Decision Making with Coupled Learning: Applications in Inventory Management and Auctions

Juan Manuel Chaneton

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Decision Making with Coupled Learning: Applications in Inventory Management and Auctions

Juan Manuel Chaneton

Operational decisions can be complicated by the presence of uncertainty. In many cases, there exist means to reduce uncertainty, though these may come at a cost. Decision makers then face the dilemma of acting based on current, incomplete information versus investing in trying to minimize uncertainty. Understanding the impact of this trade-off on decisions and performance is the central topic of this thesis.

When attempting to construct probabilistic models based on data, operational decisions often affect the amount and quality of data that is collected. This introduces an exploration-exploitation trade-off between decisions and information collection. Much of the literature has sought to understand how operational decisions should be modified to incorporate this trade-off. While studying two well-known operational problems, we ask an even more basic question: does the exploration-exploitation trade-off matter in the first place? In the first two parts of this thesis we focus on this question in the context of the newsvendor problem and sequential auctions with incomplete private information.

We first analyze the well-studied stationary multi-period newsvendor problem, in which a retailer sells perishable items and unmet demand is lost and unobserved. This latter limitation, referred to as demand censoring, is what introduces the exploration-exploitation trade-off in this problem. We focus on two questions: *i.*) what is the value of accounting for the exploration-exploitation trade-off; and, *ii.*) what is the cost imposed by having access only to sales data as opposed to underlying demand samples? Quite remarkably, we show that, for a broad family of tractable cases, there is essentially no exploration-exploitation trade-off; i.e., there is almost no value of accounting for the impact of decisions

on information collection. Moreover, we establish that losses due to demand censoring (as compared to having full access to demand samples) are limited, but these are of higher order than those due to ignoring the exploration-exploitation trade-off. In other words, efforts aimed at improving information collection concerning lost sales are more valuable than analytic or computational efforts to pin down the optimal policy in the presence of censoring.

In the second part of this thesis we examine the problem of an agent bidding on a sequence of repeated auctions for an item. The agent does not fully know his own valuation of the object and he can only collect information if he wins an auction. This coupling introduces the exploration-exploitation trade-off in this problem. We study the value of accounting for information collection on decisions and find that: *i.*) in general the exploration-exploitation trade-off cannot be ignored (that is, in some cases ignoring exploration can substantially affect rewards), but *ii.*) for a broad class of instances, ignoring exploration can indeed produce nearly optimal results. We characterize this class through a set of conditions on the problem primitives, and we demonstrate with examples that these are satisfied for common settings found in the literature.

In the third part of this thesis we study the impact of uncertainty in the context of inventory record inaccuracies in inventory management systems. Record inaccuracies, mismatches between physical and recorded inventory, are frequently encountered in practice and can markedly affect revenues. Most of the literature is devoted to analyzing the cost-benefit relationship between investing in means to reduce inaccuracies and accounting for them in operational decisions. We focus on the less explored approach of using available data to reduce the uncertainty in inventory. In practice, collecting Point Of Sale (POS) data is substantially simpler than collecting stock information. We propose a model in which inventory is regarded as a virtually unobservable quantity and POS data is used to infer its state over time. Additionally, our method also works as an effective estimator of censored demand in the presence of inaccurate records. We test our methodology with extensive numerical experiments based on both simulated and actual retailing data. The results show that it is remarkably effective in inferring unobservable past statistics and predicting future stock status, even in the presence of severe data misspecifications.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would first like to thank my advisors, Professors Omar Besbes, Ciamac Moallemi and Garrett van Ryzin for their time, guidance and patience over these years. The invaluable tools and experience gained from my PhD are due in great part to their efforts. For that I am wholeheartedly grateful.

I would also like to thank the faculty at the DRO department, whom I got to know well and who always offered help and advice when needed. I would specially like to thank Prof. Costis Maglaras for his help in the last stages of my PhD, both in academic and non-academic matters. The DRO and PhD staff holds a special place in my heart: Clara, Joyce, Cristina, Winnie, Elizabeth and Dan and their countless efforts to make our lives easier.

I would also like to thank Prof. Gustavo Vulcano for his help and guidance since my early undergrad years. He is one of the main reasons I decided, and had the opportunity to, pursue a doctorate degree in this department. For that I am deeply thankful.

The PhD would have not been the same without the help and company of my dear friends: Daniela, Charles, Nikhil, Daniel and Peter. Their friendship will be one of my most cherished rewards of my years in graduate school.

Finally, I would like to express my most devoted gratitude to Jennifer, without whom I would hardly be standing where I am today. For her love and support during the greater part of these PhD years I will always be grateful.

A mis padres, Carlos y María Sol Chaneton

Chapter 1

Introduction

Operational decisions can be challenging in the presence of uncertainty. A common source of uncertainty is the fact that the outcome of decisions usually depends on quantities and phenomena not yet realized; that is, it depends on the *future*. For example, at the time of deciding how many items to order, a retailer may not be able to accurately predict the demand for a product. Uncertainty can also stem from the lack of comprehensive knowledge of a system: the retailer might not know exactly how many units are currently available in stock for sale. Though different in nature, both forms of uncertainty affect the outcome of decisions. In many cases, the main difference lies in the extent to which they can be reduced.

In broad terms, *learning* can be defined as the act of reducing uncertainty about a system. Learning can be performed in various ways depending on the setting and objectives. A widely adopted strategy in the Operations world is to decouple the processes of learning and decision making: the decision maker first uses historical data and statistical methods to reduce the uncertainty about the future and then makes decisions based on the information gathered. If facing not a single but a sequence of decisions, it becomes necessary to incorporate new information as it becomes available. In many settings, decisions themselves affect the collection of information: if the retailer's orders are too low and he runs out of stock, he will most likely be unable to observe current demand and thus compromise his ability to predict future demand. This gives rise to *dynamic learning* problems. The main challenge is to incorporate the learning component into the decision making process, effectively balancing learning and performance.

In this thesis we explore the effects of uncertainty and learning on well-known operational problems. In Chapters 2 and 3 we focus on quantifying the effects of the coupling between decisions and information collection on two well-known problems: the sequential newsvendor problem and sequential auctions with incomplete private information. In Chapter 4 we focus on how to improve estimation and forecasting in the presence of stock uncertainty in an inventory management system.

1.1 Quantifying exploration-exploitation

In many practical situations decision makers have to act based on limited information and this uncertainty can be reduced by collecting feedback from the environment. Often decisions themselves influence the amount and quality of information collected and thus not only immediate profits are affected by current actions, but also future rewards through the information collection process.

These type of problems are commonly known as *dynamic learning problems.* The decision maker faces the following question: should he optimize his decision using the current information, or should he make a sub-optimal choice with regard to instantaneous rewards that might improve his knowledge of the system, and lead to better performance in the future? This is known as the *exploration-exploitation* trade-off. The key concept is that decisions not only affect rewards, but also information that is collected to inform future decisions.

The coupling between decision and learning gained the attention of researches and practitioners early, as they realized it can deeply affect performance. It has been widely studied in many areas, including Economics, Operations Research and Statistics, among others. One of the most general models that captures this coupling is known as the Multiarmed Bandit problem (MAB).

Multiarmed Bandit Problem As described in [Bubeck and Cesa-Bianchi, 2012], "a MAB problem is a sequential allocation problem defined by a set of actions. At each time step, a unit resource is allocated to an action and some observable payoff is obtained. The goal is to maximize the total payoff obtained in a sequence of allocations [...] Bandit problems are basic instances of sequential decision making with limited information and naturally address the fundamental tradeoff between exploration and exploitation in sequential experiments ". MAB have been widely studied in the literature under several formalizations, and different metrics have been proposed to evaluate policy performance: regret analysis, where performance is compared to an oracle that possesses full information, worst-case or adversarial, where performance is evaluated on a worst-case scenario basis, and Markovian, where reward distributions evolve stochastically following a Markov chain. Many traditional sequential decision problems can be framed in a MAB setting. Examples include inventory and supply chain management with incomplete distributional knowledge, optimal sequential experimentation, and optimal ad allocation in online settings, among many others.

A common approach in the Operations Research literature is to model incomplete knowledge in a Bayesian setting. In a Bayesian framework unknown quantities, such as unknown distribution parameters, are assumed to follow a known prior distribution, which is updated as new information becomes available. Bayesian sequential learning problems are one of the most notable examples of Markovian bandits. In these cases, the MAB problem can be formulated as a Markov Decision problem, where the state of the system is given by the current prior distribution over the unknown quantities (and potentially additional state variables), and the system evolves as new information is gathered according to Bayes rule.

Bandit problems provide a very general framework to study the interaction between learning and decision making. However, it is rarely possible to draw general conclusions, and studies usually rely heavily on the particular structure of each problem. For example, in one of the classical MAB formulations the number of actions is finite and the (observable) reward of each action is independent from the rest. In such case the outcome of a particular decision gives no information on the alternatives and hence acting greedily with respect to current information can lead to poor performance. In other settings choosing one action can inform about the reward associated with other actions. These are instances of bandit problems with *correlated* rewards. In these cases the value of explicitly exploring becomes less clear, as learning might take place implicitly.

The study of this *implicit* learning when rewards are heavily correlated is of central focus in the second and third chapters of this thesis. We study two dynamic learning problems from very different applications that nevertheless share the common feature that actions inform about other actions' rewards. In chapter 2 we study the sequential newsvendor problem with partially known demand distribution. In this problem every order results in a partial observation of demand and hence improves the knowledge of the cost associated with *all* possible orders. In chapter 3 we study the problem of an agent with incomplete private information bidding on a sequence of repeated auctions. The agent only collects information about his valuation if he wins the auction and hence his current bid can inform about the value associated with all future bids.

Much of the literature addressing these and other problems seeks to obtain insight in understanding how the exploration-exploitation trade-off should impact operational decisions. We ask an even more basic question: does the trade-off matter in the first place? In other words, to what extent is it necessary to explicitly explore? Quite remarkably we find that in the problems we analyze, the value of exploring is not significant.

1.1.1 Newsvendor problem

In many business settings, inventory management can be challenging due to uncertainty about underlying demand. A decision maker must build probabilistic models for future demand which are estimated based on past data. In practical operational situations, however, demand information is rarely accessible. Often, only access to some form of distorted demand is possible, and this distortion is usually impacted by prior operational decisions. An important example is *demand censoring*: in many retail environments, demand is only observable up to the level at which it can be fulfilled. As a result, firms most often only have access to sales data as opposed to the true, underlying demand.

In Chapter 2 we consider a multi-period newsvendor problem, where a retailer sells perishable items and has to decide how many units to order for each time period. Unmet demand is censored. We assume that the retailer does not have a complete knowledge of the underlying distribution, but rather learns about it as sales realizations are observed. We consider a Bayesian formulation where one may summarize the state of the system through the current belief over the demand distribution.

As we discuss in the literature review below, it is clear from existing studies that ignoring the exploration-exploitation trade-off is suboptimal and that one should order more than the myopic solution under fairly general conditions. We focus on more basic issues with regard to demand censoring. In particular, we consider two fundamental questions. First, what is the value of accounting for the exploration-exploitation trade-off? We evaluate this by introducing a novel metric quantifying the losses stemming from ignoring this trade-off, a quantity we will refer to as the *myopic optimality gap*. Second, what are the losses due to censoring? In other words, what is the cost of having access to sales data as opposed to demand samples? This latter quantity we refer to as the *cost of censoring*. Quite remarkably we find that there is essentially no exploration-exploitation trade-off, that is there is almost no value of accounting for the impact of ordering decisions on information collection. Moreover, we establish that losses due to demand censoring (as compared to having full access to demand samples) are limited, but these are of higher order than those due to ignoring the exploration-exploitation trade-off.

Related literature. A Bayesian approach to inventory management with observable demand was first proposed by [Scarf, 1959], leading to a formulation as a Markov Decision Process (MDP), with the state given by the current knowledge about the demand distribution. In such a case, decisions do not affect collection of information, and there is no exploration-exploitation trade-off; the main issue is to deal with the dimensionality of the state space. In its most general formulation, the problem has an infinite dimensional state; it can, however, be written as a finite dimensional MDP if the demand distribution is chosen from a parametric conjugate family. See also [Azoury, 1985] for another early reference on the topic.

When demand observations are censored, decisions and information collection are coupled, and the problem becomes more challenging. Most of the literature on inventory control with Bayesian learning has focused on understanding the censored case. The main difficulty is that Bayesian updates lose their conjugate property in the presence of censoring for most distribution families, and hence the problem becomes intractable. Many studies have therefore focused on showing structural properties of the optimal policies. [Ding *et al.*, 2002] (see also the complementary note of [Lu *et al.*, 2005b]) study the censored newsvendor problem for general distributions and show that, under certain conditions, the optimal order quantity is greater than the quantity one would order to minimize current single period cost, also referred to as the myopic quantity. This result was then generalized by [Bensoussan *et al.*, 2009a] for arbitrary parametric distributions. The main point is that the exploration-exploitation trade-off leads to collecting more information for future periods. This is, broadly speaking, as far as the general case has been characterized.

There exists, however one well known family of distributions that preserves the conjugate property even under censoring. First introduced by [Braden and Freimer, 1991], the *newsvendor* distributions have been frequently used as a benchmark to study the problem in a more tractable fashion. [Lariviere and Porteus, 1999] study the non perishable inventory problem with newsvendor demands, and they find that, if demands follow a *Weibull* distribution - a particular case of newsvendor- a state space reduction technique can be applied and the problem can be solved by backward induction. [Bisi *et al.*, 2011] analyze the same case for perishable and non perishable inventory and they develop a specific recursive formula for the optimal solution in the perishable case. They also show that the Weibull is the only case in which such space reduction can be applied.

The effects of demand censoring have been studied in a wide range of environments, focusing on different aspects of the problem. In particular, many extensions of the base newsvendor case have been analyzed recently. [Dai and Jerath, 2013] study the impact of inventory restrictions and demand censoring in sales force compensation contracts, and [Heese and Swaminathan, 2010] analyze a similar problem when the firm has complete control over the sales effort. [Chen and Han, 2013] analyze the effects of learning and censoring on the supply side. [Jain *et al.*, 2014] study the newsvendor problem with censored demands when sales transaction timing is also available, while [Mersereau, 2014] focus on the conjunction of demand censoring with inventory record inaccuracies. Most of these papers rely on strong model assumptions in order to circumvent the complexity of the analysis. From the computational point of view, heuristics have been developed in the literature to overcome the intractability of the problem for general prior/demand combinations. [Chen, 2010] and [Lu *et al.*, 2005a] propose heuristics and provide bounds on the optimal quantity.

All of the above papers focus on modeling uncertainty in a Bayesian framework. There is also a stream of literature that analyzes the design of policies under different informational assumptions, when there is no parameterization of the demand distribution. In such cases, performance is measured through regret, e.g., the gap between the performance of a policy and that of an oracle with knowledge of the demand distribution. [Kunnumkal and Topaloglu, 2008] and [Huh and Rusmevichientong, 2009] are example of such approaches. In that line of work, [Besbes and Muharremoglu, 2013] is related in spirit to the present study as it measures the impact of demand censoring on performance. However, it does so under different informational assumptions and through the lens of minimax regret, and only measures performance through the growth rate of regret.

Finally, we refer to the recent review of [Chen and Mersereau, 2013] for a broader overview of the literature on demand censoring.

1.1.2 Sequential auctions with incomplete private information

Auctions are a widely adopted allocation mechanism in many economic exchange settings, and in particular they play a central role in online advertising. In many cases, for example when an ad has had few or no impressions, advertisers may be uncertain about the value of the ad and therefore choosing the best bid represents a challenging task: a high bid has the risk of incurring too much cost on a potentially low value ad, whereas a low bid reduces the chances of the ad being shown and thus collecting information about its value. As a result, the present bid affects not only immediate rewards, but also the information collected for future decisions.

In Chapter 3 we study the exploration-exploitation trade-off in the context of sequential auctions, where the bidder has incomplete information about his own valuation of the auctioned item. We consider a sequential problem where identical copies of a good (e.g. graphical ad impressions in a web page or ads in a sponsored search service) are sold through a sequence of auctions over time. We focus on the point of view of a particular advertiser or *agent*, who is uncertain about the value of the ad and has to decide what bid to place in each auction. If the agent wins the auction he observes and collects a reward associated with the display of the ad. We consider a Bayesian formulation where the state of the system is summarized through the current belief over the ad reward distribution.

The agent's decision is not only based on his own reward beliefs, but might also depend on his beliefs about the bids of the competing players. We will assume that the agent has access to a *competing bid distribution*, which summarizes the information about the bids of the competing players and further that this distribution is stationary over time. [Iyer *et al.*, 2014] work under similar assumptions as they study a setting, commonly found in online marketplaces, in which bidders are chosen from a large pool of homogeneous agents.[1] They regard the competing bid distribution as an endogenous feature of the market that can be derived in equilibrium. From a particular agent's perspective, we regard the competing bid as exogenous and given. Data to construct or approximate such distribution is commonly available for advertisers in online ad marketplaces, either through past historical bid information or through services that compute the probability of winning for a particular bid.

In the present work, we first show that in this problem bidding myopically is suboptimal, and the optimal policy always prescribes a *higher* bid than the myopic one, a result previously shown in similar settings. With this in mind, we proceed to identify a broad range of settings where the actual gain from deviating from the myopic bid is limited. In other words, in many cases there is little value in exploring. In particular, Theorem 3.1 and Corollary 3.1 in Section 3.2 show that if the competing bid and the prior agent beliefs satisfy mild conditions, the losses stemming from ignoring exploration grow only logarithmically with the length of the time horizon and are reduced as the initial uncertainty decreases at a rate proportional to the prior variance. Our results are developed in a general framework that includes both first-price and second price auctions, and well known families of distributions.

Related Literature: There is a substantial body of literature that studies online auction markets from a game theoretic perspective, both in static (see for example [Edelman *et al.*, 2007] and [Varian, 2007]) and dynamic settings (see [Babaioff *et al.*, 2009] or [Devanur and Kakade, 2009]). A large portion of the literature focuses on studying market equilibria or designing efficient mechanisms from the auctioneer point of view. We focus our work on a

[1] In that case a particular agent is unlikely to participates in an auction with the same competitor twice and hence, since agents are homogeneous, the competing bid is roughly stationary in the short term.

particular agent's perspective, who faces a sequence of auctions and needs to collect information to assess the value of the object he is bidding for. Our model formulation resembles the single agent problem studied in [Iyer *et al.*, 2014], who focus on market behavior and equilibria rather than the extent to which exploration is necessary for a particular agent. [Kanoria and Nazerzadeh, 2014] study how to choose reserve prices in repeated second price auctions with learning. [Li *et al.*, 2010] show that the exploration-exploitation trade-off cannot be ignored from a mechanism design perspective, as the second price auctions fail to be incentive compatible when the mechanism does not perform exploration. [Hummel and McAfee, 2014] study a similar model to ours, from the auctioneers perspective, and they reach the conclusion that the required degree of exploration is low. Their results, however, mostly study the dependency on the initial uncertainty in the system, and not the dependency on the time horizon (equivalently the discount rate, as they work with infinite time horizon), a more common metric in the dynamic learning literature.

From a sequential learning perspective our work is also closely related to the extensive literature on dynamic learning problems and strategic experimentation. We refer the reader to the very good surveys [Bubeck and Cesa-Bianchi, 2012], in the machine learning literature, and [Bergemann and Valimaki, 2006] in the economics literature.

1.1.3 Contributions

We next summarize our main contributions of chapters 2 and 3.

1. We introduce a metric, the myopic optimality gap, by which to measure the extent of the exploration-exploitation trade-off. In particular, we do so by comparing the performance of an optimal policy to that of a myopic policy that chooses the action that minimizes current single-period rewards given all the accumulated information.

2. We develop upper bounds for the myopic optimality gap (also bounding the cost of censoring in the newsvendor case). We show that under broad conditions the gap becomes negligible for extreme cases of the problem primitives, and characterize the exact rate of convergence to zero. In particular, we study the the dependence on the time horizon and the intial uncertainty in the system. For the newsvendor case we

further establish the result as the service level increases to 1 or decreases to zero: quite notably, even when the inventory manager selects a low service level and demand is very often censored, the myopic optimality gap vanishes.

3. In the sequential auctions problem, we characterize a broad class of problem primitives, with minimal distributional assumptions, where the bound on the myopic optimality gap holds. We illustrate its applications with examples in different settings that include first and second price auctions.

4. In the newsvendor case, we take advantage of (widely adopted) distributional assumptions and evaluate the myopic optimality gap in an *exact* manner for a grid of **all** input parameters: time horizon, parameter uncertainty, service level, Weibull parameter. We show that it is negligible for virtually all parameter combinations, often on the order of a fraction of a percentage point. In other words, *there is almost no exploration-exploitation trade-off*.

   Similarly, we evaluate the cost of censoring in an exact manner. We show that while it is in general of the order of 10%, it is always of a higher order than the myopic optimality gap. This suggests that effort to improve information collection about lost sales is more valuable than analytic or computational effort to pin down the optimal policy in the presence of censoring.

In addition to the above, we note that the finding regarding the exploration-exploitation trade-off is also relevant from a computational point of view. Implementing myopic policies is always relatively straightforward while the computation of optimal policies may be intractable. In other words, our findings suggest that the myopic policy should be not only computationally attractive but may also be a viable candidate heuristic in cases where the optimal policy is intractable.

1.2 Estimation and forecasting with partially observable data

Using past data to reduce uncertainty is a fundamental tool for operations researchers and practitioners. A common complication is that collected data is often subject to different types of distortion and as a result only partial information is available.

Incomplete data can be attributed to disruptions in data collection and storing, or it can mean the that part of the data is simply unobservable. Demand censoring, studied in Chapter 2, is an example of the latter. There are numerous others examples in which data, due to technical difficulties or more fundamental issues, is not fully available for use: competitors' data in commerce, non-logged user data in online applications, customer preferences in marketing, are only a few examples.

When data is incomplete estimation and forecasting becomes more difficult. A large body of literature in areas such as Statistics, Operations Research, Economics and Marketing, among others, has been devoted to finding methodologies to circumvent this issue. In the supply chain and inventory management literature, demand censoring has particularly gained the attention of researchers.

In Chapter 4 we focus on settings in which not only demand is censored, but inventory information itself is incomplete or inaccurate. Inventory record inaccuracies, mismatches between physical and recorded inventory, are commonly found in practice as a result of several unobservable phenomena. We study the problems of estimation and forecasting in the presence of these inaccuracies, with inventory regarded as a partially observable quantity.

1.2.1 Estimating inaccurate inventory with transactional data

Accurate stock tracking is a crucial capability in inventory management. In many retail settings, however, stock information can get distorted to produce inaccurate inventory records. Record inaccuracies can affect normal supply chain operations and lead to poor revenue performance, and therefore means to reduce them or mitigate their impact are very valuable for retailers.

This problem is important in many retail settings. For example, in very small fast moving consumer goods shops, such as small grocery stores, the retailer commonly has very little or no resources to devote to inventory tracking and management, and hence inventory is virtually an unobservable quantity. Even in large organized retail settings, where retailers do have electronic inventory records, empirical studies show that inventory record inaccuracies are common (see for example [DeHoratius and Raman, 2008]). This problem is not only relevant to retailers, but also to suppliers. Suppliers often cannot access their customers' inventory data and, due to the complex nature of supply chains, they cannot rely on their own delivery data to provide them with an accurate picture of inventory on their customers.

The main issue with effectively tracking inventory is that there exist phenomena affecting stock that are unobservable to inventory tracking systems. In extreme cases where there is no such tracking system, even demand and supply can be regarded as unobservable inventory processes. In less rudimentary cases, other sources of inaccuracy produce mismatches between physical and recorded inventory: theft, spoilage and misplacing, among others, are typical causes. Aware of this problems, retailers must choose between investing on means to reduce this mismatch, or accounting for it in their operational decisions. As we discuss in the literature review below, a good portion of the literature has been devoted to analyzing the cost-benefit relationship between this two options, and in particular to study how accounting for record inaccuracies should affect operational decisions. We focus our work on the less explored approach of using available data to reduce the uncertainty in inventory. In particular, we rely on the fact that, typically, collecting sales transaction data is substantially simpler than collecting physical inventory data. Our main research objective is then to develop an estimation methodology that uses only point of sale (POS) data, and potentially some reduced amount of inventory information, to estimate and predict stock levels.

