \Pi^T(q+1, q+1+n+1|\gamma_1\lambda, \gamma_2\lambda; 0, 0; \alpha, \beta+(\gamma_1+\gamma_2)T)\mathbb{1}\{t > T\}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
&\Pi^T(q+2, q+1+n|\gamma_1\lambda, \gamma_2\lambda; T, T; \alpha, \beta) \\
&\quad = E_{\lambda|\alpha,\beta} E_{t|\lambda}\Bigg[\Bigg\{\frac{\gamma_1}{\gamma_1+\gamma_2}\Pi^T(q+1, q+1+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t) \\
&\qquad + \frac{\gamma_2}{\gamma_1+\gamma_2}\Pi^T(q+2, q+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t)\Bigg\}\mathbb{1}\{t \leq T\} \\
&\qquad + \Pi^T(q+2, q+1+n|\gamma_1\lambda, \gamma_2\lambda; 0, 0; \alpha, \beta+(\gamma_1+\gamma_2)T)\mathbb{1}\{t > T\}\Bigg].
\end{aligned}
$$

Note that

(i) $\Pi^T(q, q+1+n+1|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t) \leq \Pi^T(q+1, q+1+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t)$, following from the first induction assumption;

(ii) $\Pi^T(q+1, q+1+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t) \leq \Pi^T(q+2, q+n|\gamma_1\lambda, \gamma_2\lambda; T-t, T-t; \alpha+1, \beta+\gamma_1 t+\gamma_2 t)$, following from the second induction assumption;

(iii) and $\Pi^T(q+1, q+1+n+1|\gamma_1\lambda, \gamma_2\lambda; 0, 0; \alpha, \beta+(\gamma_1+\gamma_2)T) = \Pi^T(q+2, q+1+n|\gamma_1\lambda, \gamma_2\lambda; 0, 0; \alpha, \beta+(\gamma_1+\gamma_2)T)$ by definition.

Therefore, the induction is complete. □

Proof of Proposition 3.4(b). Immediately follows from part (a). □

B.6 Proof of Proposition 3.5

Proof of Proposition 3.5(a). We formulate the problem of maximizing the expected sales during period 1 as a dynamic program. Let $S_n(q)$ denote the sales at store n with test inventory level q. Then

$$E[S_n(q)] = \sum_{x=0}^{q} x\phi_n(x) + q \sum_{x=q+1}^{\infty} \phi_n(x),$$

where $\phi_n(\cdot)$ is the unconditioned pmf of demand at store n. We thus have

$$\Delta E[S_n(q)] = E[S_n(q+1)] - E[S_n(q)] = \overline{\Phi}_n(q), \tag{B.3}$$

where $\overline{\Phi}_n(q) = \sum_{x=q+1}^{\infty} \phi_n(x)$.

Let $S(\mathbf{q})$ be the expected total test sales under allocation $\mathbf{q} = (q_1, \ldots, q_N)$, i.e., $S(\mathbf{q}) = \sum_{n=1}^{N} S_n(q_n)$. Let $V_q(\mathbf{q})$ denote the maximum additional expected total test sales with q units of test inventory left to allocate given an allocation $\mathbf{q}$. The problem of allocating Q units of test inventory to maximize test sales can be formulated as a longest path problem with the following Bellman equations:

$$\begin{aligned} V_q(\mathbf{q}) &= \max_{n \in \{1,\ldots,N\}} \{\Delta E[S_n(q_n)] + V_{q-1}(\mathbf{q} + \delta_n)\} \\ &= \max_{n \in \{1,\ldots,N\}} \{\overline{\Phi}_n(q_n) + V_{q-1}(\mathbf{q} + \delta_n)\}, \end{aligned}$$

with $V_0(\mathbf{q}) = 0$ for all $\mathbf{q}$. The maximum expected total test sales with Q units of test inventory is given by $V_Q(\mathbf{0})$.