We propose a model where the inventory evolution is a hidden process that periodically reveals information about its state, through transactional data. The observable information is given only by POS transactions, which describe which products where purchased in the store and when. Our proposed methodology is motivated by settings where inventory information feedback is very infrequent or simply unavailable; this is the case of micro-retailers who do not possess the human resources to track their inventory but, with the help of information technology tools, can effectively record sales of their products. Though inspired in such extreme cases, our method is indeed applicable more generally: beside POS data our model can incorporate direct stock information, i.e. inventory inspections, into the estimation procedure and hence it is readily applicable in a broader range of settings.

We approach the problem by defining a *Hidden Markov Model* (HMM). Inventory evolution is modeled as an unobservable Markov Chain that emits observable information through POS transactions. The transitions of the chain reflect the changes in stock as it is subject to the drivers of inventory evolution: demand, supply and potentially unobserbable effects.

Fitting the parameters of the chain can be interpreted as fitting parametrized models for the drivers of inventory evolution. Of particular practical interest is the demand process. Accurately predicting demand is one the main priorities of retailers, and hence a subject of great interest for researches. In particular, a lot of attention has been given to the estimation of *censored* demand, that is, demand that is not observable when the item is out of stock. As discussed in the literature review, there has been numerous efforts to estimate demand in the presence of censoring, but most models make use of reliable stock information. In this sense, our methodology can also be regarded as an effort to estimate censored demand with inaccurate stock information. It is, to our best knowledge, one of the first efforts to achieve this goal.

Related literature. This work follows in a stream of inventory management literature that looks at inventory record inaccuracy and techniques for addressing it. Bensoussan et al. have a series papers looking at optimal control of partially observed inventory systems. In particular, [Bensoussan *et al.*, 2007] look at a system where demand is not observed and inventory is known to be either zero or greater than zero, based on employees zero-balance walks. [DeHoratius *et al.*, 2008] look at a setting where an invisible demand process affects inventory levels but is not observed by the retailer. They create a distribution of physical inventory using Bayesian updating and they propose an audit triggering mechanism based on the value of perfect information versus the cost to audit. Similarly, in the manufacturing setting, [Kök and Shang, 2007] determine the optimal dynamic order-up-to policy and physical inspection schedule, where the order quantity grows as the inter-inspection time increases, to account for increased uncertainty in the inventory level. [Chen, 2013] study the case where inventory level is known exactly, but the system can randomly switch to a faulty state where all sales are blocked, while [Mersereau, 2014] look at the case where records are inaccurate, the demand distribution is not fully known, and it is updated as new information becomes available.

Most of the aforementioned papers are mainly focused on finding optimal policies, or good heuristics, to handle the tradeoff between performing, potentially costly, inventory inspections and making decisions with inaccurate information. We focus on the *estimation* or *forecasting* phase, a necessary step for most of the optimization models in the literature. In this sense, our work is closely related to the literature on censored demand estimation. There has been numerous efforts in the Statistics literature to develop estimation methodologies under censored information, and some have focused on demand censoring in particular. This is the case, for example, of [Braden and Freimer, 1991] who propose the widely adopted Newsvendor family of distributions as a candidate for demand distributions with censored observations. On a related work [Agrawal and Smith, 1996] propose the Negative Binomial distribution as a good candidate for estimation using sales data. A good amount of research has also been devoted to study the problem of estimating product level demand and substitution patterns, in the presence of stockouts. See for example [Musalem *et al.*, 2010], [Vulcano *et al.*, 2012] or [Conlon and Mortimer, 2012]. In these cases, the estimation relies deeply on information about the availability of related products, with perhaps the absence of stockout timing information. Accurate inventory data is hence critical to obtain reliable results. Alternatively, we focus on the problem of estimating stock status and inventory evolution parameters with a minimal amount of information, given by POS data.

Because we model the stock evolution with a HMM, our work is also related to the vast literature on these models. We refer the reader to [Scott, 2002] for a survey on Bayesian methods with HMM models. Finally we refer to [Chen and Mersereau, 2013] for a comprehensive review of the literature on demand censoring and record inaccuracies.

1.2.2 Contributions

We propose a methodology to estimate inaccurate stock levels, based solely on POS transnational data and, if available, inventory inspection observations. With minimal data, our method is able to estimate past stock levels and predict future ones, particularly effectively recognizing in-stock and out-of-stock periods. In this sense, the method can be regarded as an *unsupervised learning* methodology for unonbservable inventory.

In addition, our method represents a novel tool to perform demand estimation and forecasting when demand is censored by inventory, and records are inaccurate. Our approach can effectively forecast demand based solely on POS data without relying on stock information.

With extensive numerical experiments, based on both simulated and actual retailing data, we show that our methodology is remarkably effective in inferring unobservable past statistics and predicting future stock status, even in the presence of severe data misspecifications.

Chapter 2

Dynamic learning in the newsvendor problem

2.1 Problem Formulation

We consider a multi-period newsvendor problem with a finite time horizon consisting of T periods. We assume zero lead-time. At the beginning of each period, the retailer may order units, demand is then realized and is fulfilled to the extent possible. If the decision maker runs out of stock, demand beyond the initial inventory level is lost. On the other hand, we assume that inventory is perishable, i.e., if items remain, these are discarded.

We assume that the demands are independent and identically distributed, belonging to a family of distributions parameterized by an unknown parameter θ, which takes values in the parameter set $\Theta \subset \mathbb{R}$. Given parameter θ, the demand distribution has a probability density function denoted by $f(\cdot|\theta)$ and a cumulative distribution function denoted by $F(\cdot|\theta)$. Following a Bayesian approach, we assume that the decision maker has a prior distribution over θ, with a density denoted by $\pi(\cdot)$.

The retailer incurs a cost $h > 0$ for each unit of unsold product and a penalty cost $p > 0$ for each unit of unmet demand. In other words, the single period cost given stocking decision $y \geq 0$ and realized demand $D \geq 0$ is

$$L(y, D) := h(y - D)^+ + p(D - y)^+.$$

Let $r := p/(h+p)$ denote the *critical ratio*, also referred to as the *service level* . Without loss of generality, we assume that $h = 1$ and parameterize the cost structure through $r \in (0, 1)$, i.e.,

$$L(y, D) := (y - D)^+ + \frac{r}{1-r}(D - y)^+.$$

The retailer can only observe sales, rather than full demand realizations. In period t, the observed sales are given by $\min\{D_t, y_t\}$, where D_t is the demand realized and y_t the ordered quantity. We denote by $\mathcal{F}_t$ the filtration generated by the censored demand process, that is,

$$\mathcal{F}_t := \sigma\big(\min\{D_1, y_1\}, y_1, \ldots, \min\{D_{t-1}, y_{t-1}\}, y_{t-1}\big).$$

Note that the information collected about the demand up to period t is impacted by the past stocking decisions $y_1, y_2, \ldots, y_{t-1}$.

We assume that $\mathbb{E}[D_1] < \infty$, which implies that $\mathbb{E}[L(y, D_t)] < \infty$, for all $t = 1, \ldots, T$ and any $y \geq 0$. We denote by $\mathcal{P}^c$ the set of non-anticipatory policies with respect to the censored information system, that is, policies for which the decision prescribed in period t is $\mathcal{F}_t$-measurable for all $t \geq 1$. The objective is to minimize the cumulative expected cost and the optimal value is given by

$$V_T^* = \inf_{\mu \in \mathcal{P}^c} \sum_{t=1}^{T} \mathbb{E}^\mu \left[L(y_t, D_t)\right],$$

where the expectation is taken assuming that decisions are made according to the policy μ.

Since the items are perishable, decisions across periods are only coupled through the information collected about the unknown parameter θ of the demand distribution. When deciding on an order quantity in period t, the decision maker has to balance the impact of this decision on the current single-period cost with the impact on future costs that stems from the information collection process. This leads to the *exploration-exploitation* tradeoff.

2.1.1 Quantities of Interest

Our analysis of the newsvendor problem centers on several quantities of interest.

Exploration-exploitation trade-off. Our first goal is to quantify the exploration-exploitation trade-off. As mentioned in the introduction, it has been shown under general conditions that it is optimal to "order more". That is, in a multi-period problem with censored observations, a decision maker will want to stock at a level higher than would be optimal to minimize the current single-period cost. Intuitively, it is desirable to explore by ordering more so as to reduce the likelihood of censoring and collect more information about the demand distribution for use in future periods. However, precisely pinning down the optimal order quantity is, in general, difficult. As a result, the exact form of what type of exploration ought to be conducted, and the value associated with it, are not known. To quantify the exploration-exploitation trade-off, we will isolate the value of exploring (i.e., ordering more) by comparing the performance of an optimal policy to that of a policy that sequentially minimizes current single-period expected costs, fully ignoring the impact of decisions on information collection. More formally, the latter *myopic* (i.e., full exploitation) policy prescribes in period t the order size

$$y_t^m := \underset{y \geq 0}{\operatorname{argmin}} \ \mathbb{E}\left[L(y, D_t) \middle| \mathcal{F}_t\right],$$

The associated cumulative cost is then defined as

$$V_T^m = \sum_{t=1}^{T} \mathbb{E}\left[L(y_t^m, D_t)\right].$$

The myopic policy is perhaps the simplest heuristic to consider as an alternative to the optimal policy, and leads to a natural upper bound to the optimal cost,

$$V_T^* \leq V_T^m. \tag{2.1}$$

We are interested in qualifying the relative gap between V_T^m and V_T^*, that is, the relative additional cost incurred by ignoring the exploration-exploitation trade-off and applying a full exploitation policy. We refer to this as the *myopic optimality gap* (MOG), formally defined as

$$\mathsf{MOG} := \frac{V_T^m - V_T^*}{V_T^*}.$$

The cost of censoring. Our second goal is to quantify the impact of censoring on the performance compared to what would be possible if demand samples were fully observable. Isolating the latter impact yields an estimate of the value associated with collecting information on lost sales. If demand is observable, the collection of information is now unaffected by the policy used by the decision maker. We denote by

$$\mathcal{F}_t^o := \sigma(D_1, y_1, \ldots, D_{t-1}, y_{t-1})$$

the filtration associated with the full observation process. Let $\mathcal{P}^o$ be the set of admissible policies for the full observation problem, that is, policies for which prescribed decisions at period t are $\mathcal{F}_t^o$-measurable. The optimal cumulative cost for the uncensored problem is given by

$$V_T^o = \inf_{\mu \in \mathcal{P}^o} \sum_{t=1}^{T} \mathbb{E}^\mu \left[L(y_t, D_t) \right].$$

In the uncensored case, the information collected about the demand distribution is independent of past decisions and there is no exploration-exploitation trade-off. Hence, in this case, the problem is easy to solve: the optimal decision at time t is to minimize the current single-period cost. Moreover, clearly $\mathcal{P}^c \subseteq \mathcal{P}^o$, and it follows that V_T^o is a lower bound on V_T^*, that is,

$$V_T^o \leq V_T^*. \tag{2.2}$$

In order measure the impact of censoring on performance, we introduce the *Cost of Censoring* (COC), defined as the relative difference between the optimal costs in the censored and uncensored systems, that is,

$$\mathsf{COC} := \frac{V_T^* - V_T^o}{V_T^o}.$$

A common upper bound. While we will study the MOG and the COC explicitly, we will also develop parametric bounds on those quantities through a common upper bound. We define the *myopic cost of censoring* (MCC) as

$$\mathsf{MCC} := \frac{V_T^m - V_T^o}{V_T^o}.$$

It can be interpreted as the relative increase in cost stemming from censoring, when a myopic policy is applied. Note that (2.1)–(2.2) imply that

$$\mathsf{MOG} = \frac{V_T^m - V_T^*}{V_T^*} \leq \frac{V_T^m - V_T^o}{V_T^o}, \qquad \mathsf{COC} = \frac{V_T^* - V_T^o}{V_T^o} \leq \frac{V_T^m - V_T^o}{V_T^o}.$$

That is, the MCC provides an upper bound on both the MOG and the COC.

2.1.2 Demand Distributions and Dynamic Programming Formulation

Beliefs about the demand distribution change with time, as new (potentially censored) demand realizations are observed. At period t, the current knowledge about θ can be summarized by the current prior distribution density $\pi_t(\theta) := \pi(\theta|\mathcal{F}_t)$. The current prior is updated at every period, following Bayes' rule. In particular, if, in period t, the order quantity is y and observed sales are given by z, the Bayesian update rule can be separated into two cases,

$$\pi_{t+1}(\theta|z,\pi_t) = \begin{cases} \pi^o_{t+1}(\theta|z,\pi_t) := \dfrac{f(z|\theta)\pi_t(\theta)}{\displaystyle\int_\Theta f(z|\theta)\pi_t(\theta)\,d\theta} & \text{if } z < y, \\ \pi^c_{t+1}(\theta|y,\pi_t) := \dfrac{\overline{F}(y|\theta)\pi_t(\theta)}{\displaystyle\int_\Theta \overline{F}(y|\theta)\pi_t(\theta)\,d\theta} & \text{if } z = y, \end{cases}$$

for $\theta > 0$. Here, $\overline{F}(y|\theta) := 1 - F(y|\theta)$.

The cost functions and the relative performance gaps were defined in the general setting with arbitrary demand distributions. As noted earlier in the introduction, such generality makes dynamic optimization and any type of sensitivity analysis intractable. In line with the existing literature, we will narrow down our analysis to a subfamily of demand distributions.

Newsvendor distributions. One general class of distributions that preserves the conjugate property even under censoring are the so-called *newsvendor distributions*. A random variable is said to belong to the newsvendor family if its cumulative distribution function is given by

$$F(z|\theta) := 1 - e^{-\theta d(z)}, \quad \text{for all } z > 0, \tag{2.3}$$

where $d\colon \mathbb{R}_{++} \to \mathbb{R}_{++}$ is a differentiable, nondecreasing, and unbounded function. The parameter space is defined to be $\Theta := \mathbb{R}_{++}$ and the prior distribution of θ is assumed to be a Gamma distribution with shape parameter $S \in \mathbb{R}_{++}$ and rate parameter $a \in \mathbb{R}_{++}$. Formally, the prior distribution on θ has a density given by

$$\pi(\theta|a,S) := \frac{S^a\theta^{a-1}e^{-S\theta}}{\Gamma(a)}, \quad \text{for all } \theta > 0.$$

It has been shown citepBraden91 that this family of distributions preserves its structure even under censored observations, that is, if $\pi_t(\cdot)$ is a Gamma distribution, then so are both

$\pi^o_{t+1}(\cdot|z,\pi_t)$ and $\pi^c_{t+1}(\cdot|y,\pi_t)$. In particular, given a demand level z and order quantity y, the update rules for Gamma distribution hyperparameters (a,S) are given by

$$a_{t+1} = a_t + \mathbb{I}_{\{z<y\}}, \qquad S_{t+1} = S_t + d(z \wedge y), \tag{2.4}$$

where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function and $\cdot \wedge \cdot$ denotes the minimum. Equation (2.4) offers a natural interpretation of the hyperparameters (a,S): the scale hyperparameter S accounts for the total quantity of *sales* observed in the system, while the shape hyperparameter a counts the number of fully observed demand realizations. The shape hyperparameter plays a central role in our analysis as it measures the level of information about the unknown demand parameter θ: the coefficient of variation of the the demand parameter θ given hyperparameters (a,S) is

$$\mathrm{CV}(\theta|a,S) = 1/\sqrt{a}. \tag{2.5}$$

Hence, higher values of a indicate less uncertainty regarding the value of θ.

Assumptions. For tractability, our analysis focuses on *Weibull distributions*, i.e., newsvendor distributions with

$$d(z) := z^\ell, \quad \text{for } \ell > 0.$$

For the Weibull family, the Bayesian update (2.4) is written at

$$a_{t+1} = a_t + \mathbb{I}_{\{z<y\}}, \qquad S_{t+1} = S_t + (z \wedge y)^\ell.$$

The *predictive distribution* (the distribution of demand conditional on the current belief (a,S) for the parameter θ) has density

$$m(z|a,S) = \frac{aS^a \ell z^{\ell-1}}{(S+z^\ell)^{a+1}}, \tag{2.6}$$

and cumulative distribution

$$M(z|a,S) = 1 - \frac{S^a}{(S+z^\ell)^a}, \tag{2.7}$$

for $z > 0$. The predictive distribution is integrable (i.e., the demand has finite expectation) if and only if $a\ell > 1$, and therefore we will assume that this condition is satisfied throughout the chapter. Note that it suffices to verify that the initial prior distribution of θ satisfies

$a\ell > 1$, since the update rule (2.4) then guarantees that this will continue to hold at all future times.

Finite dimensional dynamic programs. Given current hyperparameters (a, S) we denote by $V_T^o(a, S)$, $V_T^m(a, S)$, and $V_T^*(a, S)$ the optimal cost function under uncensored demand, the myopic cost function under censored demand, and the optimal cost function under censored demand, respectively. Similarly, we define $C(a, S)$ to be the optimal single-period cost given current hyperparameters (a, S), that is,

$$C(a, S) := \min_{y \geq 0} \mathbb{E}\big[L(y, D)\big]. \tag{2.8}$$

In equation (2.8) the expectation is taken with respect to the pair (D, θ), that follows a newsvendor distribution with parameters (a, S). Unless otherwise stated, this will be the convention for expectation terms in the remainder of the chapter.

With this notation in mind, we can decompose the cost functions of interest using standard dynamic programming backward induction.[1] To begin, the optimal cost function under censored demand must satisfy the Bellman equation

$$\begin{aligned} V_T^*(a, S) \quad = \quad \min_{y \geq 0} \; & \Big\{ \mathbb{E}\left[L(y, D)\right] + \mathbb{E}\left[\mathbb{I}_{\{D < y\}} V_{T-1}^*\big(a + 1, S + D^\ell\big) \right] \\ & \quad + \mathbb{P}\left\{D \geq y\right\} V_{T-1}^*\big(a, S + y^\ell\big) \Big\}, \end{aligned}$$

for all (a, S) and all $T \geq 1$, with the terminal condition $V_0^*(a, S) = 0$, for all (a, S). Similarly, the myopic cost function under censored demand must satisfy

$$V_T^m(a, S) = C(a, S) + \mathbb{E}\left[\mathbb{I}_{\{D < y^m\}} V_{T-1}^m\big(a + 1, S + D^\ell\big) \right] + \mathbb{P}\left\{D \geq y^m\right\} V_{T-1}^m\big(a, S + (y^m)^\ell\big), \tag{2.9}$$

for all (a, S) and all $T \geq 1$, with the terminal condition $V_0^m(a, S) = 0$, for all (a, S). In (2.9), the myopic decision y^m is defined to be a solution to (2.8), i.e., $y^m \in \operatorname{argmin}_{y \geq 0} \mathbb{E}\big[L(y, D)\big]$. Finally, the optimal cost function when demand samples are observable must satisfy

$$V_T^o(a, S) = C(a, S) + \mathbb{E}\left[V_{T-1}^o\big(a + 1, S + D^\ell\big) \right], \tag{2.10}$$

for all (a, S) and all $T \geq 1$, with the terminal condition $V_0^o(a, S) = 0$, for all (a, S). Here, we have used the fact that the optimal policy in this case is myopic.

[1] The existence of an optimal policy in this setting can be shown by applying Proposition 3.4 of citetbertsekas1978.

2.2 Parametric Bounds and Structural Analysis

As highlighted in Section 2.1.1, the MCC serves as an upper bound for both the MOG and the COC. In this section, we focus on deriving bounds on the MCC. In doing so, we are able to jointly consider questions relating to the exploration-exploitation trade-off and the cost-of-censoring and derive insights on the impact of problem parameters on these quantities. From an analytical perspective, this approach circumvents the necessity to determine the optimal solution to the dynamic program associated with the censored demand problem. Instead, our analysis in this section relies on the performance of myopic policies, which are more amenable to analysis.

2.2.1 Upper Bounds

We aim to develop an upper bound on $V_T^m(a,S) - V_T^o(a,S)$. Rewriting the recursion for the myopic policy in the censored case (Eq. 2.9) in a way that parallels the recursion for the observable demand case (Eq. 2.10), one obtains, for all (a,S) and $T \geq 1$,

$$V_T^m(a,S) = C(a,S) + \mathbb{E}\left[V_{T-1}^m\big(a+1, S+D^\ell\big)\right] + \Gamma_{T-1}(a,S), \tag{2.11}$$

where

$$\Gamma_{T-1}(a,S) := (1-r)V_{T-1}^m\big(a, S+y^\ell\big) - \mathbb{E}\left[V_{T-1}^m\big(a+1, S+D^\ell\big)\mathbb{I}_{\{D\geq y\}}\right].$$

Combining (2.10) and (2.11), one has that

$$V_T^m(a,S) - V_T^o(a,S) = \mathbb{E}_D\left[V_{T-1}^m(a+1, S+D^\ell) - V_{T-1}^o(a+1, S+D^\ell)\right] + \Gamma_{T-1}(a,S). \tag{2.12}$$

The correction term $\Gamma_{T-1}(a,S)$ may be interpreted as the additional cost incurred over the next period due to the presence of censoring.

As a first step towards bounding the performance difference $V_T^m(a,S) - V_T^o(a,S)$, we bound $\Gamma_{T-1}(a,S)$ as follows.

Lemma 2.1. *Suppose that the demand distribution is Weibull. For all (a,S) with $a\ell > 1$, and all $T \geq 1$,*

$$\Gamma_T(a,S) \leq \sum_{k=1}^{T}(1-r)^k\left\{C\big(a, S(1-r)^{-k/a}\big) - C_{T-k+1}^o\big(a, S(1-r)^{-k/a}\big)\right\},$$

where $C_t^o(a, S)$ is the future expected single-period cost, t periods from now, when demands are uncensored, i.e.,

$$C_t^o(a, S) := \mathbb{E}\left[C\big(a+t, S+D_1^\ell+\ldots+D_t^\ell\big)\right]. \tag{2.13}$$

Notably, Lemma 2.1 offers a bound on $\Gamma_T(a, S)$ in which all terms depend only on future costs in the *uncensored* setting. We further note that Proposition 2.1 holds for *any* newsvendor distribution, and not just the Weibull case. The proof, which can be found in the appendix, is written in general terms. We next use Proposition 2.1 recursively to provide an explicit bound on $V_T^m(a, S) - V_T^o(a, S)$ by exploiting the Weibull structure[2].

Theorem 2.1. *Suppose that the demand distribution is Weibull. Then, for any $\ell > 0, T \geq 1, r \in (0,1), a > 0, S > 0$ with $a\ell > 1$,*

$$V_T^m(a, S) - V_T^o(a, S) \leq S^{1/\ell}\left[Q(a, r, l)\right]^{1/\ell}\frac{\lambda - \lambda^{T+1}}{1-\lambda}\left[\log\left(1+\frac{T}{a-1/\ell}\right)+\frac{1}{a-1/l}\right], \tag{2.14}$$

where $\lambda := (1-r)^{1-\frac{1}{a\ell}}$ and $Q(a, r, \ell)$ is a function depending only on a, r and ℓ, such that $Q(a, r, l) = O(1/a)$ as $a \to \infty$, when r and ℓ are fixed.

The theorem provides a bound on the (absolute) myopic cost of censoring as a function of the problem parameters: the time horizon T, the "uncertainty" parameter a and the service level r. We first provide a high level overview of the proof of the result and then analyze in detail the parametric dependence.

The proof of the result is based on two steps. We first develop an intermediate bound that connects the (absolute) myopic cost of censoring to the difference between the observable demand cost and the expected total cost when the demand parameter θ is known: $V_{T-1}^o(a, S) - (T-1)C_\infty^o(a, S)$, where

$$C_\infty^o(a, S) := \mathbb{E}\left[\min_{y\geq 0}\ \mathbb{E}\left[L(y, D)\middle|\theta\right]\right]$$

is the expected optimal single-period cost assuming θ is known.[3] The latter difference can be interpreted as the increase in cost stemming from the fact that θ is unknown, in

[2] In what follows, given functions $f(\cdot)$ and $g(\cdot) > 0$ we say that $f(x) = O\big(g(x)\big)$ as $x \to a$ if $\limsup_{x\to a} |f(x)|/g(x) < \infty$.

[3] Comparing with the definition of $C_T^o(a, S)$ in (2.13), note that $C_\infty^o(a, S) = \lim_{T\to\infty} C_T^o(a, S)$.

an uncensored environment. In a second step, we bound the latter difference based on the problem parameters by analyzing information accumulation and its implications on performance in the observable demand case.

Parametric dependence. Using Theorem 2.1, one can characterize the behavior of the MCC, as a function of the various problem parameters when one approaches the extremes of the parameter region.

Proposition 2.1 (Time horizon dependence)**.** *Suppose that the demand distribution is Weibull. Then, For any $\ell > 0, r \in (0,1), a > 0, S > 0$ with $a\ell > 1$,*

$$\frac{V_T^m(a,S) - V_T^o(a,S)}{V_T^o(a,S)} = O(T^{-1}\log T) \qquad \textit{as} \qquad T \to \infty.$$

$$\frac{V_T^m(a,S) - V_T^o(a,S)}{V_T^o(a,S)} = 0 \qquad \textit{when} \qquad T = 1.$$

The result captures the dependence of the MCC on the time horizon T. The fact that the MCC is equal to zero when $T = 1$ simply stems from the fact that, in a one period problem, the censored and uncensored problems are identical, and the optimal decision is myopic. For large time horizons, the MCC vanishes at rate $O(T^{-1}\log T)$. This reflects that: (i) over time, in the censored environment, a myopic decision maker will learn the true demand parameter θ, as in the uncensored environment[4]; (ii) the cumulative effects of deviations from a myopic solution in the censored case cannot affect performance significantly, at least not more than $O(T^{-1}\log T)$ in relative terms. (i) reflects the fact that learning takes place in both observable and censored environments, and provides some initial insights on the COC. (ii) suggests that the MOG might not be significant, at least in both the extreme cases when $T = 1$ and $T \to \infty$.

Proposition 2.2 (Prior information dependence)**.** *Suppose that the demand distribution is Weibull. Then for any $\ell > 0, r \in (0,1), S > 0, T \geq 1$,*

$$\frac{V_T^m(a,S) - V_T^o(a,S)}{V_T^o(a,S)} = O(1/a) \qquad \textit{as} \qquad a \to \infty.$$

[4]Related to this fact is the work of [Bensoussan *et al.*, 2009b], who study the asymptotic properties of the myopic order quantity y_t^m as an estimator of the optimal order quantity y_t^*, for the case of exponentially distributed demands.

In other words, the MCC converges to zero at a fast rate as $a \to \infty$. This captures the behavior of the MCC in a regime in which there is very little prior uncertainty about the demand distribution parameter θ. In this case, the presence of censoring does not affect performance significantly, since there is little additional information to be captured.

Proposition 2.3 (Service level dependence)**.** *Suppose that the demand distribution is Weibull. Then for any $\ell > 0, a > 0, S > 0, T \geq 1$ with $a\ell > 1$,*

$$\frac{V_T^m(a,S) - V_T^o(a,S)}{V_T^o(a,S)} = O\left((1-r)^{1-1/a\ell}\right) \qquad as \qquad r \to 1^-,$$

$$\frac{V_T^m(a,S) - V_T^o(a,S)}{V_T^o(a,S)} = O\left(r^{1/\ell}\right) \qquad as \qquad r \to 0^+,$$

The result gives a characterization of the asymptotic properties of the MCC as the service level r becomes close to 0 or 1, that is, as the holding cost becomes arbitrarily large with respect to the penalty cost and vice versa.

If the penalty cost is large (i.e., r is close to 1), the myopic order quantity will be high (corresponding the r-fractile of the predictive distribution) as the decision-maker attempts to mitigate the large penalties associated with stockouts. This implies that the myopic policy will often observe full demand realizations, and one expects that the presence of censoring should not impact performance significantly.

What is more surprising and remarkable, however, is that even when r is close to zero, and hence the myopic quantity leads the decision-maker to be censored very often, the performance implications of being myopic are still very limited; the MCC shrinks to zero at a rate $O(r^{1/\ell})$. This stems from the fact that the holding cost is very high and every unit of unconsumed inventory becomes very expensive, making exploration (or "ordering more" than myopic) very expensive.

As a corollary of the results above, both the myopic optimality gap and the cost of censoring become negligible at the boundaries of the input parameters of the problem, namely when $T = 1$, $T \to \infty$, $a \to \infty$, $r \to 0$, and $r \to 1$. In Section 2.3, we quantify more finely the MOG and the COC, as well as their relative magnitudes, by using an exact analysis to compute these values for various ranges of the parameters of the problem.

2.2.2 Lower Bound

To complement Theorem 2.1, we next provide a lower bound on the MCC for the case of exponential demand ($\ell = 1$).

Theorem 2.2. *Suppose demands are exponential (i.e., $\ell = 1$) and $a > 1$. Then, for any $T \geq 2$, $S > 0$ and $r \in (0,1)$,*

$$V_T^m(a,S) - V_T^o(a,S) \geq (1-r)\log^2(1-r)\frac{S}{a-1}\left[\log\left(1+\frac{T-2}{a}\right)+\frac{1}{a}-\frac{T-1}{a+T-1}\right]. \quad (2.15)$$

Note that the structure of the lower bound obtained in Theorem 2.2 is similar to the upper bound given in Theorem 2.1. Indeed, by specializing to the $\ell = 1$ case in (2.14), Theorem 2.1 implies that

$$V_T^m(a,S) - V_T^o(a,S) \;\leq\; K(r,1)\,\frac{\lambda - \lambda^{T+1}}{1-\lambda}\,S\,Q(a-1)\,\left[\log\left(1+\frac{T}{a-1}\right)+\frac{1}{a-1}\right].$$

In particular, the logarithmic dependence with respect to the time horizon T is the best possible dependence one could obtain for $V_T^m(a,S) - V_T^o(a,S)$.

The proof of Theorem 2.2 is presented in the appendix. We next describe the key idea underlying the proof of the result as it leads to additional insights on the extent of information collection limitation induced by censoring.

The key idea resides in the introduction of an alternative problem with a new information structure that arises from censoring of a different nature. In particular, we define the *random rejection problem* to be one in which, during each time period, either the decision maker fully observes the realized demand or receives no information at all (this can be interpreted as the decision maker having access to full demand observations, but lacking access to a fraction of them). This revelation occurs independently of the order size or the demand realization, and is based on i.i.d. draws of a Bernoulli random variable with success probability equal to r. Note that the probability of obtaining no information in any given period is equal to the probability of observing a censored observation in the original problem when a myopic policy is applied. The main difference is that, in the original problem every demand realization provides some level of information, while in the random rejection problem some periods provide less (no) information and some periods provide more.

We develop a lower bound on the difference $V_T^m(a,S) - V_T^o(a,S)$ as follows. We establish that the optimal cost in the random rejection case is always lower than that achieved by the myopic policy in the original censored problem. Given this, one can lower bound $V_T^m(a,S) - V_T^o(a,S)$ by the difference between the cost in the random rejection system and that in the observable case. The latter two costs are much simpler to characterize as the update rules do not involve censoring. (Note that, since the information collected at each time period is independent of the decision rule, the random rejection problem is similar to the observable demand case in that an optimal policy is myopic and minimizes the expected current single-period cost.) This yields the bound in (2.15).

One way to interpret the result that the random rejection problem is "easier" than the censored problem is is that high demand realizations are more informative than low demand realizations, and therefore the absence of censoring leads to better performance even with fewer demand observations.

2.3 MOG and COC: Exact Analysis

2.3.1 Scalability

A notable feature of the problem when demand has a Weibull distribution is that the single period optimal cost possesses the so-called scalability property, namely that

$$C(a,S) = S^{1/\ell} C(a,1), \quad \text{for all } a > 1/\ell,\ S > 0.$$

As observed by citetAzoury85 and citetlariviere1999stalking this property can be extended to the optimal and full observation cost functions: for any $a > 1/\ell$, $S > 0$, $T \geq 1$

$$V_T^*(a,S) = S^{1/\ell} V_T^*(a,1), \quad V_T^o(a,S) = S^{1/\ell} V_T^o(a,1).$$

This separability allows the exact cost functions to be determined in the optimal and full observation cases given only knowledge of $V_T^*(a,1)$ and $V_T^o(a,1)$, respectively, for all values of $a > 1/\ell$. Further, exact recursions have been developed for these two quantities in the existing literature citepAzoury85,lariviere1999stalking,Bisi11, and we will use those to evaluate these costs.

A similar reasoning yields that the myopic cost function $V_T^m(a,S)$ also possesses the scalability property, and an exact recursion can be used to compute it, as summarized in the next result.

Proposition 2.4. *Suppose that demand distribution is Weibull. Then, for all $a > 1/\ell$, $S > 0$, $T \geq 1$,*

$$V_T^m(a,S) = S^{1/\ell} V_T^m(a,1).$$

In addition, for all $a > 1/\ell$,

$$V_T^m(a,1) = C(a,1) + \frac{a\ell}{a\ell - 1}\left(1 - (1-r)^{1-\frac{1}{a\ell}}\right) V_{T-1}^m(a+1,1) + (1-r)^{1-\frac{1}{a\ell}} V_{T-1}^m(a,1).$$

In the recursive equation provided by Proposition 2.4, $V_T^m(a,1)$ can be computed exactly given $V_{T-1}^m(a+1,1)$ and $V_{T-1}^m(a,1)$. This implies that $V_T^m(a,1)$ can be evaluated using backwards induction, starting with the boundary condition $V_1^m(a,1) = C(a,1)$. The value of $C(a,1)$ itself has no closed form expression, but can be approximated to arbitrary accuracy by numerical integration for any a. This is analogous to the situation for the value function of an optimal policy in the censored and full observation cases.

Hence, using the scalability property, the cost functions $V_T^m(a,S)$, $V_T^*(a,S)$, and $V_T^o(a,S)$ can be computed numerically using simplified dynamic programming recursions that are exact up to errors from numerical integration. In particular, no discretization of the state space is necessary.

2.3.2 Parametric Setup

We are interested in analyzing the behavior of the MOG and the COC across a broad range of input parameters. In the case of Weibull demand, the input parameters are given by:

- the Weibull parameter ℓ;
- the shape parameter a of the Gamma prior distribution;
- the scale parameter S of the Gamma prior distribution;
- the time horizon T; and
- the service level r.

Given the scalability property discussed in Section 2.3.1, the values of the MOG and the COC do not dependent on the scale parameter S, since both of these quantities are cost differences that are normalized relative to a baseline cost. We are therefore left with four parameters that summarize the three relevant dimensions of the problem: time (T), cost structure / service level (r), and the demand uncertainty (a, l). We consider ranges for time and service level given by

$$1 \leq T \leq 100, \qquad r \in \{0.1, \ldots, 0.9, 0.99\}.$$

The demand uncertainty parameters (a, ℓ) are more difficult to directly interpret. In order to clarify their role, it is convenient to consider two measures of demand uncertainty.

- Conditional on the prior distribution hyperparameters (a, S), we measure the aggregate uncertainty in the next period demand (with predictive distribution (2.6)–(2.7)) through the coefficient of variation of demand, i.e., $\mathrm{CV}(D|a, S)$. Note that $\mathrm{CV}(D|a, S)$ is a function of (a, ℓ), but does not depend on the shape parameter S[5]. In the exponential case ($\ell = 1$), for example, one has that $\mathrm{CV}(D|a, S) = \sqrt{a/(a-2)}$ for $a > 2$.

- The aggregate uncertainty is in part driven by the fact that, even absent uncertainty regarding θ, the demand realizations themselves are random. This can be quantified by the coefficient of variation of demand given the parameter θ, i.e., $\mathrm{CV}(D|\theta)$. Here, demand is assumed to be distributed according to the distribution (2.3). Note that $CV(D|\theta)$ depends on ℓ but does not vary with θ[6] . In the exponential case ($\ell = 1$) one has that $\mathrm{CV}(D|\theta) = 1$ for all $\theta > 0$.

We define the *uncertainty ratio* (UR) as

$$\mathrm{UR}(a, l) := \frac{\mathrm{CV}(D|a, S)}{\mathrm{CV}(D|\theta)}.$$

[5]Indeed, if the random variable D follows the predictive distribution (2.7) with parameters (a, S), then $S^{1/\ell}D$ follows (2.7) with parameters $(a, 1)$. Thus, changing the parameter S corresponds to rescaling the random variable, leaving the coefficient of variation unchanged.

[6]Indeed, if the random variable D follows the demand distribution (2.3) with parameter θ, then $\theta^{-1/\ell}D$ follows (2.3) with parameter $\theta = 1$. Thus, changing the parameter θ corresponds to a rescaling, and leaves the coefficient of variation unchanged.

$\mathrm{UR}(a, l)$ is always greater or equal than 1 and is the ratio of the overall, aggregate uncertainty in the next period demand, to the uncertainty that would remain if θ was perfectly known. From the above discussion, UR is a function of only (a, ℓ). For example, in the exponential case ($\ell = 1$), we have that $\mathrm{UR} = \sqrt{a/(a-2)}$ for $a > 2$.

We will parameterize demand uncertainty through input parameters (UR, ℓ). Note that we use UR rather than a because it is directly interpretable as the relative value of uncertainty arising from the fact that θ is unknown. For example, if $\mathrm{UR} = 3$, the demand uncertainty would be reduced by a factor of three if θ were known.

On the other hand, we interpret the Weibull parameter ℓ as measuring the uncertainty of demand realizations given knowledge of θ. From (2.3), it is clear that larger values of ℓ lead to faster decaying tails for the demand distribution. Moreover, from the above discussion, $\mathrm{CV}(D|\theta)$ depends only on ℓ. If, for example, ℓ increases, the coefficient of variation of the Weibull distribution (independent of θ) decreases, and hence there is less variation in demand realizations. The uncertainty about θ, however, remains constant, cf. (2.5). Because θ can be learned via demand observations, the potential value of exploration can be significantly affected by changing ℓ.

In what follows, we will explore the effect of demand uncertainty on the MOG and the COC.

2.3.3 Analysis of the Myopic Optimality Gap

First, we consider the behavior of the myopic optimality gap, i.e., the relative sub-optimality of the myopic policy as compared to an optimal policy.

Exponential case. To gain some intuition, we start by considering the case where demands follow an exponential distribution ($\ell = 1$). Figure 2.1 depicts the MOG as a function of the time horizon T, holding fixed $\mathrm{UR} = 1.5$ and $r = 0.8$.

We first observe that this curve confirms the results of Proposition 2.1; indeed the MOG decreases as T grows. The most striking conclusion of Figure 2.1, however, is not the shape of the function (which was expected given the previous results) but rather the magnitude of the MOG: *the MOG never exceeds .3% over all values of the time horizon tested.* This says that computing an optimal policy and finely balancing the exploration-exploitation trade-

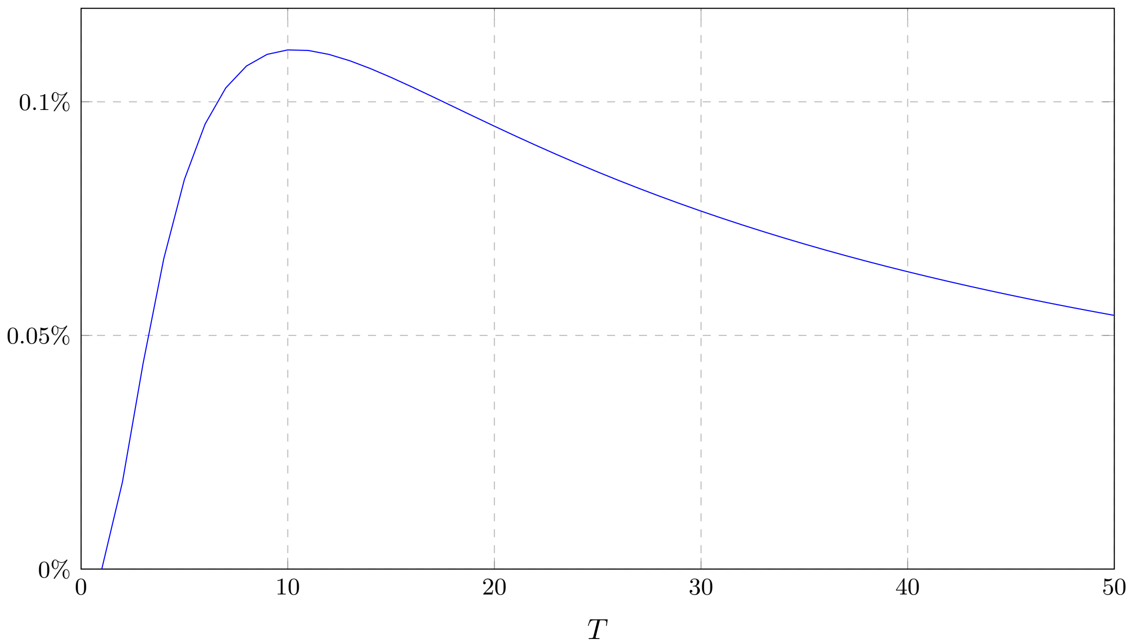

Figure 2.1: **The myopic optimality gap MOG as a function of the time horizon T.** Demand is exponential ($\ell = 1$), $\mathrm{UR} = 1.5$, and $r = 0.8$.

off stemming from demand censoring yields at most 0.3% improvements in costs compared to a myopic policy that simply orders the best quantity given current information. In other words, for all practical purposes, there is essentially no exploration-exploitation trade-off in this case.

To obtain a broader understanding of the behavior and magnitude of the MOG, we next evaluate it for different values of the uncertainty ratio UR and service level r. In Figure 2.2, for varying choices of (r, UR), we depict the worst-case myopic optimality gap $\mathsf{MOG}_{\mathrm{wc}}$ over time horizons $1 \leq T \leq 100$, i.e.,

$$\mathsf{MOG}_{\mathrm{wc}} := \max_{1 \leq T \leq 100} \frac{V_T^m(a,1) - V_T^*(a,1)}{V_T^*(a,1)}. \tag{2.16}$$

Observe that the $\mathsf{MOG}_{\mathrm{wc}}$ tends to zero for values of the service level r close to 0 or 1, consistent with Proposition 2.3. It is also clear that the $\mathsf{MOG}_{\mathrm{wc}}$ decreases as the uncertainty ratio UR decreases, consistent[7] with Proposition 2.2. However, as was the case in Figure 2.1,

[7]Note that, for the exponential case, $\mathrm{UR} = \sqrt{a/(a-2)}$, and hence $a \to \infty$ as $\mathrm{UR} \to 1$. A similar

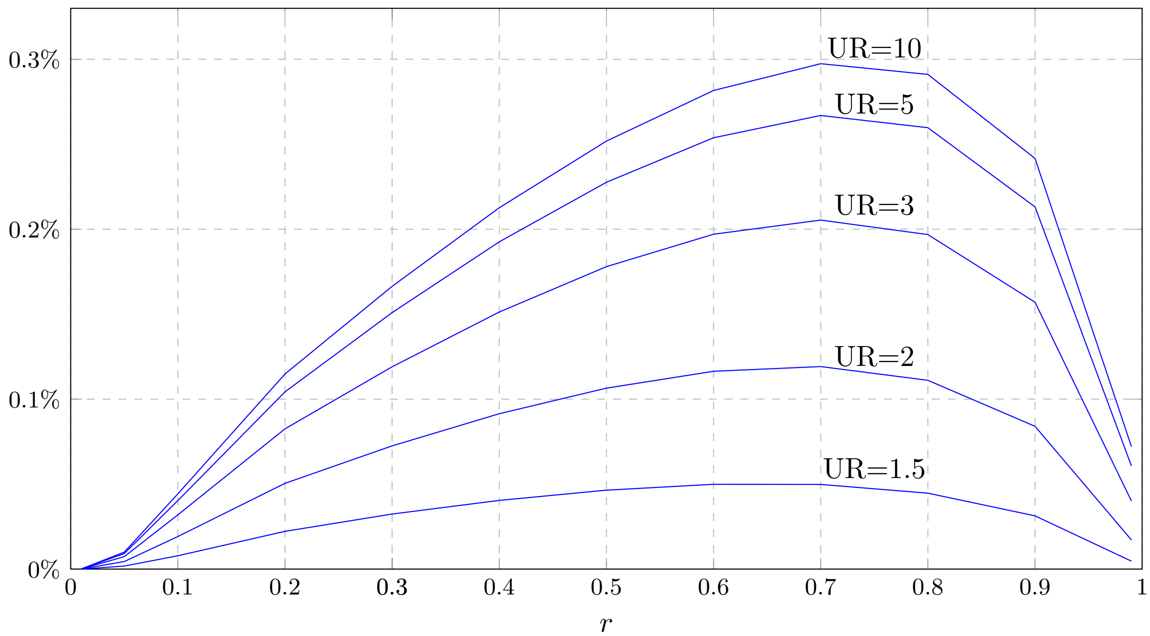

Figure 2.2: **The worst-case myopic optimality gap $\mathsf{MOG}_{\mathrm{wc}}$ as a function of r and UR.** Demand is exponential ($\ell = 1$).

the most remarkable fact is the magnitude of the $\mathsf{MOG}_{\mathrm{wc}}$: the maximum value of the MOG over all parameters tested is below 0.3%. It is notable that this holds for *any* service level r. In particular, even when r is low and censoring occurs often, there is almost no value of deviating from the myopic policy. *In summary, when demand is exponentially distributed, the value of exploration (or of "ordering more") is negligible independent of the problem parameters.*

General Weibull case. We now consider the general Weibull case. We are interested in analyzing the MOG and, in particular, understanding the sensitivity of our conclusions in the exponential case to changes in the shape parameter of the Weibull distribution, ℓ. In Figure 2.3, for different values of the Weibull parameter ℓ and varying choices of the uncertainty ratio UR and the service level r, we depict the worst case myopic optimality gap $\mathsf{MOG}_{\mathrm{wc}}$.

conclusion holds for the general Weibull case.

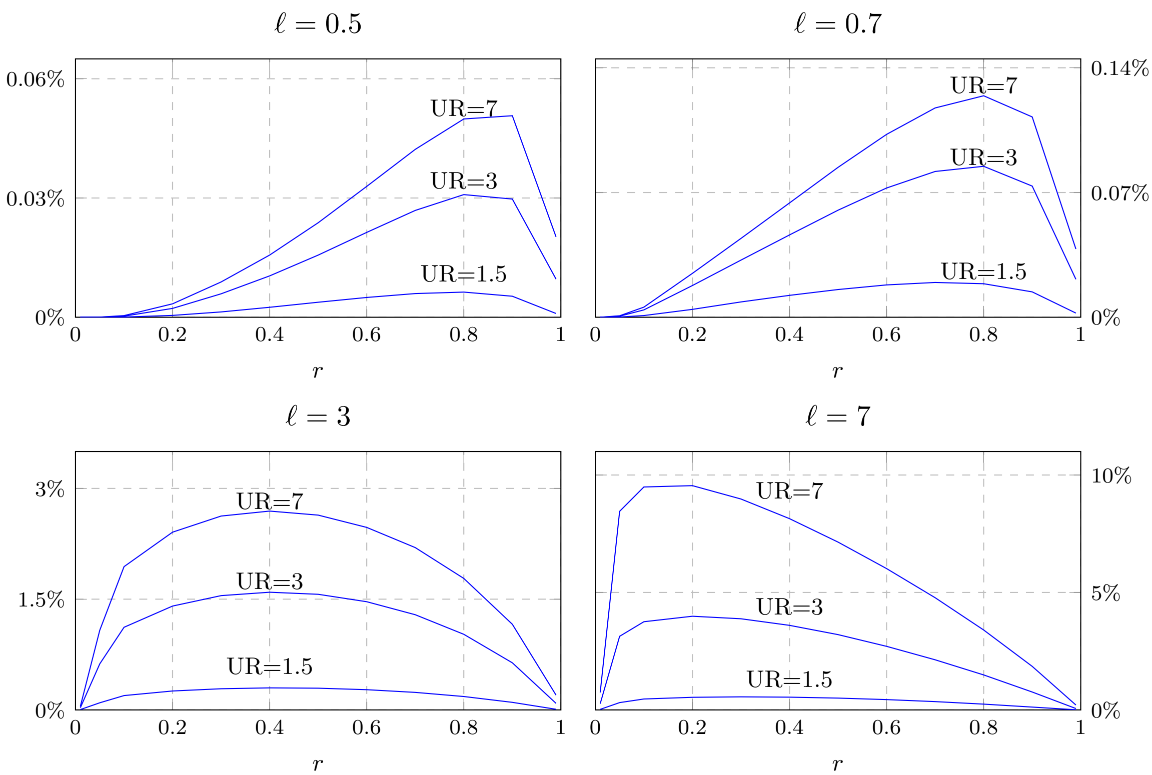

Figure 2.3: **The worst-case myopic optimality gap $\mathsf{MOG}_{\mathrm{wc}}$ as a function of ℓ, r, and UR.**

The results show that, for each value of the Weibull parameter ℓ, the overall shape of the $\mathsf{MOG}_{\mathrm{wc}}$ curve is similar to the one obtained in the exponential case. The main differences are given by the location of the worst case service level r value and the order of magnitude of the gap itself. One observes that the worst-case gap is still below 3% for $\ell \leq 3$, with the worst cases given by $\ell = 7$ and UR $= 7$. Except for those cases, the conclusions of the exponential setting seem to hold in the general Weibull case.