The proof of part (a) is by backward induction. For any allocation $\mathbf{q} = (q_1, \ldots, q_N)$, relabel the stores by $n_1, n_2, \ldots, n_N$, a permutation of $1, 2, \ldots, N$, such that $\overline{\Phi}_{n_1}(q_{n_1}) \geq \overline{\Phi}_{n_2}(q_{n_2}) \geq \cdots \geq \overline{\Phi}_{n_N}(q_{n_N})$. Note that

$$V_1(\mathbf{q}) = \max_{i=1,\ldots,N} \{\overline{\Phi}_{n_i}(q_{n_i}) + V_0(\mathbf{q} + \delta_{n_i})\} = \overline{\Phi}_{n_1}(q_{n_1}) + V_0(\mathbf{q} + \delta_{n_1}),$$

as $V_0(\mathbf{q}) = 0$ for all allocation $\mathbf{q}$. Suppose that for some $q \geq 1$ and allocation $\mathbf{q}$,

$$V_q(\mathbf{q}) = \max_{i=1,\ldots,N} \{\overline{\Phi}_{n_i}(q_{n_i}) + V_{q-1}(\mathbf{q} + \delta_{n_i})\} = \overline{\Phi}_{n_1}(q_{n_1}) + V_{q-1}(\mathbf{q} + \delta_{n_1})$$

where stores are relabeled according to $\mathbf{q}$ such that $\overline{\Phi}_{n_1}(q_{n_1}) \geq \overline{\Phi}_{n_2}(q_{n_2}) \geq \cdots \geq \overline{\Phi}_{n_N}(q_{n_N})$. Consider

$$V_{q+1}(\mathbf{q}) = \max_{i=1,\ldots,N} \{\overline{\Phi}_{n_i}(q_{n_i}) + V_q(\mathbf{q} + \delta_{n_i})\}.$$

Note that $\overline{\Phi}_{n_1}(q_{n_1}) \geq \overline{\Phi}_{n_i}(q_{n_i}) \geq \overline{\Phi}_{n_i}(q_{n_i} + 1)$ for $i = 2, \ldots, N$. Therefore,

$$V_q(\mathbf{q} + \delta_{n_i}) = \overline{\Phi}_{n_1}(q_{n_1}) + V_{q-1}(\mathbf{q} + \delta_{n_i} + \delta_{n_1})$$

for $i = 2, \ldots, N$ by the induction assumption. It follows that

$$\begin{aligned} \overline{\Phi}_{n_i}(q_{n_i}) + V_q(\mathbf{q} + \delta_{n_i}) &= \overline{\Phi}_{n_i}(q_{n_i}) + \overline{\Phi}_{n_1}(q_{n_1}) + V_{q-1}(\mathbf{q} + \delta_{n_i} + \delta_{n_1}) \\ &= \overline{\Phi}_{n_1}(q_{n_1}) + \overline{\Phi}_{n_i}(q_{n_i}) + V_{q-1}(\mathbf{q} + \delta_{n_1} + \delta_{n_i}) \\ &\leq \overline{\Phi}_{n_1}(q_{n_1}) + V_q(\mathbf{q} + \delta_{n_1}), \end{aligned}$$

where the inequality follows from the definition of $V_q(\mathbf{q} + \delta_{n_1})$. As a result, we have

$$V_{q+1}(\mathbf{q}) = \overline{\Phi}_{n_1}(q_{n_1}) + V_q(\mathbf{q} + \delta_{n_1}).$$

This completes the induction. □

Proof of Proposition 3.5(b). We prove part (b) by showing that the ordering $q_1 \geq q_2 \geq \ldots \geq q_N$ is preserved before each step of the Max-Sales algorithm. The ordering holds trivially at the beginning of the algorithm as $q_1 = \cdots = q_N = 0$. Suppose that the ordering holds before some step i. By definition $n^* = \min\{n : \overline{\Phi}_n(q_n) \leq \overline{\Phi}_m(q_m), \forall m \neq n\}$. The ordering is preserved before step $i+1$ if $n^* = 1$. When $n^* > 1$, assume that $q_{n^*-1} < q_{n^*} + 1$, then $\overline{\Phi}_{n^*}(q_{n^*}) > \overline{\Phi}_{n^*-1}(q_{n^*-1}) \geq \overline{\Phi}_{n^*}(q_{n^*-1}) \geq \overline{\Phi}_{n^*}(q_{n^*})$, where the first inequality follows from the definition of n^* and the last from the fact that $q_{n^*-1} \geq q_{n^*}$. This leads to contradiction. Hence we must have $q_{n^*-1} \geq q_{n^*} + 1$, i.e., the ordering holds before step $i+1$. □

B.7 Proof of Proposition 3.6

We show the proof only for $N = 2$. The proof extends to the $N > 2$ cases by conditioning on demand at other stores and noting the fact that stores are identical.

Let $\hat{V}(\hat{a}, \hat{S})$ denote the optimal expected profit in period 2 at a single store under some gamma posterior with parameters $(\hat{a}, \hat{S})$. The ex-ante expected profit of allocation (q_1, q_2) under a gamma prior with parameters (a, S) is given by

$$\begin{aligned}
\Pi^{NT}(q_1, q_2|a, S) &= \Pi(q_1, q_2) \\
&= \int_0^\infty \Bigg[\int_0^{q_1} \int_0^{q_2} \hat{V}(a+2, S + x_1^k + x_2^k) f(x_1|\theta) f(x_2|\theta) \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \int_{q_1}^\infty \int_0^{q_2} \hat{V}(a+1, S + q_1^k + x_2^k) f(x_1|\theta) f(x_2|\theta) \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \int_0^{q_1} \int_{q_2}^\infty \hat{V}(a+1, S + x_1^k + q_2^k) f(x_1|\theta) f(x_2|\theta) \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \int_{q_1}^\infty \int_{q_2}^\infty \hat{V}(a, S + q_1^k + q_2^k) f(x_1|\theta) f(x_2|\theta) \mathrm{d}x_1 \mathrm{d}x_2 \Bigg] \pi(\theta|a, S) \mathrm{d}\theta.
\end{aligned}$$

Following Bisi et al. (2011), we can write $\hat{V}(a, S) = \hat{S}^{\frac{1}{k}} \hat{v}(\hat{a})$, where $\hat{v}(\hat{a}) = \hat{V}(\hat{a}, 1)$. Taking the derivatives, we have

$$\begin{aligned}
\frac{\partial \Pi(q_1, q_2)}{\partial q_1} &= q_1^{k-1} \left[A(S + q_1^k)^{\frac{1}{k} - a - 1} + (B - A)(S + q_1^k + q_2^k)^{\frac{1}{k} - a - 1} \right], \\
\frac{\partial^2 \Pi(q_1, q_2)}{\partial q_1 \partial q_2} &= k \left(\frac{1}{k} - a - 1 \right) (B - A) q_1^{k-1} q_2^{k-1} (S + q_1^k + q_2^k)^{\frac{1}{k} - a - 2},
\end{aligned}$$

$$\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} = Aq_1^{k-2}(S+q_1^k)^{\frac{1}{k}-a-2}\left[(k-1)S - kaq_1^k\right]$$
$$+ (B-A)q_1^{k-2}(S+q_1^k+q_2^k)^{\frac{1}{k}-a-2}\left[(k-1)(S+q_2^k) - kaq_1^k\right], \qquad \text{(B.4)}$$

where A and B are constants given by $A = S^a ka\left[\frac{a+1}{a+1-1/k}\hat{v}(a+2) - \hat{v}(a+1)\right]$ and $B = S^a k\left(a - \frac{1}{k}\right)\left[\frac{a}{a-1/k}\hat{v}(a+1) - \hat{v}(a)\right]$ that have $B > A > 0$. We also have $\frac{\partial\Pi(q_1,q_2)}{\partial q_2} = \frac{\partial\Pi(q_2,q_1)}{\partial q_2}$ and $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_2^2} = \frac{\partial^2\Pi(q_2,q_1)}{\partial q_2^2}$ as $\Pi(q_1,q_2)$ is symmetric with respect to q_1 and q_2.

It can be verified that $\frac{\partial\Pi(q_1,q_2)}{\partial q_1} > 0$ for all $q_1 > 0, q_2 \geq 0$, and that $\frac{\partial\Pi(q_1,q_2)}{\partial q_2} > 0$ for all $q_1 \geq 0, q_2 > 0$. Therefore, it suffices to consider allocations that satisfy $q_1 + q_2 = Q$.

Proof of Proposition 3.6(a). When $0 < k \leq 1$, from (B.4) we have $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} < 0$ and $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} < 0$ for all $q_1 > 0, q_2 > 0$. Furthermore, we have $(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_2^2})(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2}) - (\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1\partial q_2})^2 > 0$ for all $q_1 > 0$, $q_2 > 0$. Hence, $\Pi(q_1,q_2)$ is jointly concave in q_1 and q_2 for $q_1 > 0$, $q_2 > 0$. Also, $\Pi(q_1,q_2)$ is continuous at points with $q_1 = 0$ and/or $q_2 = 0$. Therefore, allocation $\mathbf{q}^* = (q_1^*, q_2^*)$ with $q_1^* = q_2^* = Q/2$ maximizes $\Pi(q_1,q_2)$. □