We next analyze in more detail the intuition behind these results.

Analysis of MOG results. To understand what are the main drivers of our results, we start by analyzing, on a sample path basis, the evolution in time of the myopic and optimal order quantities. We measure the level of learning achieved be either policy as the distance between the prescribed order quantity and the optimal order when θ is known. Specifically, if we let $\{y_t^m(\mathbf{D}_1^{t-1})\}_{t=1,\ldots,T}$ and $\{y_t^*(\mathbf{D}_1^{t-1})\}_{t=1,\ldots,T}$ denote the sequences of myopic and optimal orders for a particular sample path $\mathbf{D}_1^T := (D_1, \ldots, D_T)$, and $y(\theta) := F^{-1}(r|\theta)$ the optimal order when θ is known, we define the error terms

$$\mathsf{MyopicError}_t := \mathbb{E}\left[\left|\frac{y_t^m(\mathbf{D}_1^{t-1}) - y(\theta)}{y(\theta)}\right|\right], \text{ and } \mathsf{OptError}_t := \mathbb{E}\left[\left|\frac{y_t^*(\mathbf{D}_1^{t-1}) - y(\theta)}{y(\theta)}\right|\right].$$

To assess the impact that these error have in costs, we consider an alternative error measure. Let $C(y|\theta) := \mathbb{E}[L(D-y)|\theta]$ be the expected cost of ordering y, when θ is known. We then define

$$\mathsf{CostError}_t := \mathbb{E}\left[\frac{C(y_t^*(\mathbf{D}_1^{t-1})|\theta) - C(y_t^m(\mathbf{D}_1^{t-1})|\theta)}{C(y_t^*(\mathbf{D}_1^{t-1}))|\theta)}\right],$$

that is, $\mathsf{CostError}_t$ represents the relative cost distance of the myopic and the optimal policies in therms of the true expected cost given θ.

For brevity, we consider a subset of the examples given in the exponential and Weibull cases. In particular we fix $T = 50$, $\mathrm{UR} = 7$, and consider the $\ell = 1$ (exponential) and $\ell = 7$ cases for two values of r, $r = 0.2$ and $r = 0.8$. For each case we estimate the values of $\mathsf{MyopicError}_t$, $\mathsf{OptError}_t$ and $\mathsf{CostError}_t$ through montecarlo simulation. The results are depicted in Figure 2.4.

Consider first the $\ell = 1$ case. From the top-left picture two main conclusions can be drawn: first, there is non-trivial learning taking place, as there is at least a twofold reduction in $\mathsf{MyopicError}_t$ and $\mathsf{OptError}_t$ for all cases; since both policies prescribe higher orders for $r = 0.8$ the learning rate is steeper in that case. Second, the learning curves of the optimal and myopic policies quickly become indistinguishable. In other words, both policies learn at very similar rates. This fact is reflected on the cost picture, at the bottom-left of Figure 2.4. The cost difference actually favors the mypopic policy in the first periods (i.e. $\mathsf{CostError}_t > 0$) and then this is compensated very slowly by the optimal policy in the rest of the time horizon. The result is a very slight overall advantage of the optimal policy: less than a half percent point in total cost, as shown in Figure 2.2.

Consider now the case of $\ell = 7$. Recall that in these examples $\mathrm{UR} = 7$ and hence this case corresponds to the the highest curve in the $\mathsf{MOG}_{\mathrm{wc}}$ plots at the bottom-right of Figure 2.3. These are the only cases where the MOG can become significant. In these cases the problem primitives satisfy two conditions: (i) there is little noise associated with demand (i.e., ℓ is large) and (ii) there is high uncertainty about the unknown demand parameter θ

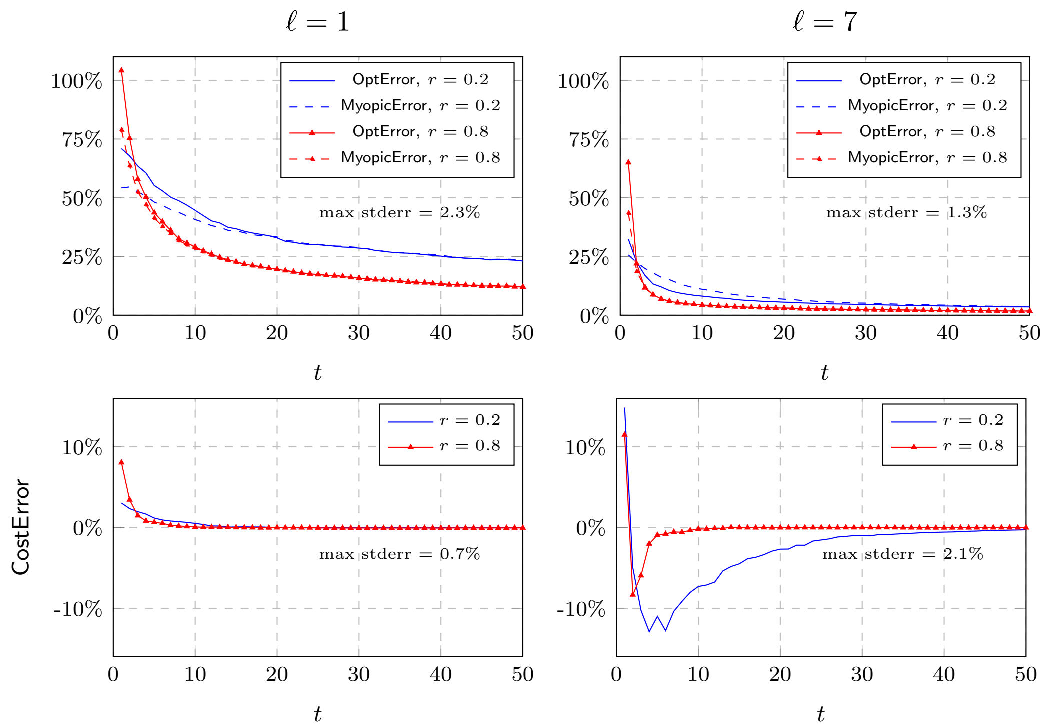

Figure 2.4: **$\mathsf{OptError}_t$, $\mathsf{MyopicError}_t$ and $\mathsf{CostError}_t$ as a function of t, for $T = 50$, UR $= 7$ $\ell = 1, , 7$, and $r = 0.1, 0.8$.**

(i.e., UR is large). Suppose that we keep UR fixed, and hence the ratio between the two sources of uncertainty in the system is constant. When ℓ is large, the coefficient of variation of the demand distribution is small and hence most of the uncertainty in the system comes from the prior distribution, that is, from the fact that θ is unknown. We are then in a situation where, if θ were known, demand itself would be highly predictable (roughly speaking, almost deterministic) and the demand-supply mismatch costs very low. This is evidenced in the steepness of the learning curves in the top-right picture of Figure 2.4: particularly in the $r = 0.8$ case, most of the uncertainty on the optimal order $y(\theta)$ vanishes after a few periods under both policies. Note that, in contrast to the $\ell = 1$ case, for $r = 0.2$ the learning rate of the myopic order is markedly slower then that of the optimal order. This stems from the fact that, even though very few demand observations suffice to nearly learn θ, the myopic censoring probability is high due to the low r and hence the optimal policy can make a difference by ordering more at the beginning. This effect is clearly observed in the evolution of $\mathsf{CostError}_t$ in the bottom-right of Figure 2.4: one can observe how the optimal policy incurs a substantially higher cost at the first periods ($\mathsf{CostError}_t > 10\%$) in order to quickly learn θ and then this produces a substantial advantage in the following periods, where $\mathsf{CostError}_t < 0$. Overall the total cost difference, for the $r = 0.2$ case, is no longer negligible, close to 10%, as shown in the bottom-right picture of Figure 2.3.

To summarize, the prior discussion shows that the MOG can indeed be significant, but only if the problem satisfies very specific conditions: (i) there is little noise associated with demand, (ii) there is high uncertainty about the unknown demand parameter θ and (iii) the holding cost h is high relative to the penalty cost p. In these situation there is a high potential gain from exploring, and hence one may not ignore the exploration-exploitation trade-off. From a practical point of view, however, when the three conditions above are satisfied, one faces a problem of a different nature than the typical newsvendor problem we started with: in these cases, demand is, roughly speaking, deterministic and can be essentially learned exactly with very few uncensored observations. Aside from these examples, in some sense *degenerate* cases, the results in this section show that the MOG is almost uniformly negligible.

2.3.4 Analysis of the Cost of Censoring

We next assess the cost of censoring, that is, the relative cost of going from an uncensored to a censored environment. As in Section 2.3.3, we will consider the worst-case cost of censoring COC_{wc} over values of the time horizon $1 \leq T \leq 100$, i.e.,

$$\mathsf{COC}_{\text{wc}} := \max_{1 \leq T \leq 100} \frac{V_T^*(a,1) - V_T^o(a,1)}{V_T^o(a,1)}. \tag{2.17}$$

In Figure 2.5, we plot COC_{wc} as a function of the uncertainty ratio UR and the service level r, for different values of the Weibull parameter ℓ.

Figure 2.5 shows that the overall shape and the behavior in extreme cases of the cost of censoring is similar to that of the myopic optimality gap . The main difference between the results in Figure 2.5 and those shown in Figure 2.3 is given by the order of magnitude of the gaps. The COC_{wc} appears to be at least an order of magnitude larger than the MOG_{wc}.[8] To

[8] As in Section 2.3.3, one can observe that the most extreme cases are given in the lower right plot of Figure 2.5, particularly with higher values of UR. One can apply the same reasoning as before to explain

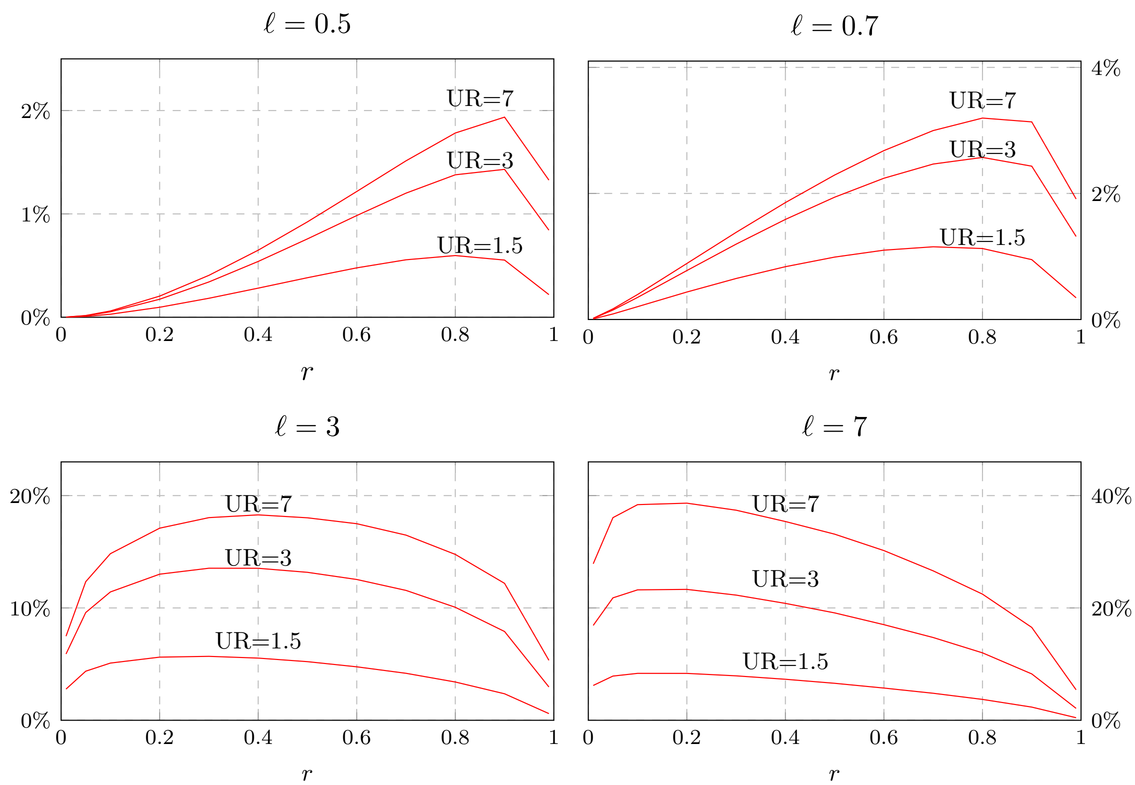

Figure 2.5: **The worst-case cost of censoring COC_{wc} as a function of ℓ, r and** UR**.**

make clearer comparison, Figure 2.6 combines the results of Figures 2.3 and 2.5 in single plot: the red lines represent the worst case cost of censoring COC_{wc} and the blue areas represent the worst case myopic optimality gap MOG_{wc}, displayed as an increment over the cost of censoring. As we can see from the results, the MOG is not only low in absolute terms, but also relative to the COC.

Implications. At a high level, most practitioners are well aware of censoring but rarely fully recognize the exploration-exploitation trade-off, focusing more on attempting to record lost sales. The comparison above is informative in the following sense. It shows that the exploration-exploitation trade-off and the need for forward looking policies introduced by demand censoring (with the computational complexity that might be associated with it) is, for all practical purposes, a second order problem compared to the value that might be

the high gap values in this context: if there is high θ uncertainty and very low demand noise, an uncensored demand observation essentially reveals future demand values, leading to a problem of a different nature.

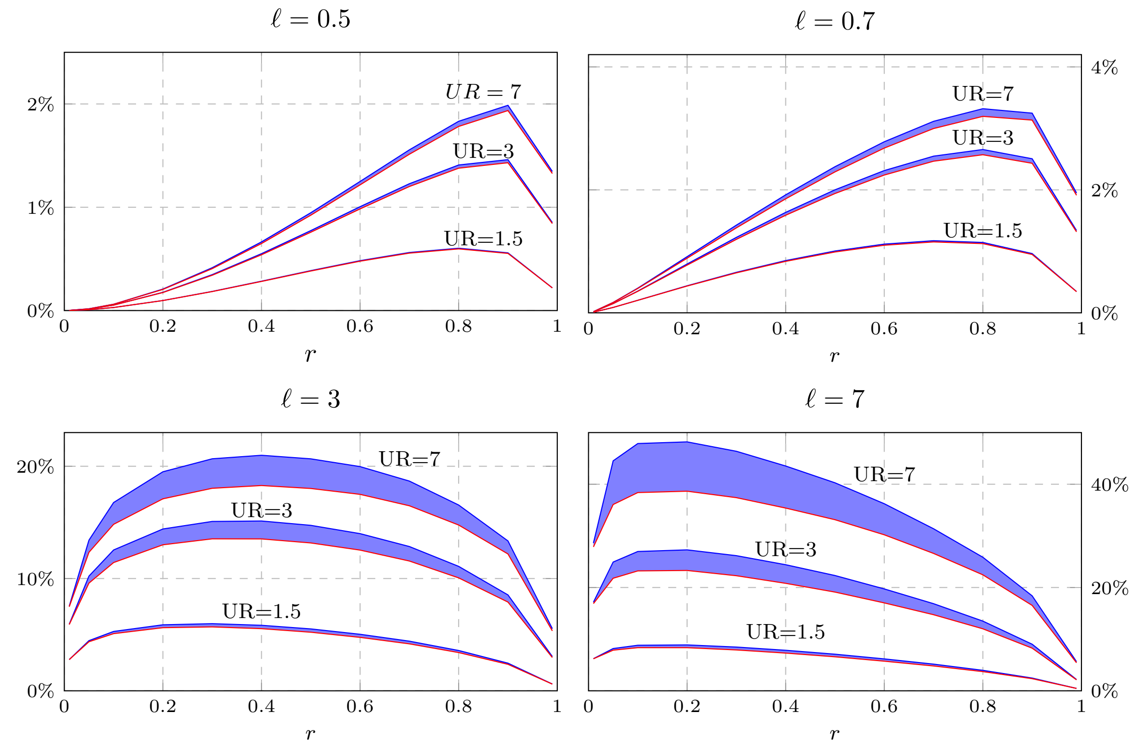

Figure 2.6: **Comparison of the worst-case cost of censoring COC_{wc} and the worst case myopic optimality gap MOG_{wc} as a function of ℓ, r and UR.** The red lines represent the COC_{wc} and the blue areas represent the MOG_{wc}, displayed as an increment over the COC_{wc}.

generated by investing in processes and technology to uncensor (even partially) demand.

2.4 Concluding Remarks

In the present chapter, we study the implications of demand censoring in inventory problems, focusing on the perishable, or newsvendor, case. From a performance point of view, we study the overall impact of censoring on costs, and we find that, for one of the few known tractable cases, the extra incurred cost is low. We also study how censoring affects decisions, and in particular how the exploration-exploitation tradeoff introduced by censoring affects an (otherwise optimal) myopic policy. We find that, for practical purposes, there is virtually no tradeoff in this case: being myopic is essentially as good as optimal. Operationally, this surprising fact implies that there is no need to develop and apply optimal policies. We also find that, even though the cost of censoring is low, it is an order of magnitude higher than the myopic optimality gap. This means that if a decision maker faces the choice of either investing in improving decision policies or in collecting data on lost sales, effort would be better rewarded by focusing on the latter.

Due to analytical reasons, we focused our study on a particular family of distributions, widely adopted in the literature. On the one hand, this restricts the scope of our conclusions. On the other hand, this is virtually the only known tractable case of this problem. This work is a first step towards the quantification of the fine trade-off associated with information collection introduced by demand censoring. While challenging to formalize and establish, we conjecture that the general conclusion that demand censoring and, more specifically, its effects on policy decisions and subsequent performance, can be, to some extent, disregarded, might extend beyond the Weibull demand case we studied. This is a worthwhile avenue of future research. Indeed, it is worth noting that in more general cases with non-conjugate families of distributions, the dynamic optimization problem becomes infinite dimensional and obtaining an optimal policy is highly intractable. However, the myopic policy is the easiest policy to apply in practice, and is always a viable heuristic.

Chapter 3

Dynamic learning in sequential auctions

3.1 Problem Formulation

We consider an agent facing a sequence of T auctions for identical copies of a good. If the agent wins an auction he is awarded with item and collects a reward associated with it. The agent seeks to maximize the cumulative reward over the T auctions, and is assumed to be risk neutral. In each auction there exists a set of competing players who bid for the item. We summarize the information about the competing players at period t in a *competing bid* $\widehat{b}_t$ that represents the maximum of their bids. We assume that $\widehat{b}_1, \ldots, \widehat{b}_T$ are independent and identically distributed random variables, with support $[0, \overline{b}]$, where $\overline{b} \in \mathbb{R}_+ \cup \{+\infty\}$, and a cdf denoted by $q(\cdot)$.

We denote the reward collected by the agent, if he wins the auction at period t, by X_t. This reward should be interpreted as the benefit that the agent obtains by winning the auction. For example, if the auctioned item is a slot for a graphical ad in a webpage, X_t represents the revenue associated with the displayed ad. We assume that the rewards in each auction are independent and identically distributed, belonging to a family of distributions parametrized by an unknown parameter $\nu \in \mathcal{N} \subset \mathbb{R}^n$.[1] Following a Bayesian approach, we

[1] Following the graphical ad example above, ν could represent the (unknown) click probability of the ad.

assume that the bidder holds a prior distribution over ν, denoted by $\pi(\cdot)$. Given ν, the cdf and pdf of the rewards are denoted by $F(\cdot|\nu)$ and $f(\cdot|\nu)$ respectively. We denote by μ_ν the expected reward conditional on ν, that is $\mu_\nu := \int x f(x|\nu)dx$, and by $\mu(\pi)$ the predictive expected reward, i.e. $\mu(\pi) := E_\pi[X] := \int x \int f(x|\nu)\pi(\nu)d\nu dx$. Unless otherwise stated, in the rest of this chapter the expectation $\mathbb{E}_\pi[\cdot]$ is taken with respect to $(\{X_t\}, \nu, \{\widehat{b}_t\})$, where ν has pdf $\pi(\cdot)$, X_t has (conditional) cdf $F(\cdot|\nu)$ and $\widehat{b}_t$ has cdf $q(\cdot)$. For simplicity, in some cases we will denote by X and $\widehat{b}$ generic versions of the reward and competing bid, equally distributed to X_t and $\widehat{b}_t$ respectively.

For simplicity, we further assume that the agent is facing *second price* auctions. This assumption is not critical for our results, but it facilitates exposition. In section 3.2.2 we detail how our framework can be adapted to a more general setting. According to a second price auction structure, the expected single period reward, when the agent's bid is b, is given by

$$\mathbb{E}_\pi \left[\mathbb{I}_{\{b \geq \widehat{b}\}} \left(X - \widehat{b} \right) \right].$$

Note that we do not model explicitly a reserve price in the structure of the auction, but such reserve price may be encoded in $\widehat{b}$.

The bidder only observes (and collects) rewards if he wins the auction. We denote by $\mathcal{F}_t$ the filtration generated by the partially observed reward process, that is

$$\mathcal{F}_t := \left\{ (b_1, \mathbb{I}_{\{b_1 \geq \widehat{b}_1\}} X_1), \ldots, (b_{t-1}, \mathbb{I}_{\{b_1 \geq \widehat{b}_{t-1}\}} X_{t-1}) \right\}.$$

We assume that $\mathbb{E}_\pi [X_1] < \infty$, which implies that $\mathbb{E}_\pi \left[\mathbb{I}_{\{b \geq \widehat{b}\}} \left(X_t - \widehat{b}_t \right) \right] < \infty$, for all $t = 1, \ldots, T$ and any $b \in [0, \bar{b}]$. We denote by $\mathcal{P}$ the set of $\mathcal{F}_t$-measurable policies for all $t \geq 1$. The objective is to maximize the cumulative expected reward. The optimal value is given by

$$V_T^*(\pi) = \sup_{\beta \in \mathcal{P}} \sum_{t=1}^{T} \mathbb{E}^\beta \left[\mathbb{I}_{\{b_t \geq \widehat{b}_t\}} \left(X_t - \widehat{b}_t \right) \right], \tag{3.1}$$

where the expectation is taken assuming that decisions are made according to the policy β.