Proof of Proposition 3.6(b). When $k > 1$,

(i) $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} > 0$ and $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} > 0$ for all $q_1 < Q_0, q_2 < Q_0$ where constant $Q_0 = \left[\frac{(k-1)S}{ka}\right]^{\frac{1}{k}}$. Also, $(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2})(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_2^2}) - (\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1\partial q_2})^2 > 0$ for all $q_1 < Q_0$, $q_2 < Q_0$. Hence, $\Pi(q_1,q_2)$ is jointly convex on $q_1 < Q_0$, $q_2 < Q_0$. As a result, for all $Q < Q_0$, $\Pi(q_1,q_2)$ is jointly convex on $\{(q_1,q_2) : q_1 + q_2 = Q\}$, and a single-store allocation $(Q,0)$ or $(Q,0)$ maximizes $\Pi(q_1,q_2)$.

(ii) $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2} < 0$ for all $q_1 > \bar{Q}(q_2)$ and $\frac{\partial^2\Pi(q_1,q_2)}{\partial q_2^2} < 0$ for all $q_2 > \bar{Q}(q_1)$ where function $\bar{Q}(q) = \left[\frac{(k-1)(S+q^k)}{ka}\right]^{\frac{1}{k}}$. Also, $(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1^2})(\frac{\partial^2\Pi(q_1,q_2)}{\partial q_2^2}) - (\frac{\partial^2\Pi(q_1,q_2)}{\partial q_1\partial q_2})^2 > 0$ for all $q_1 > \bar{Q}(q_2)$, $q_2 > \bar{Q}(q_1)$. Hence, for $q_1 > \bar{Q}(q_2)$, $q_2 > \bar{Q}(q_1)$, $\Pi(q_1,q_2)$ is jointly concave in q_1 and q_2, thus $(Q/2, Q/2)$ is a local maximal. Also, as $Q \to \infty$, $\Pi(Q/2, Q/2|a,S) \to \int_0^\infty\int_0^\infty\int_0^\infty \hat{V}(a+2, S+x_1^k+x_2^k)f(x_1|\theta)f(x_2|\theta)\mathrm{d}x_1\mathrm{d}x_2\pi(\theta)\mathrm{d}\theta \geq \Pi(q_1,q_2)$ for all $q_1 \geq 0$ and $q_2 \geq 0$.

□

APPENDIX C

Appendix for Chapter 4

C.1 Proof of Lemma 4.1

Proof of Proposition 4.1(a). Follows immediately by taking the first derivative of $u(\hat{q}, s, \theta)$ with respect to $\hat{q}$. $\square$

Proof of Proposition 4.1(b). Follows immediately by taking the first derivative of $u(\hat{q}, s, \theta)$ with respect to θ. $\square$

Proof of Proposition 4.1(c). $u(\hat{q}, s, \mu_\theta + \hat{q} - q) = \frac{\hat{q}(\sigma_\theta^2 + s\sigma_T^2) + \hat{q}\sigma_T^2}{\sigma_\theta^2 + (s+1)\sigma_T^2} = \hat{q}$. $\square$

C.2 Proof of Lemma 4.2

Proof. We prove by induction. Note that $o_1^*(\hat{q}, s) = 1$ as $F_\Theta(p - \hat{q})V_0(\hat{q}, s) + \int_{p-\hat{q}}^{\infty}[p + V_0(u(\hat{q}, s, \theta), s + 1)]\mathrm{d}F_\Theta(\theta) = p\overline{F}_\Theta(p - \hat{q}) > 0 = V_0(\hat{q}, s)$. Suppose that $o_t^*(\hat{q}, s) = 1$, i.e., $V_t(\hat{q}, s) = F_\Theta(p - \hat{q})V_{t-1}(\hat{q}, s) + \int_{p-\hat{q}}^{\infty}[p + V_{t-1}(u(\hat{q}, s, \theta), s + 1)]\mathrm{d}F_\Theta(\theta)$. Since $V_t(\hat{q}, s) \geq V_{t-1}(\hat{q}, s) \geq 0$ by definition, we have $F_\Theta(p-\hat{q})V_t(\hat{q}, s) + \int_{p-\hat{q}}^{\infty}[p + V_t(u(\hat{q}, s, \theta), s+1)]\mathrm{d}F_\Theta(\theta) \geq F_\Theta(p - \hat{q})V_{t-1}(\hat{q}, s) + \int_{p-\hat{q}}^{\infty}[p + V_{t-1}(u(\hat{q}, s, \theta), s + 1)]\mathrm{d}F_\Theta(\theta) = V_t(\hat{q}, s)$. The equality follows from the induction assumption. Therefore, $o_{t+1}^*(\hat{q}, s) = 1$, which completes the induction. $\square$

C.3 Proof of Proposition 4.1

Proof of Proposition 4.1(a). We prove by induction. The result trivially holds for $t = 0$. Suppose $V_t(\hat{q}_1, s, \theta) \leq V_t(\hat{q}_2, s, \theta)$ for some t, all $\hat{q}_1 < \hat{q}_2$, s, and θ. We show that $V_{t+1}(\hat{q}_1, s, \theta) \leq V_{t+1}(\hat{q}_2, s, \theta)$ for all $\hat{q}_1 < \hat{q}_2$, s, and θ. Consider the following regions of θ:

(i) $\theta < p - \hat{q}_2$. In this region, $V_{t+1}(\hat{q}_1, s, \theta) = \mathsf{E}_\Theta V_t(\hat{q}_1, s, \Theta) \leq \mathsf{E}_\Theta V_t(\hat{q}_2, s, \Theta) = V_{t+1}(\hat{q}_2, s, \theta)$.

(ii) $p - \hat{q}_2 \leq \theta < p - \hat{q}_1$. In this region, $V_{t+1}(\hat{q}_1, s, \theta) = \mathsf{E}_\Theta V_t(\hat{q}_1, s, \Theta) \leq \mathsf{E}_\Theta V_t(\hat{q}_2, s, \Theta) \leq V_{t+1}(\hat{q}_2, s, \theta)$.

(iii) $\theta \geq p - \hat{q}_1$. In this region,

$$\begin{aligned} V_{t+1}(\hat{q}_1, s, \theta) &= \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}_1, s, \theta), s+1, \Theta), \mathsf{E}_\Theta V_t(\hat{q}_1, s, \Theta)\} \\ &\leq \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}_2, s, \theta), s+1, \Theta), \mathsf{E}_\Theta V_t(\hat{q}_2, s, \Theta)\} \\ &= V_{t+1}(\hat{q}_2, s, \theta), \end{aligned}$$

where the inequality uses the fact that $u(\hat{q}_1, s, \theta) \leq u(\hat{q}_2, s, \theta)$ as a result of Lemma 4.1(a).

These together complete the induction. □

Proof of Proposition 4.1(b). The proof is by induction. The result trivially holds for $t = 0$. Suppose $V_t(\hat{q}, s, \theta_1) \leq V_t(\hat{q}, s, \theta_2)$ for some t, all $\theta_1 < \theta_2$, and all $\hat{q}$ and s. We show that $V_{t+1}(\hat{q}, s, \theta_1) \leq V_{t+1}(\hat{q}, s, \theta_2)$ for all $\theta_1 < \theta_2$ and all $\hat{q}$ and s. Consider the following regions of $\hat{q}$:

(i) $\theta_1 < \theta_2 < p - \hat{q}$. In this region, $V_{t+1}(\hat{q}, s, \theta_1) = \mathsf{E}_\Theta V_t(\hat{q}, s, \Theta) = V_{t+1}(\hat{q}, s, \theta_2)$.

(ii) $\theta_1 < p - \hat{q} \leq \theta_2$. In this region, $V_{t+1}(\hat{q}, s, \theta_1) = \mathsf{E}_\Theta V_t(\hat{q}, s, \Theta) \leq V_{t+1}(\hat{q}, s, \theta_2)$.

(iii) $p - \hat{q} \leq \theta_1 < \theta_2$. In this region,

$$\begin{aligned} V_{t+1}(\hat{q}, s, \theta_1) &= \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}, s+1, \theta_1), s, \Theta), \mathsf{E}_\Theta V_t(\hat{q}, s, \Theta)\} \\ &\leq \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}, s+1, \theta_2), s, \Theta), \mathsf{E}_\Theta V_t(\hat{q}, s, \Theta)\} \\ &= V_{t+1}(\hat{q}, s, \theta_2), \end{aligned}$$

where the inequality follows from Lemma 4.1(a).