3.1.1 Dynamic Programming Formulation and Myopic Value

The value function defined in (3.1) can be alternatively written using Bellman's equation as follows:

$$V_T^*(\pi) = \max_{b\in[0,\bar{b}]} \mathbb{E}_\pi \left[\mathbb{I}_{\{b\geq\widehat{b}\}} \left(X - \widehat{b}\right)\right] + q(b)\mathbb{E}_\pi \left[V_{T-1}^*(\pi \oplus X)\right] + (1 - q(b))V_{T-1}^*(\pi), \quad (3.2)$$

with boundary conditions $V_0^*(\pi) := 0$ for all π. Here $\pi \oplus X$ represents the updated beliefs after observing the reward X, that is

$$\pi \oplus X(\nu) := \frac{\pi(\nu)f(X|\nu)}{\int \pi(\widetilde{\nu})f(X|\widetilde{\nu})d\widetilde{\nu}}.$$

The first term in equation (3.2) represents the present value of placing a bid b, while the second and third terms represent the future value of bidding b, purely associated to the *learning* component of the problem. That is, a bid that is optimal in terms of today's rewards might not be optimal if one takes into account future rewards. This represents the exploration-exploitation tradeoff in this problem. It is not hard to prove in this case that, in a multiperiod problem, it is always optimal to bid more than what one would in a single auction. Let $Q(\pi)$ denote the optimal single period auction reward, that is,

$$Q(\pi) \quad := \quad \max_{b\in[0,\bar{b}]} \mathbb{E}_\pi \left[\mathbb{I}_{\{b\geq\widehat{b}\}} \left(X - \widehat{b}\right)\right] = \max_{b\in[0,\bar{b}]} q(b)\mu(\pi) - p(b), \quad (3.3)$$

where $p(b) := \mathbb{E}_\pi[\,\widehat{b}\,\mathbb{I}_{\{\widehat{b}\leq b\}}]$ represents the expected payments if the agent places a bid of b. Note that, because this is a second price auction, the agent's optimal strategy is to bid his own expected reward. That is, the bid that maximizes (3.3) is given by $b^m = \mu(\pi)$. We call this bid the *myopic* bid, as it represents the optimal action for an agent facing a single period auction. The following lemma shows that, in a multiperiod problem, it is always optimal to bid higher than the myopic bid.[2]

Lemma 3.1. *Let $b^*(\pi, T)$ be the optimal bid in (3.2). For any prior distribution π and any T,*

$$b^*(\pi, T) \geq \mu(\pi).$$

[2]Similar version of this lemma have been proven in [Li *et al.*, 2010] and [Iyer *et al.*, 2014].

Proof. By differentiating with respect to b in the the right side of (3.2) one obtains

$$\frac{\partial}{\partial b}\mathbb{E}_\pi\left[\mathbb{I}_{\{b\geq\widehat{b}\}}\left(X-\widehat{b}\right)\right]+q'(b)\left(\mathbb{E}_\pi[V^*_{T-1}(\pi\oplus X)]-V^*_{T-1}(\pi)\right).$$

Because $\frac{\partial}{\partial b}\mathbb{E}_\pi\left[\mathbb{I}_{\{b\geq\widehat{b}\}}\left(X-\widehat{b}\right)\right]$ equals 0 at $b=\mu(\pi)$ and $q'(b)\geq 0$ for all $b\in[0,\bar{b}]$, it suffices to show that

$$\mathbb{E}_\pi[V^*_{T-1}(\pi\oplus X)]-V^*_{T-1}(\pi)\geq 0.$$

We proceed by induction. The case $T-1=0$ is trivial, since both terms equal 0. Suppose now that the result is true for $T-1$. Consider now $V^*_T(\pi\oplus X)$ and suppose one picks $b:=b^*(\pi,T)$ as a bid. Clearly the reward is worse than the optimal value, that is,

$$\begin{aligned}V^*_T(\pi\oplus X) \quad\geq\quad & \mathbb{E}_{\pi\oplus X}[\mathbb{I}_{\{b^*(\pi,T)\geq\widehat{b}\}}(X'-\widehat{b})]+q(b^*(\pi,T))\mathbb{E}_{\pi\oplus X}\left[V^*_{T-1}(\pi\oplus X\oplus X')\right]\\ & +(1-q(b^*(\pi,T)))V^*_{T-1}(\pi\oplus X),\end{aligned}$$

where X' is a random variable iid to X. Note that, because of the tower property of expectations, one has that

$$\mathbb{E}_\pi\left[\mathbb{E}_{\pi\oplus X}\left[\mathbb{I}_{\{b^*(\pi,T)\geq\widehat{b}\}}(X'-\widehat{b})\right]\right]=\mathbb{E}_\pi\left[\mathbb{I}_{\{b^*(\pi,T)\geq\widehat{b}\}}(X-\widehat{b})\right].$$

Therefore,

$$\begin{aligned}\mathbb{E}_\pi[V^*_T(\pi\oplus X)]-V^*_T(\pi) \quad\geq\quad & q(b^*(\pi,T))\mathbb{E}_\pi\left[V^*_{T-1}(\pi\oplus X\oplus X')-V^*_{T-1}(\pi\oplus X)\right]+\\ & (1-q(b^*(\pi,T)))\mathbb{E}_\pi\left[V^*_{T-1}(\pi\oplus X)-V^*_{T-1}(\pi)\right].\end{aligned}$$

By the induction hypothesis both terms inside the expectations are nonnegative and therefore the result is proven. □

Let us now formally define the *myopic* value function, which results from placing the myopic bid on every auction, given the current information:

$$V^m_T(\pi) \quad:=\quad Q(\pi)+q(\mu(\pi))\mathbb{E}_\pi\left[V^m_{T-1}(\pi\oplus X)\right]+(1-q(\mu(\pi)))V^m_{T-1}(\pi), \qquad (3.4)$$

with boundary conditions $V^m_0(\pi):=0$ for all π.

3.1.2 An information relaxation bound

Because the optimal policy is hard to characterize analytically, we will rely on an information relaxation bound to make the analysis more tractable. We consider an alternative informational system, where the agent always *observes* the reward X, regardless of whether he wins the auction or not. We call this the *full observation* system. The filtration generated by this process is given by

$$\mathcal{F}_t^o := \{(b_1, X_1), \ldots, (b_{t-1}, X_{t-1})\}.$$

Note that in this case decisions and information collection are independent and hence the optimal policy is to play myopically in every period. Let $V_T^o(\pi)$ denote the optimal value function in the full observation system, which can be defined using Bellman's equation as

$$V_T^o(\pi) = \mathbb{E}_\pi\left[\mathbb{I}_{\{\mu(\pi)\geq\widehat{b}\}}\big(X - \widehat{b}\big)\right] + \mathbb{E}_\pi[V_{T-1}^o(\pi \oplus X)].$$

Note that, because the myopic policy is an admissible policy in the problem defined in (3.1) and $\mathcal{F}_t \subset \mathcal{F}_t^o$, one can conclude that

$$V_T^m(\pi) \leq V_T^*(\pi) \leq V_T^o(\pi).$$

In particular, this implies that

$$V_T^*(\pi) - V_T^m(\pi) \leq V_T^o(\pi) - V_T^m(\pi). \tag{3.5}$$

3.2 Main Results

Lemma 3.1 shows that in this problem the exploration always occurs in one direction: it is always optimal to bid *more* in order to collect more information for the future. Note, however, that even by playing myopically and ignoring the learning component of the decision the bidder still implicitly learns, if the probability of winning remains high enough. This raises the question of to what extent, or in which cases, the exploration-exploitation trade-off matters.

There are, nevertheless, situations where the myopic policy is indeed expected to perform poorly. In particular, when the distributions of ν and the competing bid $\widehat{b}$ define a region in

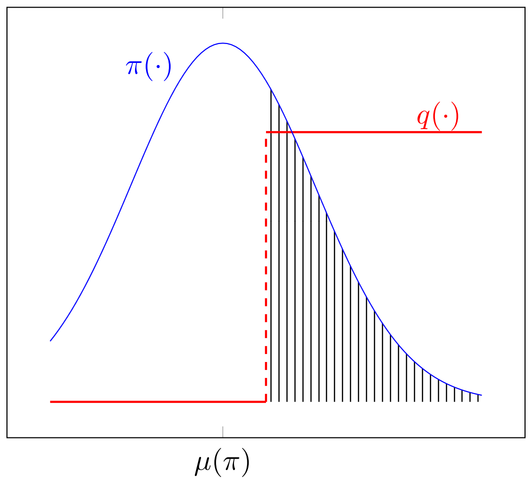

Figure 3.1: **Example of prior and competing bid distributions where the myopic policy produces zero rewards, but there exist potential gains from overbidding.** Because $\mu(\pi) < \widehat{b}$, the myopic policy loses all auctions and obtains 0 rewards, but there are potential gains from overbidding to learn if the true valuation lies above $\widehat{b}$, which occurs with probability equal to the shaded area.

the bid space where the myopic policy can get "stuck", that is, a region with zero probability of (myopically) winning an auction. For example, consider an auction where the competing bid is deterministic, but the agent's expected reward is unknown. If the current (predictive) expected reward $\mu(\pi)$ lies below the competing bid, the myopic policy will lose every auction and its cumulative reward will be 0. Note however that, because of the uncertainty around ν, there is potential incentive to bid higher than the myopic bid $\mu(\pi)$. This situation is depicted in Figure 3.1 (for simplicity we assume that $\mathbb{E}[X|\nu] = \nu$). The shaded area represents the probability of ν belonging to a region where the bidder would profit from the auctions. Though extreme, since it is unlikely that the competing bid be deterministic and known, this example illustrates the disadvantage of playing myopically when the current expected reward has a low winning probability, but there is still high uncertainty around its true value. In those cases the myopic policy essentially stops learning and the potential gain from exploration grows.

The example of Figure 3.1 illustrates situations where the myopic policy performs poorly. In Theorem 3.1 and Corollary 3.1 below we develop bounds that help characterize cases where being myopic indeed produces good results. In particular, we develop a bound on the difference between the myopic value and the full observation value, $V^o - V^m$, which in turn serves as an upper bound on $V^m - V^*$, as shown in (3.5).

Theorem 3.1. *Let $X_1^t := (X_1, \ldots, X_t)$, $\pi_t^o := \pi \oplus X_1 \oplus \ldots \oplus X_t$ and $\mathrm{Var}_{\widetilde{\nu}\sim\pi}(\mu_{\widetilde{\nu}})$ denote the variance of $\mu_{\widetilde{\nu}}$, when $\widetilde{\nu}$ has a pdf π. Suppose that $q(\cdot)$ has a bounded first derivative. Then there exists $K > 0$ such that for any T and prior distribution π,*

$$V_T^o(\pi) - V_T^m(\pi) \leq K \sum_{t=0}^{T-2} \mathbb{E}_{X_1^t}\left[(1 - q(\mu(\pi_t^o)))\frac{\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\mu_{\widetilde{\nu}})}{q(\mu(\pi_t^o))}\right].$$

Proof. We start by defining some additional notation. Unless otherwise stated, in this proof all the expectations are taken with π as a prior distribution, that is, $\mathbb{E}[\cdot] \equiv \mathbb{E}_\pi[\cdot]$. We will denote by

$$Q_\infty(\pi) := \mathbb{E}\left[\max_b \mathbb{E}\left[\mathbb{I}_{\{b\geq\widehat{b}\}}\left(X - \widehat{b}\right)\Big|\nu\right]\right] = \mathbb{E}\left[\max_b q(b)\mathbb{E}\left[X\big|\nu\right] - p(b)\right] \tag{3.6}$$

the expected single auction value when ν is known by the agent. Let us also define

$$\Gamma_T(\pi) := \mathbb{E}[V_T^m(\pi \oplus X)] - V_T^m(\pi)$$

as the (myopic) value of a reward observation.

We have

$$\begin{aligned}
&V_T^o(\pi) - V_T^m(\pi) \\
&= \mathbb{E}[V_{T-1}^o(\pi \oplus X)] - q(\mu(\pi))\mathbb{E}[V_{T-1}^m(\pi \oplus X)] - (1 - q(\mu(\pi)))V_{T-1}^m(\pi) \\
&= \mathbb{E}[V_{T-1}(\pi \oplus X)] - \mathbb{E}[V_{T-1}^m(\pi \oplus X)] + (1 - q(\mu(\pi)))\left[\mathbb{E}[V_{T-1}^m(\pi \oplus X)] - V_{T-1}^m(\pi)\right] \\
&= \mathbb{E}[V_{T-1}(\pi \oplus X)] - \mathbb{E}[V_{T-1}^m(\pi \oplus X)] + (1 - q(\mu(\pi)))\Gamma_{T-1}(\pi) \\
&= \sum_{t=0}^{T-2} \mathbb{E}\left[(1 - q(\mu(\pi_t^o)))\Gamma_{T-1-t}(\pi_t^o)\right] \\
&\leq \sum_{t=0}^{T-2} \mathbb{E}\left[\frac{1 - q(\mu(\pi_t^o))}{q(\mu(\pi_t^o))}\left(Q_\infty(\pi_t^o) - Q(\pi_t^o)\right)\right] \qquad (3.7)
\end{aligned}$$

where $\pi_t^o := \pi \oplus X_1 \oplus \ldots \oplus X_t$ and the last inequality comes from Lemma B.1 in Appendix B. We now aim to bound the summand in (3.7). Recall from (3.3) and (3.6) that $Q(\pi_t^o) = \max_b\{q(b)\mu(\pi_t^o) - p(b)\}$ and $Q_\infty(\pi) = \int \max_b\{q(b)\mu_{\widetilde{\nu}} - p(b)\}\pi(\widetilde{\nu})d\widetilde{\nu}$. Let us define

$$g(z) := \max_b\{q(b)z - p(b)\}.$$

Therefore, term t in (3.7) becomes

$$\begin{aligned} & \mathbb{E}\Big[\frac{1-q(\mu(\pi_t^o))}{q(\mu(\pi_t^o))}\Big(\int g(\mu_{\widetilde{\nu}})\pi_t^o(\widetilde{\nu})d\widetilde{\nu} - g(\mu(\pi_t^o))\Big)\Big] \\ = \ & \mathbb{E}\Big[\frac{1-q(\mu(\pi_t^o))}{q(\mu(\pi_t^o))} \int \big(g(\mu_{\widetilde{\nu}}) - g(\mu(\pi_t^o))\big)\pi_t^o(\widetilde{\nu})d\widetilde{\nu}\Big], \end{aligned} \tag{3.8}$$

Consider the Taylor expansion of $g(\cdot)$ at $\mu(\pi_t^o)$ and note that, due to the envelope theorem, $g'(z) = q(z)$. Thus, we have

$$\begin{aligned} g(\mu_{\widetilde{\nu}}) - g(\mu(\pi_t^o)) &= q(\mu(\pi_t^o))(\mu_{\widetilde{\nu}} - \mu(\pi_t^o)) + \frac{1}{2}q'(\xi)(\mu_{\widetilde{\nu}} - \mu(\pi_t^o))^2 \\ &\leq q(\mu(\pi_t^o))(\mu_{\widetilde{\nu}} - \mu(\pi_t^o)) + K(\mu_{\widetilde{\nu}} - \mu(\pi_t^o))^2, \end{aligned} \tag{3.9}$$

where the inequality comes from the fact that $q'(\cdot)$ is bounded. Next, note that

$$\begin{aligned} \mu(\pi_t^o) &:= \int x \int f(x|\widetilde{\nu})\pi_t^o(\widetilde{\nu})d\widetilde{\nu}dx \\ &= \int\int x f(x|\widetilde{\nu})dx \, \pi_t^o(\widetilde{\nu})d\widetilde{\nu} \\ &= \int \mu_{\widetilde{\nu}}\pi_t^o(\widetilde{\nu})d\widetilde{\nu}, \end{aligned} \tag{3.10}$$

and therefore, by integrating on (3.9) one obtains

$$\int \big(g(\mu_{\widetilde{\nu}}) - g(\mu(\pi_t^o))\big)\pi_t^o(\widetilde{\nu})d\widetilde{\nu} \leq K \int (\mu_{\widetilde{\nu}} - \mu(\pi))^2\pi_t^o(\widetilde{\nu})d\widetilde{\nu} := K \,\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\widetilde{\nu}),$$

where the last term comes from the fact that $\mu(\pi_t^o) = \int \mu_{\widetilde{\nu}}\pi_t^o(\widetilde{\nu})d\widetilde{\nu} =: \mathbb{E}_{\widetilde{\nu}\sim\pi_t^o}[\mathbb{E}[X|\widetilde{\nu}]]$. By combining this with (3.8) the result is proven. $\square$

The numerator in the expectation of Theorem 3.1 $\big[1 - q(\mu(\pi_t^o))\big]\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\mu_{\widetilde{\nu}})$ can be interpreted as the losses due to the fact that the agent only observes a fraction of the reward realizations: $1 - q(\mu(\pi_t^o))$ represents the fraction of lost auctions when playing myopic and the posterior variance $\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\mu_{\widetilde{\nu}})$ the estimation error associated with the missed reward observations. The denominator $q(\mu(\pi_t^o))$ can be associated with the fact that when the probability of winning an auction is low, the myopic loss grows.

The posterior variance shrinks as t grows and the uncertainty around ν, and μ_ν, is reduced. The denominator $q(\mu(\pi_t^o))$, on the other hand, becomes closer to $q(\mu_\nu)$, as $\mu(\pi_t^o) \to \mu_\nu$, and could render the expectation unbounded if $1/q(\mu_\nu)$ is not integrable. This is the case of the example of Figure 3.1. In that example, the expectations in Theorem 3.1 are equal to infinity for every t. In order to avoid this type of situations, more restrictions are needed on the competing bid and reward distributions. In Corollary 3.1 we make mild extra assumptions to ensure that the bound in Theorem 3.1 is finite, and we show that under these assumptions the myopic loss growths only at a logarithmic rate in time, and shrinks as the prior uncertainty is reduced at a rate proportional to the prior variance.

Corollary 3.1. *Let* $\pi_k^o := \pi \oplus X_1 \oplus \ldots \oplus X_k$. *Suppose the assumptions in Theorem 3.1 hold and that* $1/q(x) = O(\phi(x))$ *as* $x \to 0$, *where* $\phi(\cdot)$ *is a convex function such that* $\mathbb{E}_\pi[\phi(\mu_\nu)] := \int \phi(\mu_\nu)\pi(\nu)d\nu < \infty$. *If* $\mathrm{Var}_{\nu\sim\pi_k^o}(\mu_\nu) = O(1/k)$ *as* $k \to \infty$ *almost surely, then*

a)

$$V_T^o(\pi) - V_T^m(\pi) = O(\log(T)) \text{ as } T \to \infty.$$

b)

$$\mathbb{E}\pi\left[V_T^o(\pi_k^o) - V_T^m(\pi_k^o)\right] = O(1/k) \text{ as } k \to \infty.$$

Proof. First note that

$$\begin{aligned}
\mathbb{E}_{X_1^t}\left[q(\mu(\pi_t^o))^{-1}\right] &\leq O(1)\mathbb{E}_{X_1^t}\left[\phi\left(\int \mu_{\widetilde{\nu}}\pi_t^o(\widetilde{\nu})d\widetilde{\nu}\right)\right] \\
&\leq O(1)\mathbb{E}_{X_1^t}\left[\int \phi(\mu_{\widetilde{\nu}})\pi_t^o(\widetilde{\nu})d\widetilde{\nu}\right] \\
&= O(1)\int \phi(\mu_{\widetilde{\nu}})\pi(\widetilde{\nu})d\widetilde{\nu},
\end{aligned}$$

where the first inequality comes from (3.10) and the fact that $1/q(\cdot) = O(\phi(\cdot))$, the second inequality comes from Jensen's inequality, and the last equality from the law of iterated expectations. Now, consider the summand in the bound of Theorem 3.1:

$$\mathbb{E}_{X_1^t}\left[(1 - q(\mu(\pi_t^o)))\frac{\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\mu_{\widetilde{\nu}})}{q(\mu(\pi_t^o))}\right] \leq O(1/t)\mathbb{E}_{X_1^t}\left[q(\mu(\pi_t^o))^{-1}\right] = O(1/t).$$

Summing over t gives part a). For part b), note that if $\pi := \pi_k^o$ for same k, then $\pi_t^o := \pi_{k+t}^o$ and since $\mathrm{Var}_{\widetilde{\nu}\sim\pi_{k+t}^o} = O(1/(k+t)) = O(1/k)$, we have

$$\mathbb{E}_{X_1^t}\left[(1 - q(\mu(\pi_t^o)))\frac{\mathrm{Var}_{\widetilde{\nu}\sim\pi_t^o}(\mu_{\widetilde{\nu}})}{q(\mu(\pi_t^o))}\right] = O(1/k),$$

with a constant independent from t, which proves b). ∎

Part a) of Corollary 3.1 shows that the myopic loss grows at most logarithmically with the size of the time horizon. Part b) shows that the rate at which the myopic loss is reduced as the original uncertainty around ν decreases is proportional to the prior variance. Corollary 3.1 requires that $1/q(\mu_\nu)$ is bounded by an integrable function near 0, which is equivalent to requiring that $\mathbb{E}[1/q(\mu_\nu)] < \infty$, precluding situations like that of Figure 3.1. This is a necessary condition for the summand in Theorem 3.1 to be finite; by further requiring convexity of $\phi(\cdot)$ it becomes sufficient as well. The convergence rate of the posterior variance $1/k$ is inspired on a frequentist approach: $1/k$ is the rate of convergence of the variance of the sample average for independent observations. In a Bayesian setting it is harder to ensure in general that $1/k$ is indeed the rate of convergence of the posterior variance, but is readily verifiable for many common distribution families.

Though not true in general, it is possible to verify that $1/q(\cdot)$ is itself convex, if $q(\cdot)$ is the cdf of common families of distributions, such as Normal or Gamma. In particular, if $1/q(\cdot)$ is *polynomically bounded* near 0, i.e. $1/q(x) = O(x^{-\alpha})$ for some $\alpha \in \mathbb{R}$, then the convexity condition in Corollary 3.1 is satisfied. This case is particularly relevant to our problem because $q(\cdot)$ is defined as the cdf of the competing bid in the auction. Suppose the auction is composed of j bidders indexed by $i = 0, \ldots, j-1$, where $i = 0$ is the index of the agent. Let us denote by $\widehat{b^i}$ the bid of bidder i, and assume that the bids are independent and identically distributed, with cdf denoted by $G(\cdot)$. In this case, $q(x) := P\{\max_{i=1\ldots j-1} \widehat{b^i} \leq x\} = G^{j-1}(x)$. Thus, if $1/G(\cdot)$ is polynomically bounded, then so is $1/q(\cdot)$. Therefore, $q(\cdot)$ such that $1/q(x) = O(x^{-\alpha})$ represents a natural choice when applying Corollary 1 to concrete examples. We next present two cases where we apply it for well known families of reward distributions, with polynomically bounded competing bids.

Example 3.1: Beta-Bernoulli rewards and competing bids. Suppose the reward, conditional on ν, is distributed as $X|\nu \sim$Bernoulli(ν) and $\nu \sim$Beta(s_1, s_2). This can be interpreted as follows: the agent bids for an ad with an uncertain click-through-rate (or click probability) represented by ν; if a user clicks on the ad the advertiser obtains a fixed profit, normalized to 1. The Beta-Bernoulli example has been repeatedly adopted in the literature (in online advertising and in other areas of application - see for example [Li *et al.*, 2010] or [Hummel and McAfee, 2014]-), mainly for tractability purposes. Under this assumptions, $\mu_\nu = \nu$, $\mathbb{E}[\nu] = s_1/(s_1+s_2)$ and $\mathrm{Var}(\nu) = E[\nu](1-\mathbb{E}[\nu])/(s_1+s_2+1)$. After observing $X_1, \ldots, X_t$ the Bayesian updates are given by

$$\begin{aligned} s_1' &:= s_1 + \sum_{k=1}^{t} X_k, \\ s_2' &:= s_2 + t - \sum_{k=1}^{t} X_k, \end{aligned}$$

which implies

$$\begin{aligned} \mu(\pi_t^o) &= \frac{s_1 + \sum_{k=0}^{t} X_k}{s_1+s_2+t}, \\ \mathrm{Var}_{\nu\sim\pi_t^o}(\mu_\nu) &= \frac{\mu(\pi_t^o)(1-\mu(\pi_t^o))}{s_1+s_2+t} \le \frac{1}{s_1+s_2+t} = O(1/t). \end{aligned}$$

Suppose that the competing bidders' rewards are also Beta-Bernoulli distributed. If we further assume that they play myopically with respect to their beliefs, the competing bid distribution will be given by $q(x) := G^{j-1}(x)$, where $G(\cdot)$ is a the cdf of a Beta distribution. It is not to hard to show, using elementary calculus, that $1/G(\cdot)$ is polynomically bounded and hence $1/q(x) = O\left(x^{-\alpha}\right)$ for some $\alpha \in \mathbb{R}$. Let us further assume that $s_1 > \alpha$ and note that $x^{-\alpha}$, which bounds $1/q(x)$, is a convex function. Furthermore,

$$\mathbb{E}_{(s_1,s_2)}[\nu^{-\alpha}] := \int_0^1 \nu^{-\alpha}\frac{\nu^{s_1-1}(1-\nu)^{s_2-1}}{\Gamma(s_1,s_2)}d\nu = \int_0^1 \frac{\nu^{s_1-\alpha-1}(1-\nu)^{s_2-1}}{\Gamma(s_1,s_2)} < \infty.$$

Therefore, if $s_1 > \alpha$, $\pi(\cdot)$ and $q(\cdot)$ in this example satisfy the conditions of Corollary 3.1, which implies

$$\begin{aligned} V_T^o(\pi) - V_T^m(\pi) &= O(\log(T)), \text{ and} \\ \mathbb{E}\left[V_T^o(\pi_k^o) - V_T^m(\pi_k^o)\right] &= O(1/k). \end{aligned}$$

Example 3.2: Gamma-Weibull rewards with polynomically bounded competing bid distribution. Suppose $X|\nu \sim \mathrm{Weibull}(\nu,\ell)$, and $\nu \sim \mathrm{Gamma}(a,S)$, where ℓ is known. This implies that $\mu_\nu \propto 1/\nu$, $E[\mu_\nu] \propto S/(a-1)$ and $\mathrm{Var}(\mu_\nu) \propto E[\mu_\nu]^2/(a-2)$, for $a > 2$, where the proportionality symbol denotes that the quantity is multiplied by some function of ℓ only. The prior and likelihood distributions are given by

$$\begin{aligned} \pi(\nu|a,S) &:= \frac{S^a \nu^{a-1} e^{-S\nu}}{\Gamma(a)}, \\ f(x|\nu,\ell) &:= \nu\ell x^{\ell-1}\exp\{-\nu x^\ell\}. \end{aligned}$$

After t observations $X_1, \ldots, X_t$, the Bayesian updates are given by

$$\begin{aligned} S' &= S + \sum_{k=1}^{t} X_k^{\ell}, \\ a' &= a + t, \end{aligned}$$

which implies

$$\begin{aligned} \mu(\pi_t^o) &= \frac{S + \sum_{k=0}^{t} X_k^{\ell}}{a + t - 1}, \\ \mathrm{Var}_{\nu\sim\pi_t^o}(\mu_\nu) &= \frac{\mu(\pi_t^o)^2}{a + t - 2}. \end{aligned}$$

In principle it is not possible to bound $\mathrm{Var}_{\nu\sim\pi_t^o}(\mu_\nu)$, as $\mu(\pi_t^o)$ is an unbounded random variable. Note, however, that if we assume $1/q(x) = O(x^{-\alpha})$, which implies[3]

$$\frac{\mathrm{Var}_{\nu\sim\pi_t^o}(\mu_\nu)}{q(\mu(\pi_t^o))} = O(\mu(\pi_t^o)^{2-\alpha})\frac{1}{a + t - 2},$$

by simply decreasing the value of α by two we can proceed as if $\mathrm{Var}_{\nu\sim\pi_t^o}(\mu_\nu) = O(1/t)$.

Next, note that

$$\mathbb{E}[\mu_\nu^{-(\alpha-2)}] \propto \mathbb{E}[\nu^{\alpha-2}] := \int_0^\infty \frac{S^a \nu^{a+\alpha-3} e^{-S\nu}}{\Gamma(a)} < \infty, \text{ if } a + \alpha > 2$$

and hence, we can apply Corollary 3.1 to conclude that

$$\begin{aligned} V_T^o(\pi) - V_T^m(\pi) &= O(\log(T)), \text{ and} \\ \mathbb{E}\left[V_T^o(\pi_k^o) - V_T^m(\pi_k^o)\right] &= O(1/k). \end{aligned}$$

3.2.1 Decoupling rewards: Learning click rate and user value

Our development so far, and the two examples presented above, are based on the premise that there is a reward associated with *winning* the auction, which we denote by X. In a pay-per-impression online auction environment this is equivalent to assuming that the reward is associated with the *impression* of an ad, whereas, in reality, advertisers benefit only if the user clicks on it. One can embed this feature to our model by aggregating the

[3]Similarly to Example 3.1 it is not hard to show that $\mu_\nu = 1/\nu \sim$InverseGamma$(a, 1/S)$ has a polynomically bounded cdf and hence in this case also, if the competing bidders have the same prior as the agent and they bid myopically, there exists α such that $1/q(x) = O(x^{-\alpha})$.

click event and the reward collected from a user in a single random variable. By letting $X := \epsilon Y$, where $\epsilon \in \{0, 1\}$ is the click event and $Y \geq 0$ is the *value* of a visiting user, one can apply our model and results so far without changes.

However, with this approach one is implicitly assuming that what the agent observes upon winning an auction is his compound reward X, and in principle he is unable to distinguish a user that did not click the ad from one that clicked on it, but produced no revenue; both events signal $X = 0$. In practice, however, advertisers can observe whether a user clicks on a displayed ad or not, and they can separately estimate what is the value associated with a visitor.

We propose an extension to our model that, with mild modifications, can account for the fact that click rates and user value can be estimated separately. Essentially, one needs to decouple the reward associated with an ad impression from the feedback that the advertiser gets by winning the auction. Consider again the formulation of our model in Section 3.1, and suppose that X_t represents what the agent *observes* when he wins the auction, and let $R(X_t)$ be the reward associated with this observation. In the online advertising setting one would define $X_t := (\epsilon_t, \epsilon_t Y_t)$ and $R(X_t) := \epsilon_t Y_t$. The single auction reward now becomes

$$\mathbb{E}\left[\mathbb{I}_{\{b \geq \widehat{b}\}}\left(R(X) - \widehat{b}\right)\right].$$

Note that the information collection process is unchanged with respect to the original formulation, except that now X and ν are (potentially) multidimensional random variables. Therefore, the expression for the optimal and myopic value functions are identical to (3.1), (3.2) and (3.4), except for the single period reward, which is replaced by the expression above. If we further denote

$$\begin{aligned}\mu_\nu &:= \mathbb{E}[R(X)|\nu] \\ \mu(\pi) &:= \mathbb{E}[R(X)],\end{aligned}$$

the statement and the proof of Theorem 3.1 and Corollary 3.1 are also unchanged.

We now illustrate with an example how this extension can be used to model a more realistic online advertising setting, as discussed above.

Example 3.3: Beta-Bernoulli click rate with Exponential user value. Let $\epsilon \in \{0, 1\}$ denote the event of a user clicking on a displayed ad, and let $\nu_\epsilon := \mathbb{P}\{\epsilon = 1\}$, where $\nu_\epsilon \sim$Beta(s_1, s_2). Let us denote by Y the value associated with a user that clicked on the ad, and assume that $Y \sim$Exponential(ν_Y), with $\nu_Y \sim$Gamma(a,S). We then define $\nu := (\nu_\epsilon, \nu_Y)$, $X := (\epsilon, \epsilon Y)$ and $R(\epsilon, Y) := \epsilon Y$. Assuming that the click event and the user value distributions are independent one can write

$$\begin{aligned}
\mu_\nu &:= \mathbb{E}[R(X)|\nu] = \nu_\epsilon \, \nu_Y^{-1} \\
\mu(\pi) &:= \mathbb{E}[R(X)] = \mathbb{E}[\epsilon]\mathbb{E}[Y] = \frac{s_1}{s_1+s_2}\frac{S}{a-1}
\end{aligned}$$

Note that now $\pi(\nu)$ is a bi-dimensional parameter distribution, with hyperparameters (s_1, s_2, a, S). After observing $X_1 = (\epsilon_1, \epsilon_1 Y_1), \ldots, X_t = (\epsilon_t, \epsilon_t Y_t)$, the updates are given by

$$\begin{aligned}
s_1' &= s_1 + \sum_{k=1}^{t} \epsilon_k \\
s_2' &= s_2 + t - \sum_{k=1}^{t} \epsilon_k \\
a' &= a + \sum_{k=1}^{t} \epsilon_k \\
S' &= S + \sum_{k=1}^{t} Y_k \epsilon_k
\end{aligned}$$

Because in this case it is not possible to show that the posterior variance is $O(1/t)$ for any realization of $X_1, \ldots, X_t$, we cannot apply Corollary 3.1 directly. One can, however, apply similar arguments to arrive to the same conclusion: if the competing bid $q(\cdot)$ is such that $1/q(x) = O(x^{-\alpha})$, with $a + \alpha > 2$ and $s_1 > \alpha + 1$, then

$$\mathbb{E}_{X_1^t}\left[q^{-1}(\mu(\pi_t^o)) \operatorname{Var}_{\widehat{\nu}\sim\pi_t^o}(\mu_{\widehat{\nu}})\right] = O(1/t).$$

This is shown in Lemma B.2 in Appendix B. We thus have shown that for this case also,

$$\begin{aligned}
V_T^o(\pi) - V_T^m(\pi) &= O(\log(T)), \text{ and} \\
\mathbb{E}\left[V_T^o(\pi_k^o) - V_T^m(\pi_k^o)\right] &= O(1/k).
\end{aligned}$$

3.2.2 First price auctions

So far we have studied our problem in the framework of second price auctions, where the optimal (myopic) bid is given by the current expected reward of the agent. Our approach, however, can be extended to account for other types of auctions. In particular, with mild modifications we can generalize our approach to work with the following auction reward structure:

$$\mathbb{I}_{\{b \geq \widehat{b}\}} \left(X - \psi(b, \widehat{b}) \right),$$

where $\psi(b, \widehat{b})$ are the payments of the agent, for a bidding profile $(b, \widehat{b})$. Here $\psi(b, \widehat{b}) := \widehat{b}$ would define a second price auction and $\psi(b, \widehat{b}) := b$ a first price one. A myopic bidder would then solve the following problem:

$$\max_{[0,\bar{b}]} \mathbb{E} \left[\mathbb{I}_{\{b \geq \widehat{b}\}} \left(X - \psi(b, \widehat{b}) \right) \right] = \max_{[0,\bar{b}]} q(b)\mu(\pi) - p(b),$$

where, as before, $\mu(\pi) := E[X]$, and $p(b)$ is now more generally defined as $\mathbb{E}[\mathbb{I}_{\{b \geq \widehat{b}\}} \psi(b, \widehat{b})]$. We denote by $b^m(\mu(\pi))$ the (myopic) optimal bid that solves the problem above. Note that the structure of the single period auction is very similar to the one defined in (3.3), except that now the mypoic bid as a function of $\mu(\pi)$ is no longer defined as $b^m(\mu(\pi)) := \mu(\pi)$.

All the results and examples of sections 3.1 and 3.2 (which, for the purpose of brevity, we will not restate and prove here) follow directly from replacing the optimal myopic bid by $b^m(\mu(\pi))$ and the myopic probability of winning $q(\mu(\pi))$ by $q(b^m(\mu(\pi)))$. The assumptions in Theorem 3.1 and Corollary 3.1 are now stated with respect to the function $q \circ b^m(\cdot)$, rather than $q(\cdot)$.

Consider now the case of a first price auction, where $b^m(\mu(\pi))$ is defined as the solution to the following equation on b:

$$q'(b)(\mu(\pi) - b) = q(b).$$

Note that the definition of $b^m(\cdot)$ is now dependent on $q(\cdot)$ and might not lead to a closed form in general. However, one can still apply Corollary 3.1 as in the examples of this chapter, by simply assuming a slightly stronger condition on $q(\cdot)$. In Lemma B.3 in Appendix B we show that, if $q'(x) = \Theta(x^{\alpha-1})$ as $x \to 0$ for some $\alpha \in \mathbb{R}$, then $1/q \circ b^m(x) = O(x^{-\alpha})$.[4] Therefore, we can replicate the arguments of Examples 3.1, 3.2 and 3.3 and reach the conclusion that, for first price auctions as well, the myopic loss grows at most logarithmically in time and decreases at rate $1/k$ if the prior variance decreases at that rate.

[4] A function $g(x) = \Theta(\varphi(x))$ as $x \to a$ if $C_2\varphi(x) \leq g(x) \leq C_1\varphi(x)$ for all x in a neighborhood of a and some fixed constants C_1 and C_2.

3.3 Numerical Examples

In this section we perform numerical experiments to evaluate the myopic performance and compare it to that of the optimal policy. For computational reasons, we focus on second price auctions with Beta-Bernoulli distributions, as in Example 3.1. That is, we let $X \sim$Bernoulli(ν), with $\nu \sim$Beta(s_1, s_2). We consider a competing bid $\widehat{b}$ defined as the maximum of $k-1$ iid competing bids, each of which is distributed according to a Beta$(\widehat{s}_1, \widehat{s}_2)$ distribution. The competing bid cdf is thus given by $q(x) := G^{k-1}(x)$, where G is the cdf of a Beta$(\widehat{s}_1, \widehat{s}_2)$ distribution.

We consider three scenarios for the competing bid distribution: $(\widehat{s}_1, \widehat{s}_2) = (1, 3)$, $(\widehat{s}_1, \widehat{s}_2) = (2, 2)$ and $(\widehat{s}_1, \widehat{s}_2) = (3, 1)$, and three auction sizes: $k = 3$, $k = 4$ and $k = 5$. We focus on cases that satisfy the conditions of Corollary 3.1, which in this setting translates into requiring that $s_1 > \alpha := (k-1)\,\widehat{s}_1$. In each case, we let $(s_1, s_2) := (k\,\widehat{s}_1, k\,\widehat{s}_2)$, implying that $\mathbb{E}[\nu] = \mathbb{E}[\widehat{b}^i]$, where $\widehat{b}^i$ represents the bid of an individual competing bidder. This means that the agent's myopic bid in the first auction is equal to the myopic bid of the competing players.

For each combination of $\widehat{s} := (\widehat{s}_1, \widehat{s}_2)$ and k we compute the *myopic optimality gap* (MOG), defined as $(V_T^* - V_T^m)/V_T^m$, as a function of the time horizon T. The results are depicted in figure 3.3. The plots on the left side of Figure 3.3 show the MOG values for the different experiments, while the right side plots depict the corresponding cumulative distribution functions of the prior distribution and the competing bid, $\Pi(\cdot)$ and $q(\cdot)$, respectively.

The results show that the myopic optimality gap is low in most examples, and it grows as the number of players in the auction increases: for larger k, the competing bid distributions (stochastically) grows, and the winning probability under the myopic policy diminishes, as shown in the CDFs plots. Similarly, for fixed k, the worst case is given by $(\widehat{s}_1, \widehat{s}_2) = (3, 1)$. Even though in this case the starting mean reward of the agent is higher, so is the mean reward of the competitors and as a result the winning probability decreases.

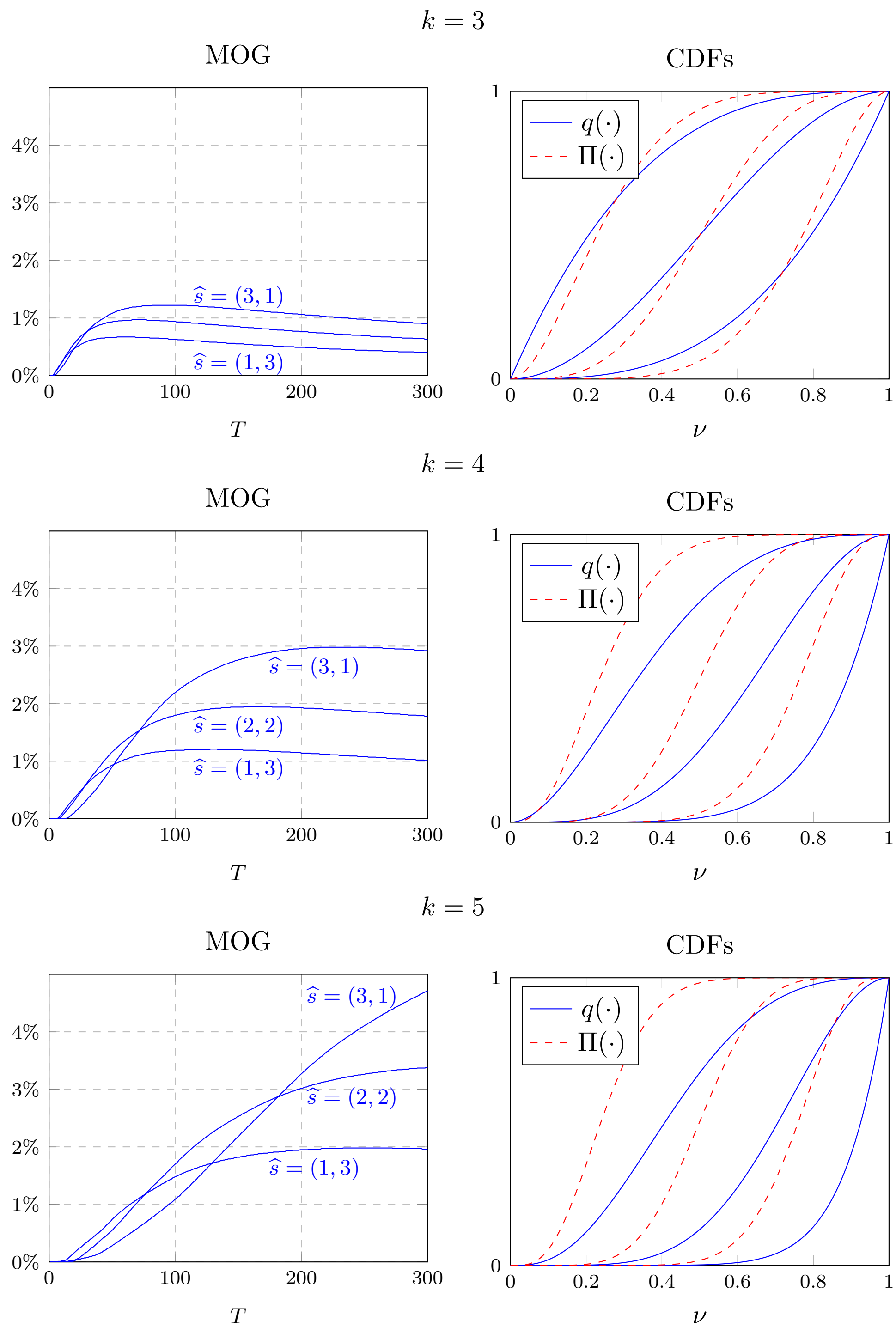

Figure 3.2: **Numerical examples results.** On the left, the myopic optimality gaps as a function of T for different distributional settings. On the right, the corresponding cumulative distribution functions.

3.4 Concluding Remarks

In the present chapter, we study the implications of incomplete private information in sequential auctions. We analyze the problem from an agent's perspective, seeking to maximize his rewards as he learns his own valuation of the auctioned item.

We study the exploration-exploitation trade-off in this problem an find that, even though in general terms it cannot be ignored, there exists a wide range of problem settings where bidding myopically with respect to information collection produces near optimal results. We characterize conditions under which this is indeed the case, and we are able to show that the myopic optimality gap decreases at a fast rate as the length of the time horizon grows and as the initial uncertainty on the item valuation is reduced. Our results are developed in a general framework that includes both first-price and second price auctions, and well known families of distributions.

Chapter 4

Estimating inaccurate inventory with transactional data

4.1 Model Specification

We consider the stock evolution of a product sold by a retailer, consumed by customers arriving to the store, and periodically replenished by a supplier. The retailer does not track the available inventory of the product, except for potential periodic inspections, which are assumed to be infrequent. We assume that the retailer can reliably record *sales* of the product in a database (with, for example, a bar-code scanner in the cash register). We also assume that the data contains, or can be used to generate, *non-purchase* events, that is, events where a customer arrived to the store but left without purchasing the product (though it is not known whether this is due to a stockout of the product or the customer simply deciding not to purchase it). In Section 4.1.2 we explain in more detail how to extract non-purchase events from transactional data. The input data can then be summarized as a discrete sequence of *transaction events*: purchases and non-purchases of the product under consideration.

We model the stock evolution of the product as a finite state, discrete time hidden Markov chain, with a finite time horizon of $T + 1$ periods. Each state of the chain represents a stock level, with 0 representing the out of stock state, and S the maximum stock. The transitions in the chain correspond to events in the stock evolution: purchases make the chain transition downwards and replenishments make the chain transition upwards. Every time there is a transition, the system reveals whether it corresponds to a purchase transaction event or a non-purchase transaction. The sequence of transaction events is denoted by $d = (d_1, \ldots, d_T) \in \{\mathbf{0}, \mathbf{1}\}^T$, where purchases are denoted by $\mathbf{1}$, and non-purchases by $\mathbf{0}$. For expository purposes, in what follows we consider d as the only available data for the estimation, and in Section 4.2.2 we explain how the model can be modified to incorporate direct inventory information. The sequence of (hidden) states of the chain is denoted by $h = (h_0, h_1, \ldots, h_T)$.

We denote by $Q \in \mathbb{R}^{S+1,S+1}$ the transition matrix of the chain, and by $E \in \mathbb{R}^{S+1,S+1,2}$ the emission matrix of the system, which defines the probability of emitting $\mathbf{0}$ or $\mathbf{1}$ for each transition. That is,

$$\begin{aligned} Q_{i,j} &:= \mathbb{P}\left\{h_{t+1} = j \mid h_t = i\right\}, \\ E_{i,j,k} &:= \mathbb{P}\left\{d_t = k \mid h_{t-1} = i, h_t = j\right\}. \end{aligned}$$

The structure of Q and E can have a deep impact in the quality of the estimation, and the usefulness of the method in general. This is particularly relevant in a case where one only has access to a very limited amount of information, and there are potentially numerous underlying sets of parameters that can explain the data. In the next section, we define the parameters that govern the inventory evolution of our model, and how Q and E can be constructed as a function of them.

4.1.1 Construction of the transition and emission matrices

We denote by θ the probability that a customer arriving to the store purchases an item, when there is availability. For simplicity, we assume that every customer can purchase at most one unit of the product under consideration, though our model can be extended to account for multiple item purchases, if that information is available in the data. The supplier visits the store periodically to replenish the stock of the product. In order to make the estimation procedure more efficient (i.e. preserve the conjugate property of the parameters prior distribution) we further assume that the probability of having a purchase and a replenishment simultaneously is negligible, and hence replenishments always correspond to non-purchase events.[1] There is a fixed probability of a supplier arrival at each non-purchase event, denoted by μ. Figure 4.1 depicts an example of the hidden Markov chain, with transition matrix Q constructed as a function of θ and μ. In this case, every transition emits either **0** or **1** with probability 1 and hence all the components of E are in $\{0, 1\}$. Note that

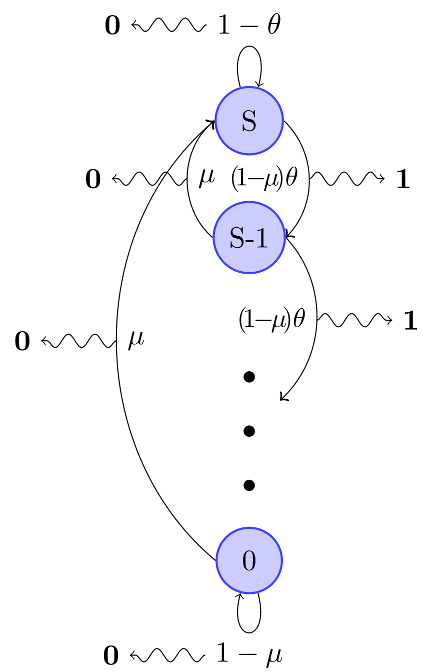

Figure 4.1: Example of the hidden Markov chain (self transition of state $S-1$ is omitted for display purposes).

in the out of stock state ($h_t = 0$) purchase events are not observable and, similarly, in the maximum stock state ($h_t = S$) replenishments are not observable.