These together complete the induction. □

Proof of Proposition 4.1(c). This is a direct corollary of Proposition 4.1(b). □

C.4 Proof of Proposition 4.3

Proof. The proof is by induction. The proposition holds trivially if $\theta_t^*(\hat{q}, s) = \theta_{t+1}^*(\hat{q}, s) = p - \hat{q}$. Suppose that for some $t > 1$, $\theta_{t-1}^*(\hat{q}, s) = p - \hat{q}$ and $\theta_t^*(\hat{q}, s) > p - \hat{q}$. We show that

$\theta_t^*(\hat{q}, s) \le \theta_{t+1}^*(\hat{q}, s)$, or, equivalently, $p + G_t(u(\hat{q}, s, \theta_t^*(\hat{q}, s)), s+1) \le G_t(\hat{q}, s)$:

$$
\begin{aligned}
G_t(\hat{q}, s) &= F_\Theta(\theta_t^*(\hat{q}, s)) G_{t-1}(\hat{q}, s) + p\overline{F}_\Theta(\theta_t^*(\hat{q}, s)) + \int_{\theta_t^*(\hat{q},s)}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1) \mathrm{d}F_\Theta(\theta) \\
&\ge F_\Theta(\theta_t^*(\hat{q}, s))[p + G_{t-1}(u(\hat{q}, s, \theta_t^*(\hat{q}, s)), s+1)] + p\overline{F}_\Theta(\theta_t^*(\hat{q}, s)) \\
&\qquad + \int_{\theta_t^*(\hat{q},s)}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1) \mathrm{d}F_\Theta(\theta) \\
&= p + F_\Theta(\theta_t^*(\hat{q}, s)) G_{t-1}(u(\hat{q}, s, \theta), s+1) + \int_{\theta_t^*(\hat{q},s)}^{\infty} G_{t-1}(u(\hat{q}, s, \theta), s+1) \mathrm{d}F_\Theta(\theta) \\
&\ge p + G_t(u(\hat{q}, s, \theta_t^*(\hat{q}, s)), s+1).
\end{aligned}
$$

This completes the induction. □

C.5 Proof of Proposition 4.4

Proof of Proposition 4.4(a). Follows from a similar argument in the proof of Proposition 4.1(a). □

Proof of Proposition 4.4(b). The proof is by induction. Suppose $V_t(\hat{q}, s, x, \theta) \le V_t(\hat{q}, s, x+1, \theta)$.

$$
\begin{aligned}
V_{t+1}(\hat{q}, s, x, \theta) &= \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}, s, \theta), s+1, x-1, \Theta), \mathsf{E}_\Theta V_t(\hat{q}, s, x, \Theta)\} \\
&\le \max\{p + \mathsf{E}_\Theta V_t(u(\hat{q}, s, \theta), s+1, x, \Theta), \mathsf{E}_\Theta V_t(\hat{q}, s, x+1, \Theta)\} \\
&= V_{t+1}(\hat{q}, s, x+1, \theta).
\end{aligned}
$$

This completes the induction. □

Proof of Proposition 4.4(c). Follows from a similar argument in the proof of Proposition 4.1(b).

□

Proof of Proposition 4.4(d). This is a direct corollary of Proposition 4.1(c). □

C.6 Proof of Proposition 4.6

Proof of Proposition 4.6(a). Follows from a similar argument in the proof of Proposition 4.3. □

Proof of Proposition 4.6(b). The proof is by induction. □

C.7 Bayesian Updating Accounting for Selection Biases

Let $\pi(q)$ denote consumers' prior belief on q. The posterior updated based on a review r is given by

$$\begin{aligned}\hat{\pi}(q|r) &= \frac{\pi(q)f_\Theta(r+\mu_\theta-q|r+\mu_\theta-q+\hat{q}-p\geq 0)}{\int_{-\infty}^{\infty}\pi(q)f_\Theta(r+\mu_\theta-q|r+\mu_\theta-q+\hat{q}-p\geq 0)\mathrm{d}\theta}\\ &= \frac{\pi(q)f_\Theta(r+\mu_\theta-q)\mathbf{1}_{q\leq r+\mu_\theta+\hat{q}-p}}{\int_{-\infty}^{\infty}\pi(t)f_\Theta(r+\mu_\theta-t)\mathbf{1}_{t\leq r+\mu_\theta+\hat{q}-p}\mathrm{d}t}.\end{aligned}$$

One can verify that the posterior $\hat{\pi}(q|r)$ is truncated Normal if $\pi(q)$ is Normal. Further updating requires keeping track of the entire review history and there is no finite dimension sufficient statistic.

BIBLIOGRAPHY

Agrawal, N. and Smith, S. A. (2009). Multi-location inventory models for retail supply chain management. In Agrawal, N. and Smith, S. A., editors, *Retail Supply Chain Management*, pages 207–235. Springer Science+Business Media, LLC.

Agrawal, N. and Smith, S. A. (2013). Optimal inventory management for a retail chain with diverse store demands. *European Journal of Oper. Res.*, 225(3):393–403.

Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.*, 31(9):1150–1160.

Azoury, K. S. and Miyaoka, J. (2009). Optimal policies and approximations for a bayesian linear regression inventory model. *Management Sci.*, 55(5):813–826.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Application.* Prentice-Hall, Englewood Cliffs, NJ.

Bensoussan, A., Çakanyıldırım, M., Minjárez-Sosa, J., Royal, A., and Sethi, S. (2008). Inventory problems with partially observed demands and lost sales. *J. Optim. Theory Appl.*, 136(3):321–340. 0022-3239.

Bensoussan, A., Çakanyıldırım, M., and Sethi, S. P. (2007). A multiperiod newsvendor problem with partially observed demand. *Math. Oper. Res.*, 32(2):322–344.

Bernstein, F., Kök, A. G., and Xie, L. (2015). Dynamic assortment customization with limited inventories. *Manufacturing & Service Oper. Management*. Forthcoming.

Besbes, O. and Scarsini, M. (2014). On information distortion in online ratings. Working Paper.

Besbes, O. and Zeevi, A. (2011). On the minimax complexity of pricing in a changing environment. *Oper. Res.*, 59(1):66–79.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026.

Bisi, A., Dada, M., and Tokdar, S. (2011). A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing & Service Oper. Management*, 13(4):525–533.

Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, volume 1, pages 93–102.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272.

Braden, D. J. and Freimer, M. (1991). Informational dynamics of censored observations. *Management Sci.*, 37(11):1390–1404.

Brown, D. B., Smith, J. E., and Sun, P. (2010). Information relaxations and duality in stochastic dynamic programs. *Oper. Res.*, 58(4-Part-1):785–801.

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

Chen, L. (2010). Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Oper. Res.*, 58(2):396–413.

Chen, L. and Mersereau, A. J. (2015). Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy. In Agrawal, N. and Smith, S. A., editors, *Retail Supply Chain Management*, pages 79–112. Springer Science+Business Media, New York.

Chen, L. and Plambeck, E. L. (2008). Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Oper. Management*, 10(2):236–256.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354.

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4):23–45.

Ding, X., Puterman, M. L., and Bisi, A. (2002). The censored newsvendor and the optimal acquisition of information. *Oper. Res.*, 50(3):517–527.

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4):1007–1016.

Elmaghraby, W. and Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Sci.*, 49(10):1287–1309.

Eppen, G. D. and Iyer, A. V. (1997). Backup agreements in fashion buying – the value of upstream flexibility. *Management Sci.*, 43(11):1469–1484.

Fisher, M. and Rajaram, K. (2000). Accurate retail testing of fashion merchandise: Methodology and application. *Marketing Sci.*, 19(3):266–278.

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

Frazier, P. I. and Powell, W. B. (2010). Paradoxes in learning and the marginal value of information. *Decision Analysis*, 7(4):378–403.

Gallien, J., Mersereau, A. J., Garro, A., Mora, A. D., and Vidal, M. N. (2015). Initial shipment decisions for new products at Zara.

Ginebra, J. (2007). On the measure of the information in a statistical experiment. *Bayesian Anal.*, 2(1):167–211.

Godes, D. and Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Sci.*, 31(3):448–473.

Golrezaei, N., Nazerzadeh, H., and Rusmevichientong, P. (2014). Real-time optimization of personalized assortments. *Management Sci.*, 60(6):1532–1551.

Graves, S. C. (1999). A single-item inventory model for a nonstationary demand process. *Manufacturing Service Oper. Management*, 1(1):50–61.

Harrison, J. M. and Sunar, N. (2014). Investment timing with incomplete information and multiple means of learning. Working Paper.

Hu, N., Zhang, J., and Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Commun. ACM*, 52(10):144–147.

Huh, W. T., Levi, R., Rusmevichientong, P., and Orlin, J. B. (2011). Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.*, 59(4):929–941.