Record inaccuracies: One issue with the Markov chain defined in Figure 4.1 is that downward transitions are potentially too restrictive: the stock goes down if and only if the system records a purchase of the product. If the assumptions made so far are a good approximation to the process that generates the data, this would not necessarily represent a problem. In practice, however, one would expect both complex inventory dynamics and

[1] In most applications, non-purchase events are substantially more common than purchase events and hence this represents a mild assumption.

record inaccuracies, and hence such a restrictive condition enforced on the model can lead to poor estimations results. To see why this might be the case, consider, as an example, the case where the system records an unusually long sequence of non-purchases; this could be due to a delay on the supplier side, an unexpected increase in demand, failure to retrieve items from a warehouse, or any other circumstantial cause. In such a case, one would expect the method to categorize the state of the system as being effectively out of stock. However, conditional on observing a long sequence of $\mathbf{0}$'s, and based on the hidden Markov chain defined in Figure 4.1, the procedure will infer that the system is eventually in a high level of stock, because no purchase were recorded, and a replenishment is likely to arrive at some point.

We propose a simple way to address this issue, and at the same time make the model more flexible and robust to noise in the data. In particular, we allow for a fraction of the purchases not to be observed by the system, that is, to be recorded as non-purchases. We do this by defining a new parameter ϵ that represents the probability of the system correctly recording a purchase. In other words, when there is a purchase, there is a probability $1-\epsilon$ that it will be missclassified as a non-purchase. With this in mind, the emission matrix E is defined as

$$\begin{aligned} E_{i,i-1,\mathbf{1}} &= \epsilon &&, \text{for } i = 1,\ldots,S \\ E_{i,i-1,\mathbf{0}} &= 1-\epsilon &&, \text{for } i = 1,\ldots,S \\ E_{i,j,k} &= 0 &&, \text{otherwise.} \end{aligned}$$

Beside making the model more flexible, the inclusion of ϵ as a parameter has a concrete interpretation: it represents the fraction of unregistered or "invisible" demand in the system. This is tightly related to the literature on inaccurate stock optimization, where the rate of invisible demand is assumed to be known. The estimation procedure can then also be regarded as a way to generate the input parameters of an inaccurate inventory optimizer.

Restocking quantity uncertainty: In the example shown in Figure 4.1 all the replenishes increase the stock to its maximum value of S units, which is assumed to be known. In practice, it is likely that there exist some level of uncertainty about either the value of the maximum stock S, or the quantity that is added to the stock when a replenishment occurs. We incorporate these sources of uncertainty in our model in the following manner: we assume that, conditional on a replenishment occurring, the new stock level is drawn independently from an unknown distribution, to be estimated by the procedure. More precisely, let us denote by $\underline{\mu}$ the minimum possible restocking value, and

$$\mu_i := \mathbb{P}\{h_t = i \mid h_{t-1} < \underline{\mu} \text{ and a replenishment occurs}\},$$

for $i = \underline{\mu}, \ldots, S$.

We assume that restocking events strictly increase the inventory level.[2] The restocking probabilities for every state of the chain can then be written as

$$\mathbb{P}\{\text{restocking level} = k \text{ when } h_t = i\} = \begin{cases} \mu_k & \text{if } i < \underline{\mu} \leq k \\ \dfrac{\mu_k}{\mu_{i+1} + \ldots + \mu_S} & \text{if } k > i \geq \underline{\mu} \\ 0 & \text{otherwise.} \end{cases}$$

The second line above states that the probability of restocking to k, when the inventory level is above $\underline{\mu}$, is the conditional probability of restocking to k, given that restockings are always strictly incremental.

The inclusion of $\{\mu_i\}$ gives the estimation procedure the capability of resolving potential uncertainty regarding the maximum stock level or the replenishment policy, very likely to exist in practice. Consider, for example, a case where one is uncertain about the value of S. Suppose, however, that it is possible to guess a reasonable range where S is likely to be, that is, $S \in [\underline{S}, \ldots, \overline{S}]$. Then, by setting $S := \overline{S}$ and $\underline{\mu} := \underline{S}$, the procedure will estimate the actual value of S using the available data. It is also possible that there is some inherent noise in the replenishment process, and hence there is no *true* restocking value to be estimated. In such a case, the output values of $\{\mu_i\}$ will reflect this, by defining a *restocking distribution.*

Figure 4.2 (partially) depicts the hidden Markov chain with invisible sale transactions and

[2] This assumption is inconsistent with some common inventory policies, such us the well known (s, S) policy. The reason to enforce this restriction is simplicity and computational efficiency, as discussed in Section 4.2.1.

restocking quantity uncertainty. The full definition Q is given by

$$Q_{ij} = \begin{cases} (1-\mu)(1-\theta) & \text{if } \; 0 < i < S \quad j = i \\ (1-\mu)\theta & \text{if } \; 0 < i < S \quad j = i-1 \\ \mu\, \mu_j & \text{if } \; 0 \le i < \underline{\mu} \quad j \ge \underline{\mu} \\ \mu \dfrac{\mu_j}{\mu_{i+1} + \ldots + \mu_S} & \text{if } \; \underline{\mu} \le i < S \quad j \ge i+1 \\ 1-\mu & \text{if } \; i = 0 \quad j = 0 \\ \theta & \text{if } \; i = S \quad j = S-1 \\ 1-\theta & \text{if } \; i = S \quad j = S \\ 0 & \text{otherwise} \end{cases}$$

As a summary of this section, Table 4.1.1 presents the list of all the parameters considered

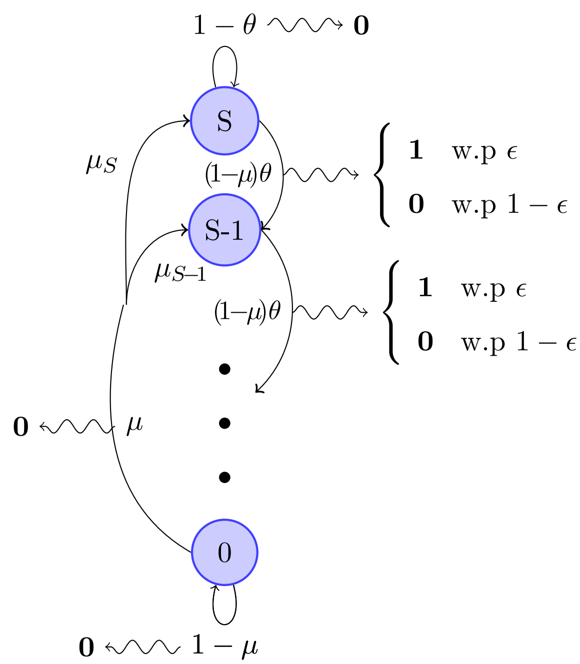

Figure 4.2: Example of the hidden Markov chain with invisible sale transactions and restocking quantity uncertainty.

in our model.

Parameter	Description
θ	Purchase probability.
ϵ	Probability of the system correctly recording a sale.
μ	Probability of supplier arrival (replenishment).
$\{\mu_i\}_{i=\underline{\mu},\ldots,S}$	Distribution of replenishment quantities.

Table 4.1: Parameters of the model.

4.1.2 Defining non-purchase events

We now turn our attention to the definition of non-purchase events. Non-purchase events should be broadly interpreted as a measure of time between purchases of the product: a long sequence of $\mathbf{0}$'s represents a long period of time where no purchase was recorded, and hence a signal that the product might be out of stock. There is more than one way of defining non-purchase events, depending on the available data. If, for example, the data contains purchases of the product with timestamps, under some mild assumptions the non-purchases can be randomly generated as a function of the inter-purchase time. We propose a simpler approach, based on aggregate store purchases, which is easy to implement in practice and robust with respect to demand fluctuations and trends.

It is reasonable to assume that a system which records a retailer's sales does so for most products in the store. Our approach regarding non-purchase events is to define them as the aggregate purchases of the store, excluding those of the product under consideration. That is, the purchase events ($\mathbf{1}$) are defined as the purchases of the product, and the non-purchase events ($\mathbf{0}$) are defined as the purchases of other products in the store. The rationale is that the purchase probability θ can be interpreted as the fraction of the customers of the store that want the product under consideration. Note that, even though we are assuming that this fraction remains constant in time, the demand rate itself is not assumed to be constant. That is, demand can be subject to any type of trend or seasonal effects and the estimation methodology will produce equally effective predictions. Formally speaking, we are assuming that aggregate demand follows a Poisson process with an arbitrary rate function $\lambda(t)$, and hence demand for the product under consideration follows a Poisson process with rate $\theta\lambda(t)$. By looking only at transaction events, the method can estimate the parameters of the model regardless of the definition of $\lambda(t)$. This is an effective way of decoupling the demand estimation process into two steps, which can be approached separately: the *aggregate demand estimation* step and the *product level* estimation step. The aggregate demand estimation step (the estimation of $\lambda(t)$) can be performed with traditional estimation techniques, such us time series analysis or smoothing. At this level, using POS data can be a good approximation to demand, since aggregate demand is less likely to be affected by the stock level of individual products. The product level estimation step, the focus of our work, requires more careful analysis because sales and demand can no longer be regarded as equivalent. It is important to note, however, that the aggregate demand estimation is not required for our proposed estimation methodology to be carried out. Moreover, the output of the estimation at a product level can be applied in practice as a prediction methodology *without* the need of the aggregate demand parameters. To see how, first note that once the sequences of **1** events (purchases of the product) and **0** events (purchases of other products) are defined, the method can jointly estimate both the historical (unobservable) stock levels and the parameters of the model. Once this estimation is performed, with data up to the time of the execution, the practitioner can use the current estimated stock distribution of the product and the parameters of the model to build a (probabilistic) path of the stock evolution as new events are observed. However, note that this path depends on the sequence of **0**'s and **1**'s only, and not on the actual timing of these events. If, for example, the system observes a large number of contiguous **0**'s, it will likely predict the item to be out of stock, regardless of when those non-purchases were recorded. In this way, the retailer can have an effective way of tracking stock levels in time, without the need of actual time information, by simply observing the sales of the store.

4.2 Estimation Methodology

We propose a Bayesian estimation method, based on the well known Markov Chain Montecarlo family of estimation algorithms. In a Bayesian setting, the parameters of the model are regarded as random variables, endowed with a prior distribution that summarizes the initial uncertainty that exists around them. The output of the estimation is given by the posterior distribution of the parameters, conditional on the observed data. We propose a Gibbs Sampling method to perform the estimation.

As opposed to traditional non-Bayesian estimation methodologies, such as maximum likelihood, the output of the Gibbs Sampler is not a set of definite estimates but rather a *sample* from a posterior (output) distribution. By computing particular statistics of the posterior sample, such as the sample mean or mode, one can obtain concrete estimators for the parameters of the model. Similarly, by computing dispersion statistics, such as the sample variance of the posterior distribution, one con measure the accuracy of such estimators.

Figure 4.3 describes the Gibbs Sampler. The method works by alternatively sampling from the hidden states of the Markov chain, given the observed data and the current parameters, and from the model parameters, given a sample path of the chain and the data. This is described by steps 1 and 2 in the loop of Figure 4.3. The samples generated by the Gibbs Sampler form a Markov Chain, whose stationary distribution can be shown to be the desired posterior distribution of the parameters, given the observed data. It can also be shown that, under mild conditions, the chain converges to its stationary distribution, though it is commonly hard to bound the number of steps needed to guarantee the mixing. In practice, most applications simply rely on a *burnout* period, where the first samples are discarded. In our numerical examples we will follow this approach.

- **Input:** $d \in \{\mathbf{0}, \mathbf{1}\}^T$
- **Initialization:**

  Set arbitrary $(\theta, \epsilon, \mu, \{\mu_i\})^1 \in (0,1)^{S-\underline{\mu}+4}$
- **Loop:** for $n = 1, \ldots, N$
  1. Sample $h^n \sim p\,(h\,|(\theta, \epsilon, \mu, \{\mu_i\})^n, d)$
  2. Sample $(\theta, \epsilon, \mu, \{\mu_i\})^{n+1} \sim p\,(\theta, \epsilon, \mu, \{\mu_i\}\,|h^n, d)$.
- **Output:** $\{h^n\}, \{(\theta, \epsilon, \mu, \{\mu_i\})^n\}$

Figure 4.3: Gibbs Sampling Method.

4.2.1 Posterior sampling

We now turn our attention to the sampling step of the Gibbs Sampler, and describe in more detail steps 1 and 2 of the loop in Figure 4.3. We start by describing step 2: sampling from the posterior distribution of the model parameters.

Posterior sampling of parameters. The main input to this step is the prior distribution of the parameters. In particular, we define the joint prior distribution as $S - \underline{\mu} + 4$ independent Uniform(0,1) distributions. That is,

$$p(\theta, \epsilon, \mu, \{\mu_i\}) := \mathbb{I}_{\left\{(0,1)^{S-\underline{\mu}+4}\right\}},$$

where $p(\cdot)$ denotes the prior distribution, and $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function.

The choice of uniform priors, a common one in the Bayesian HMM literature, reflects maximum initial uncertainty about the model parameters. It also provides a good choice from a computational efficiency point of view. Note that step 2 in the loop of Figure 4.3 requires sampling from the posterior distribution of $(\theta, \epsilon, \mu, \{\mu_i\})$ *conditional* on a sample path h^n. This greatly simplifies the sampling process: if the chain is observable, and because of the Markovian property, every visit to a particular state provides and independent realization of the event of the chain "choosing" one outwards transition over the others. To illustrate how this fact facilitates the computation of the posteriors consider, for example, the purchase events. Let x be the number of times that the chain visited a state where a purchase is observable (i.e. non-zero stock levels), and y the number of observed purchases. That is,

$$\begin{aligned} x &:= |\{t : h_t > 0\}|, \\ y &:= |\{t : h_{t+1} = h_t - 1\}|. \end{aligned}$$

The likelihood of the observed transitions is given by

$$p(h|\theta) = \theta^y (1-\theta)^{x-y},$$

and hence the posterior distribution of θ is given by

$$p(\theta|h, d) = p(\theta|h) = \frac{p(h|\theta)p(\theta)}{\int_0^1 p(h|\widetilde{\theta})p(\widetilde{\theta})d\widetilde{\theta}} = \frac{\theta^y(1-\theta)^{x-y}}{\int_0^1 \widetilde{\theta}^y(1-\widetilde{\theta})^{x-y}d\widetilde{\theta}} \sim \text{Beta}(1+y, 1+x-y).$$

This implies that, if the path of the chain is observable, the posterior distribution of θ belongs to a known family of distributions, and is easy to sample from[3]. A similar approach can be used to calculate the posterior distribution of the rest of the parameters. However, in order to preserve their conjugate structure, some extra assumptions are needed. In particular, it is necessary to avoid situations where two or more events are collapsed into a single chain transition. For example, suppose that for some state j (beside the maximum stock level S) the restocking quantity is equal to 0, that is, if a replenishment arrives the stock level remains unchanged. In such case there exist two events that trigger a self transition in j: a replenishment event and a non-purchase event. Suppose we "observe" one self transition in state j and want to compute the posterior distribution of θ conditional on the replenishment probability λ. The posterior is given by

$$p(\theta \mid h_t = h_{t-1} = j, \lambda) = \frac{\lambda + (1-\lambda)(1-\theta)}{\int_0^1 \lambda + (1-\lambda)(1-\widetilde{\theta}) d\widetilde{\theta}}.$$

Note that the resulting distribution does not belong to any well known family of distributions and hence the posterior sampling becomes difficult. In order to circumvent this issue we assume, as specified in section 4.1.1, that all the replenishment transitions (except at the maximum stock) *strictly increment the stock*.

Even under the assumption of strictly increasing replenishments, there exists one other model feature that threatens the conjugate property of the parameters. Recall from section 4.1.1 that if a state i is in the replenishment region, i.e. $i \geq \underline{\mu}$, the replenishment probability distribution is given by

$$\mathbb{P}\{\text{restocking level} = k \text{ when } h_t = i\} = \frac{\mu_k}{\mu_{i+1} + \ldots + \mu_S}, \quad \text{for } k \geq i+1.$$

Note that this likelihood form does not lead to a conjugate structure on the parameters $\{\mu_{\underline{\mu}}, \ldots, \mu_S\}$. One possible approach to assess this issue is to use some technique to sample from a non-standard family of distributions, such as the Metropolis-Hastings algorithm, and embed it in the Gibbs Sampler. We opt for a simpler solution that consists of ignoring a fraction of the data, but without introducing any bias in the estimation. Our approach works as follows: when sampling from the posterior distribution of $\{\mu_{\underline{\mu}}, \ldots, \mu_S\}$ we only

[3] This is due to the fact that *Uniform* and *Beta* are conjugate distributions.

consider states that lie outside the replenishment region, that is states $i < \underline{\mu}$. For these states, the replenishment distribution is given by

$$\mathbb{P}\{\text{restocking level} = k \text{ when } h_t = i\} = \mu_k,$$

and hence the conjugate structure is preserved. We stress the fact that ignoring a fraction of the data in this manner does not introduce a bias in the estimation, because the selection of the data is independent of the data itself. The only cost one has to incur is a less accurate estimation, with the benefit of a substantially more time efficient algorithm.

The precise definition of all the posterior parameter distributions and how to compute them is detailed in Appendix C.

Posterior sampling of states. Step 1 in the loop of the Gibbs Sampler involves sampling from the hidden states of the Markov chain, given the parameters and the observed data. Because the distribution of the states of the chain is conditional on the data, it is harder to generate a sample path from the chain.[4] Fortunately, there exists an efficient algorithm to perform this task, known as the Forward Filtering Backward Sampling (FFBS) algorithm (see, for example, [Scott, 2002]), which uses clever recursions to calculate the transition probabilities of the chain given the observed data. We next detail the derivation of the recursion formulas and the algorithm, specialized to our particular case.

The FFBS algorithm works by performing two passes. The first (forward) pass sequentially calculates the transition probabilities of the chain, given the data up to time $t-1$, using the transition probabilities up to time $t-2$. This is based on the following recursive formula: let

$$p_t(r,s) \quad := \quad \mathbb{P}(h_{t-1} = r, h_t = s \mid d_1^t), \tag{4.1}$$

$$\pi_t(s) \quad := \quad \mathbb{P}(h_t = s \mid d_1^t) = \sum_{r=0}^{S} p_t(r,s), \tag{4.2}$$

[4] This difference is similar to that of computing the likelihood of a particular sample path of a chain (calculated as a simple product) versus computing the likelihood of a sample path *given* some observed data, which is harder and requires clever computing techniques, such us the well known Viterbi algorithm.

where $d_1^t := (d_1, \ldots, d_t)$. Then

$$p_t(r,s) \quad = \quad \frac{\mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1})}{\sum_{r,s} \mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1})} = \frac{\pi_{t-1}(r)\, Q_{r,s}\, E_{r,s,d_t}}{\sum_{r,s} \pi_{t-1}(r) Q_{r,s} E_{r,s,d_t}}. \tag{4.3}$$

Based on these formulas, the forward pass calculates $p_{t,r,s}$ and $\pi_t(s)$ for all t, r, s. The backwards (sampling) pass starts at the end of the time horizon by sampling from $\pi_T(\cdot)$, and uses each sampled state to generate the previous one. The FFBS algorithm is described in Figure 4.4.

INPUT: d, Q, E, π_0.

FORWARD PASS:

- For each $t = 1, \ldots, T$
  - Set

$$\begin{aligned} p_t(r,s) &= \frac{\pi_{t-1}(r)\, Q_{r,s}\, E_{r,s,d_t}}{\sum_{r,s} \pi_{t-1}(r) Q_{r,s} E_{r,s,d_t}} \\ \pi_t(s) &= \sum_{r=0}^{S} p_t(r,s) \end{aligned}$$

BACKWARD PASS:

- Sample $\widehat{h}_T \sim \pi_T(\cdot)$
- For $t = T-1, \ldots, 0$
  - Sample $\widehat{h}_t \sim p_t(\cdot, \widehat{h}_{t+1})$

OUTPUT: $\{\widehat{h}_t\}_{t=0,\ldots,T}$

Figure 4.4: FFBS Algorithm.

4.2.2 Incorporating inventory information

The description of our methodology has so far been based on using purely transactional data. However, with minor modifications one can directly incorporate inventory information in the estimation. Suppose that the retailer performs physical inventory inspections. Let us denote by $(t_1, \ldots, t_K)$ the set of time periods where the inventory was inspected, and by $(\widehat{h}_{t_1}, \ldots, \widehat{h}_{t_K})$ the set of observed inventory levels. We can redefine equations (4.1) and (4.2) in the following way:

$$\begin{aligned} p_t(r,s) &:= \mathbb{P}(h_{t-1} = r, h_t = s \mid d_1^t, \widehat{h}_1^{t-1}), \\ \pi_t(s) &:= \mathbb{P}(h_t = s \mid d_1^t, \widehat{h}_1^{t-1}) = \sum_{r=0}^{S} p_t(r,s), \end{aligned}$$

where $\widehat{h}_1^t := \{\widehat{h}_{t_k} : t_k \leq t\}$ represents the set of inventory inspections up to time t.

To adapt the recursion given in (4.3), we consider two cases. First suppose there is an inventory inspection at time $t-1$, that is, $h_t = \widehat{h}_{t-1}$. Then

$$\begin{aligned} p_t(r,s) &= \frac{\mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1}, h_{t-1} = \widehat{h}_{t-1}, \widehat{h}_1^{t-2})}{\sum_{r,s} \mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1}, h_{t-1} = \widehat{h}_{t-1}, \widehat{h}_1^{t-2})} \\ &= \frac{\mathbb{I}_{\{r = \widehat{h}_{t-1}\}} \mathbb{P}(h_t = s, d_t \mid h_{t-1} = \widehat{h}_{t-1})}{\sum_s \mathbb{P}(h_t = s, d_t \mid h_{t-1} = \widehat{h}_{t-1})} \\ &= \frac{\mathbb{I}_{\{r = \widehat{h}_{t-1}\}} \, Q_{r,s} \, E_{r,s,d_t}}{\sum_s Q_{\widehat{h}_{t-1},s} \, E_{\widehat{h}_{t-1},s,d_t}}. \end{aligned}$$

Now, suppose there is no inventory inspection at time $t-1$, that is $\widehat{h}_1^{t-1} = \widehat{h}_1^{t-2}$. Then

$$\begin{aligned} p_t(r,s) &= \frac{\mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1}, \widehat{h}_1^{t-2})}{\sum_{r,s} \mathbb{P}(h_{t-1} = r, h_t = s, d_t \mid d_1^{t-1}, \widehat{h}_1^{t-2})} \\ &= \frac{\pi_{t-1}(r) \, Q_{r,s} \, E_{r,s,d_t}}{\sum_{r,s} \pi_{t-1}(r) Q_{r,s} E_{r,s,d_t}}. \end{aligned}$$

Therefore, by replacing this definitions of $p_t(r,s)$ and $\pi_t(r)$ in the FFBS algorithm of Figure 4.4, the inventory inspections are successfully incorporated in our methodology.

As a final remark of this section, note that the derivation above was specialized to events of the type $\{h_t = \widehat{h}_t\}$. It is not hard to modify the recursions to make them slightly more general, and incorporate events of the type $\{h_t \in A_t\}$, for some set of states A_t. As an example of how this could be useful, consider the case of "Zero Balance Walk" inspections, as studied in [Bensoussan *et al.*, 2007], where the inspections only reveal whether the product is in stock or not, that is $\{h_t > 0\}$. By generalizing the recursion, one can directly incorporate this type of partial inspections into the estimation procedure.

4.3 Numerical Experiments

In this section, we present numerical study results for our proposed methodology. We present an extensive simulation study, where we first evaluate performance with data generated from a perfectly specified model, and then we evaluate the method robustness with respect to several misspecifications. We finally present a numerical study with actual retailing data, and analyze the predictive power of our methodology in a real world environment.