Ifrach, B., Maglaras, C., and Scarsini, M. (2012). Monopoly pricing in the presence of social learning. Working Paper.

Ifrach, B., Maglaras, C., and Scarsini, M. (2015). Bayesian social learning from consumer reviews. Working Paper.

Iida, T. and Zipkin, P. H. (2006). Approximate solutions of a dynamic forecast-inventory model. *Management Sci.*, 8(4):407–425.

Jain, A., Rudi, N., and Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Oper. Res.*, 63(1):134–150.

Jing, B. (2011). Social learning and dynamic pricing of durable goods. *Marketing Sci.*, 30(5):851–865.

Karlin, S. (1960). Dynamic inventory policy with varying stochastic demands. *Management Sci.*, 6(3):231–258.

Karlin, S. and Rubin, H. (1956). Distributions possessing a monotone likelihood ratio. *Journal of the American Statistical Association*, 51(276):637–643.

Kök, A. G., Fisher, M. L., and Vaidyanathan, R. (2015). Assortment planning: Review of literature and industry practice. In Agrawal, N. and Smith, S. A., editors, *Retail Supply Chain Management*, volume 223 of *International Series in Operations Research & Management Science*, pages 175–236. Springer US.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(4):613–658.

Lariviere, M. A. and Porteus, E. L. (1999). Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.*, 45(3):346–363.

Li, X. and Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing*, 70(3):74–89.

Lovejoy, W. S. (1990). Myopic policies for some inventory models with uncertain demand distributions. *Management Sci.*, 36(6):724–738.

Lovejoy, W. S. (1992). Stopped myopic policies in some inventory models with generalized demand processes. *Management Sci.*, 38(5):688–707.

Lu, X., Song, J.-S., and Regan, A. (2006). Inventory planning with forecast updates: Approximate solutions and cost error bounds. *Oper. Res.*, 54(6):1079–1097.

Lu, X., Song, J.-S., and Zhu, K. (2008). Analysis of perishable-inventory systems with censored demand data. *Oper. Res.*, 56(4):1034–1038.

Moe, W. W. and Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Sci.*, 31(3):372–386.

Morton, T. E. and Pentico, D. W. (1995). The finite horizon nonstationary stochastic inventory problem: Near-myopic bounds, heuristics, testing. *Management Sci.*, 41(2):334–343.

Osadchiy, N., Gaur, V., and Seshadri, S. (2013). Sales forecasting with financial indicators and experts' input. *Production and Operations Management*, 22(5):1056–1076.

Papanastasiou, Y., Bakshi, N., and Savva, N. (2014). Scarcity strategies under quasi-baysian social learning. Working Paper.

Papanastasiou, Y. and Savva, N. (2014). Dynamic pricing in the presence of social learning. Working Paper.

Scarf, H. (1959). Bayes solutions of the statistical inventory problem. *The Annals of Mathematical Statistics*, 30(2):490–508.

Scarf, H. E. (1960). Some remarks on Bayes solutions to the inventory problem. *Naval Research Logistics Quarterly*, 7(4):591–596.

Sethi, S. P. and Cheng, F. (1997). Optimality of (s, S) policies in inventory models with Markovian demand. *Oper. Res.*, 45(6):931–939.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.

Song, J.-S. and Zipkin, P. (1993). Inventory control in a fluctuating demand environment. *Oper. Res.*, 41(2):351–370.

Song, J.-S. and Zipkin, P. H. (1996). Managing inventory with the prospect of obsolescence. *Oper. Res.*, 44(1):215–222.

Tang, C. S., Rajaram, K., Alptekinoglu, A., and Ou, J. (2004). The benefits of advance booking discount programs: Model and analysis. *Management Sci.*, 50(4):465–478.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2015). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, Boca Raton, FL.

Treharne, J. T. and Sox, C. R. (2002). Adaptive inventory control for nonstationary demand and partial information. *Management Sci.*, 48(5):607–624.

Wang, T., Atasu, A., and Kurtulus, M. (2012). A multiordering newsvendor model with dynamic forecast evolution. *Manufacturing & Service Oper. Management*, 14(3):472–484.

Yu, M., Debo, L., and Kapuscinski, R. (2013). Strategic waiting for consumer-generated quality information: Dynamic pricing of new experience goods. Working Paper.

Zipkin, P. H. (2000). *Foundations of Inventory Management*. McGraw Hill, New York.