4.3.1 Simulation Experiments

4.3.1.1 Perfect data specification examples

In the computational results of this section the data is generated according to the hidden Markov model used in our proposed methodology. Because the data is generated exactly following our assumptions, this examples can be used to assess the effectiveness of the method in estimating the true model parameters, under perfect modeling assumptions. From a practical point of view, we are also interested in understanding what is the impact of not having (frequent) access to the actual inventory level; in other words, how effective is this methodology in overcoming the fact that stock is only partially observable. To this end we consider a benchmark estimation methodology with the same model assumptions as before, but where the state of the chain is fully observable. This represents an ideal scenario with no stock tracking inaccuracies, and perfectly known inventory at every point in time. Recall from section 4.2.2 that our proposed methodology can incorporate stock inspection data, so one could potentially use our proposed Gibbs Sampler to carry out this benchmark estimation. However, for the case where the chain is fully observable the estimation of the parameters becomes easier. For example, following a similar Bayesian approach as before, the posterior distribution of the purchase probability θ is given by

$$p(\theta|d) = \text{Beta}(1 + y, 1 + x - y),$$

where $x := |\{t : h_t > 0\}|$ and $y := |\{t : h_{t+1} = h_t - 1\}|$. Note that in this case h_t is observable and hence there is no need to generate sample-paths from the chain. A similar argument can be applied to obtain posterior distributions of the rest of the parameters of the model.

Figure 4.5 shows the output of the first set of experiments. The data was generated using the following parameters: $S = 10$, $\theta = 0.5$, $\epsilon = 0.95$, $\mu = 0.05$, $\underline{\mu} = 8$ and $(\mu_8, \mu_9, \mu_{10}) = (0, 1, 0)$. The Gibbs Sampler was run for 5000 iterations, with the first 1000 discarded as a burnout period. Each plot in Figure 4.5 depicts the output of the estimation (expressed as 95% confidence interval) as a function of the time horizon (data size) T for the observable and unobservable stock cases. The dashed horizontal line represents the true value of the parameter.

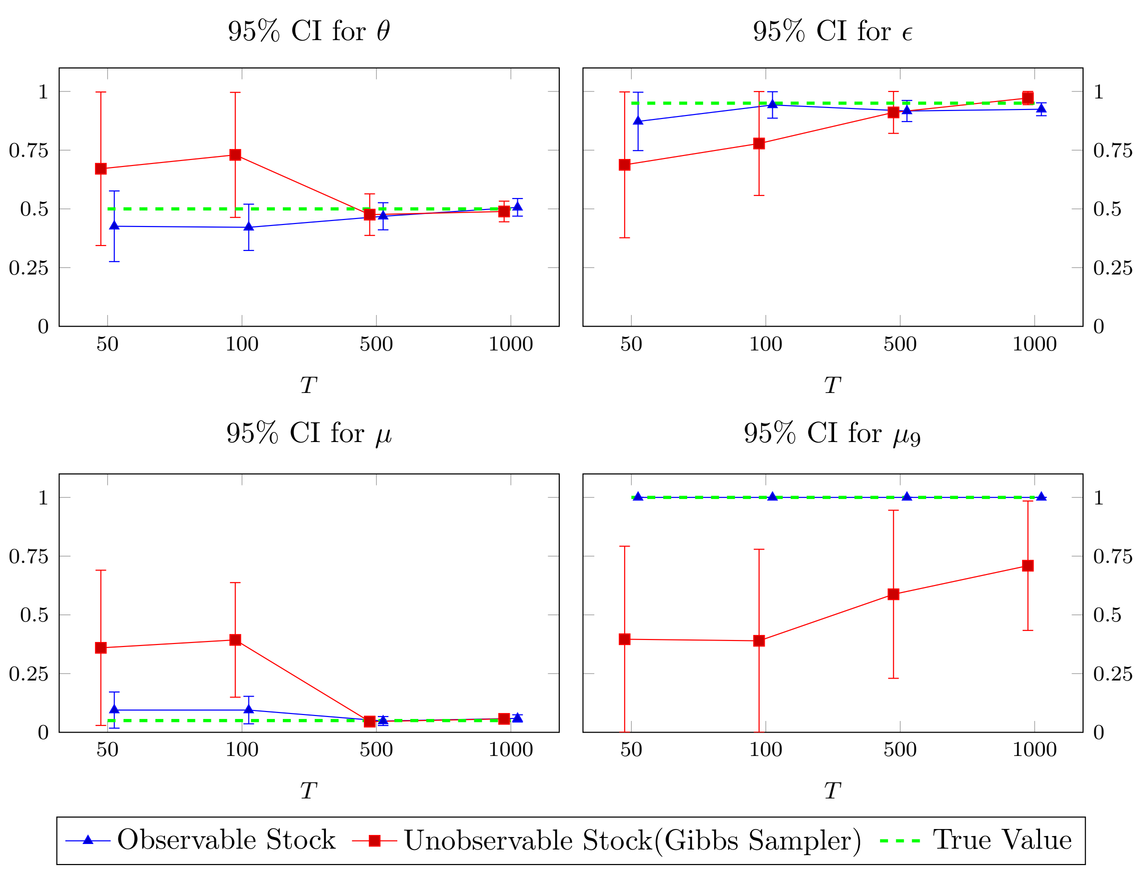

Figure 4.5: Estimation results with observable and unobservable stock as a function of the time horizon.

The results show that the unobservable stock makes the estimation harder when there is a reduced number of data points, but our methodology is able to overcome this difficulty and produce comparable results to the full observation case for most parameters. Note that the methodology gives a very good estimate with as few as 500 observations, except for the replenishment value probability μ_9. Recall from Section 4.2.1 that this behavior is expected, due to the approach we use to estimate $\{\mu_k\}$. Note, however, that this does not affect the quality of the estimation for the rest of the parameters.

The parameters of the model are one of the two main estimation targets, the other one is the unobservable states of the chain. Estimating the sample path of the chain is not only necessary for our procedure, but also very relevant from a practical point of view. Having a clear picture of the past stock levels can give the retailer very valuable insight on his current inventory management practices. Figure 4.6 shows estimation of the unobservable stock levels for the $T = 100$ and $T = 500$ examples of Figure 4.5. For display purposes, the inventory levels have been classified in three regions: low ($0 \leq h_t < S/3$), medium ($S/3 \leq h_t < 2S/3$) and high ($2S/3 \leq h_t \leq S$), denoted by R_L, R_M and R_H respectively. The vertical lines represent the posterior probabilities of the stock being in a particular region, while the horizontal thick line represents the actual stock level. One can readily observe how the estimation in the $T = 500$ case improves substantially over the $T = 100$ case on the entire time horizon, the former giving a very accurate picture of the unobservable sate of the chain in most periods. It is interesting to note how for the first 100 time periods, where the simulated data is equal in both cases, the $T = 500$ case gives substantially better results than the $T = 100$ case. This improvement comes only from adding extra *future* data, which is weakly coupled with the first periods, but nevertheless helps substantially improve the estimation.

4.3.1.2 Misspecified Simulation Examples

In this section we define a simulation model that more closely approximates a typical inventory system evolution. We aim to understand how well our methodology adapts to misspecifications in the data and, at the same time, assess its performance in a more realistic setting. We consider an inventory system with periodic inventory review and an (s, S) replenish policy. The provider of the product visits the store at fixed time intervals (e.g. weekly) and inspects the available stock of the product. If the stock is at or below s, it gets

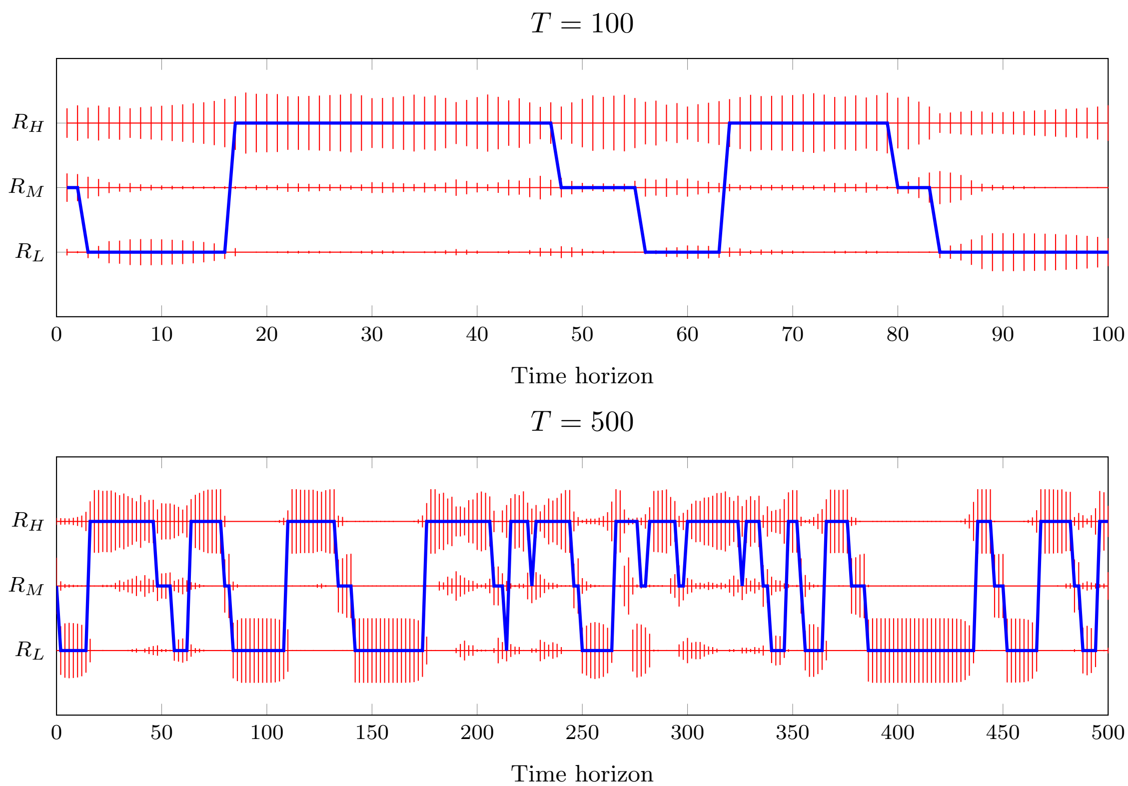

Figure 4.6: Estimation of unobservable stock levels for $T = 100$ and $T = 500$.

replenished up to level S, otherwise it remains unchanged. In order to make the scenario more realistic, we further assume that the inspection time is not fully deterministic, but rather subject to noise. Finally, on the demand side, we assume customers arrive according to a Poisson process and they purchase the product under consideration with probability θ.

We consider two sets of policy parameters: $(s, S) = (5, 5)$ and $(s, S) = (15, 30)$. In each case we fix the purchase probability at $\theta = 0.3$ and consider four different review rates $\{\widetilde{\mu}_1, \widetilde{\mu}_2, \widetilde{\mu}_3, \widetilde{\mu}_4\}$, where review rate is defined as the reciprocal of the review period. The values of the review rates are given in the following table:

(s, S)	$\widetilde{\mu}_1$	$\widetilde{\mu}_2$	$\widetilde{\mu}_3$	$\widetilde{\mu}_4$
(5,5)	0.04	0.06	0.08	0.10
(15,30)	0.0075	0.0085	0.0095	0.01

As stated above, we consider noisy realizations of the review periods, reflecting the fact that, in practice, inventory reviews might not always follow a precise schedule (with weekly reviews, for example, the time of day may vary). More specifically, the review period, i.e. the time between inventory reviews, will be given by $(1-K)\times 1/\widetilde{\mu}_k$, where K is a random variable uniformly distributed in $(0; 0.05)$, independently drawn for every inventory review. In words, we allow for a 5% error in the actual timing of the review.

The numerical results of this section are organized in two classes: we fist consider *in-sample* results, where we run our methodology with a generated dataset and analyze how well unobservable statistics of the data can be approximated by the algorithm. We then consider *out-of-sample* or *predictive* experimental results, where we use a dataset to fit (or train) our model and evaluate how well our methodology predicts inventory status as new data becomes available. The size of the generated dataset of transactions is given by $T = 3000$ in all cases.

In-sample results: In the in-sample results we simulate a $T = 3000$ transaction horizon for each of the instances described above. We then compute 95% confidence intervals for the purchase probability θ and two useful unobservable statistics: the average percentage of out of stock time periods $(T+1)^{-1}\sum_{t=0}^{T}\mathbb{I}_{\{h_t=0\}}$ and the average stock level, expressed as a fraction of the maximum stock S: $(T+1)^{-1}\sum_{t=0}^{T} h_t/S$. These quantities are of particular interest for retailers, as they can be used to estimate two of the main drivers of inventory costs: lost sales and holding costs.[5]

Figure 4.7 shows the results for the $(s, S) = (5, 5)$ case. The two pictures on the left display the estimation results for the purchase probability, the out-of-stock percentage time and the average (fractional) stock level, expressed as 95% confidence intervals, and their actual values. The pictures on the right display the stock level estimation for the first 500 periods and their true values. Overall, the results show that the estimations are remarkably accurate, especially considering the level of misspecification in the data generation. With very limited information the algorithm is able to closely approximate the purchase probability, which is the equivalent in our model to estimating demand, and give a very good picture of the unobservable statistics of the data.

[5] The average holding cost can be calculated by multiplying the unit holding cost of an item and the average stock level. The average lost sales cost can be approximated by multiplying the unit penalty cost of loosing a sale and the average lost sales in the time horizon: $\theta(T+1)^{-1}\sum_{t=0}^{T}\mathbb{I}_{\{h_t=0\}}$.

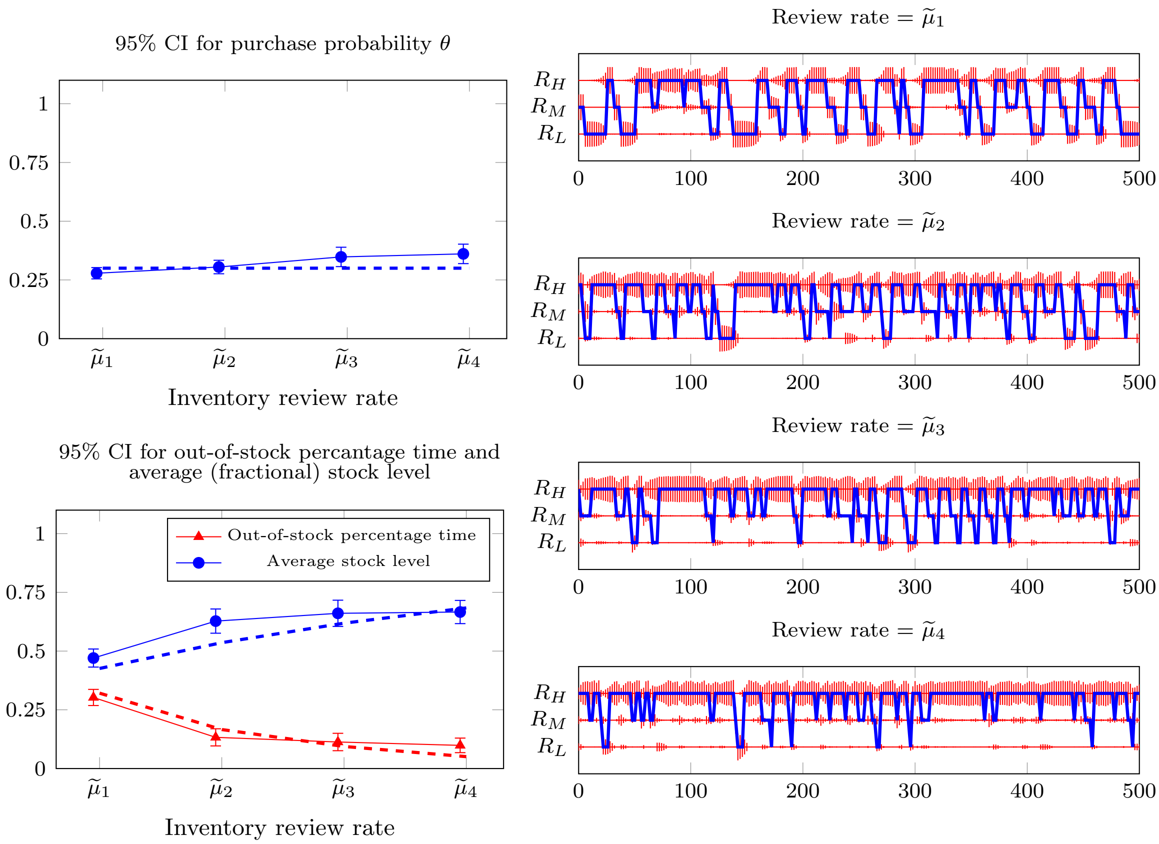

Figure 4.7: In-sample results for $(s, S) = (5, 5)$.

Our next set of results corresponds to the $(s, S) = (15, 30)$ case, presented in Figure 4.8. One can readily see from the upper left picture how the estimation of the purchase probability is substantially less accurate than in the $(s, S) = (5, 5)$ case. This is due to the fact that high stock levels make the estimation harder: with high S the system will spend most of its time in the intermediate states $[1, S-1]$, and the estimation requires frequent visits to 0 and S to produce more accurate results. To understand why, note that when the system is mostly in intermediate states, where both replenishes and purchases are possible, there exists a degree of freedom in the parameters that could explain the data; roughly speaking the likelihood depends more on the ratio θ/μ than θ and μ, and the estimation needs visits to 0 or S to resolve it. Despite these limitations, the model can still produce a very accurate estimation for the out-of-stock fraction of time and, though overestimated, a reasonable estimation of the average stock level. Moreover, as we will observe in the out-of-sample experiments, the algorithm will still be able to produce accurate predictions despite not having precise estimates of the full set of parameters.

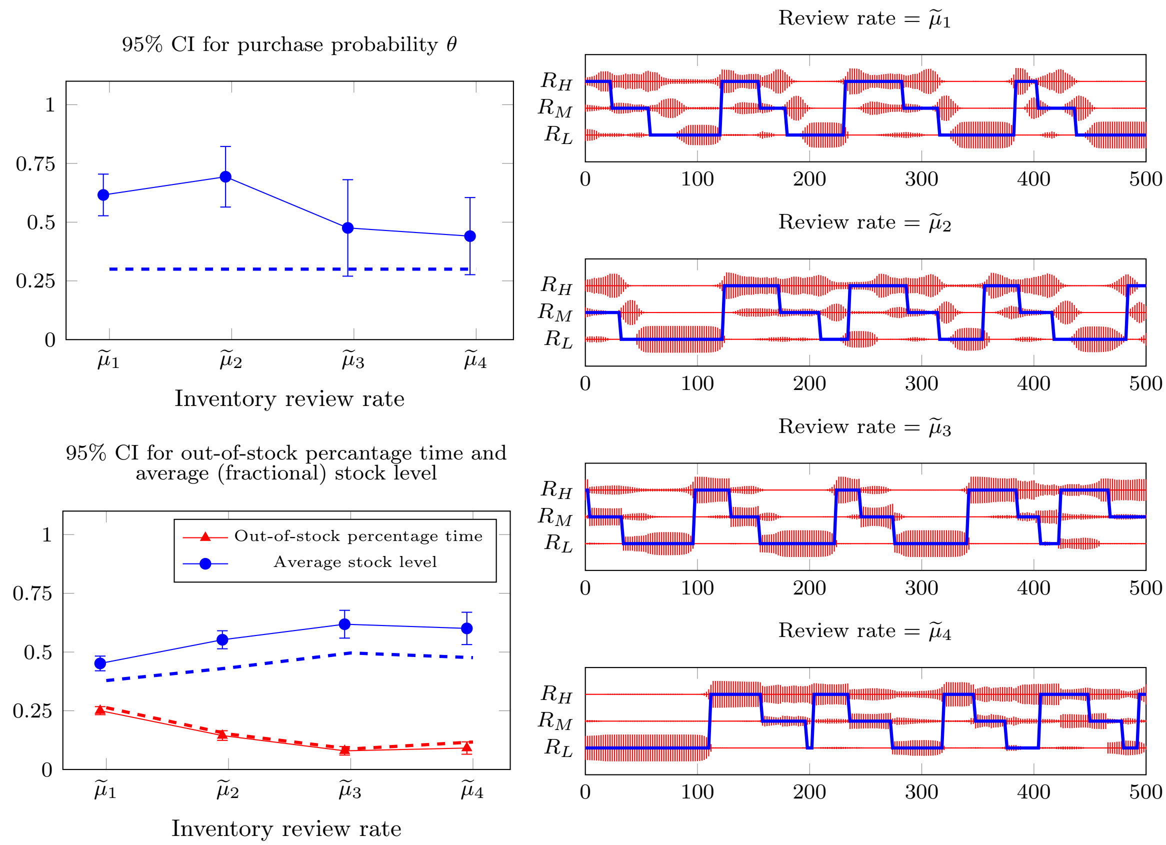

Figure 4.8: In-sample results for $(s, S) = (15, 30)$.

Out-of-sample results: We now turn our attention to out-of-sample experiments, where the output of the algorithm is tested on data that was not used to fit the model. We aim to evaluate the ability of the algorithm to predict the unobservable state of the system; in particular we focus on *out-of-stock* (OOS) predictions. The experiments work as follows: we generate a dataset of observations and we run the algorithm to compute estimates of the model parameters using the first 3000 periods. With the estimated parameters[6], we use the forward filtering phase of the FFBS algorithm (see the "Forward Pass" in Figure 4.4) to compute estimates of $p(h_t = 0|d_1^{t-1})$ for t greater than 3000. Note that the probabilities are

[6] To compute a concrete estimator of the parameters, we calculate the posterior sample mean using the output of the Gibbs Sampler.

calculated using only past data, emulating what would occur in a real world application. Following a typical classification approach, we then set thresholds for $p(h_t = 0|d_1^{t-1})$ to classify t as an in-stock or out-of-stock period and compare the result to the actual state. As a benchmark method, we consider a natural classification procedure that makes out-of-stock predictions based on the consecutive number of non-purchase periods. We call this method the *Naive* classification method. It works simply by setting a threshold $\gamma \in \mathbb{N}$ and classifying a period as out of stock if more than γ consecutive $\mathbf{0}$ were observed. Intuitively, the number of consecutive non-purchase periods represents the main signal of out-of-stock status, and hence it is natural to consider a classification scheme based on such signal. Using this simple approach, however, has the disadvantage that no past data is used to help recognize more complex patterns than a long sequence of consecutive non-purchases.

Based on the simulation model and instances described at the beginning of this section, we computed the output of our Gibbs Sampler and the Naive method and constructed Receiver Operating Characteristic (ROC) curves to test the effectiveness of the predictions. Each point in the curve is associated with a threshold value: the abscissa represents the fraction of in-stock periods that where (wrongly) classified as out-of-stock and the ordinate is the fraction of out-of-stock periods that where (correctly) classified as out-of-stock by the algorithm. ROC curves are commonly used to assess the trade-off between true positives and false negatives in a classification methodology. The area under the ROC curve (AUC) is also commonly computed as a measure of the overall quality of the method: a perfect classification would give an AUC equal to 1 and a method that randomly classifies the periods would give an AUC of 0.5 (in this case the ROC curve is given by a straight 45° line, usually included in ROC plots for reference).

Figures 4.9 and 4.10 show results for the $(s, S) = (5, 5)$ and $(s, S) = (15, 30)$ instances, respectively. One can readily observe from all the results that our proposed methodology outperforms the Naive method in every instance. The differences are small and moderate for the large stock examples, where both methods perform well. These instances seem to have more easily recognizable patterns of in-stock and out-of-stock periods: long periods of in-stock state are followed by periods of out-of-stock state (see for example the path plots on Figure 4.8). The $(s, S) = (5, 5)$ example, on the other hand, gives different results.

Figure 4.9 shows that the Naive method performs very poorly on these cases, and is widely outperformed by our Gibbs Sampler, which is able to attain comparable performance to that of the high stock examples. This implies that, even though the low stock instances represent harder examples, our methodology manages to recognize complex patterns in the data and accurately differentiate between in-stock and out-of-stock periods.

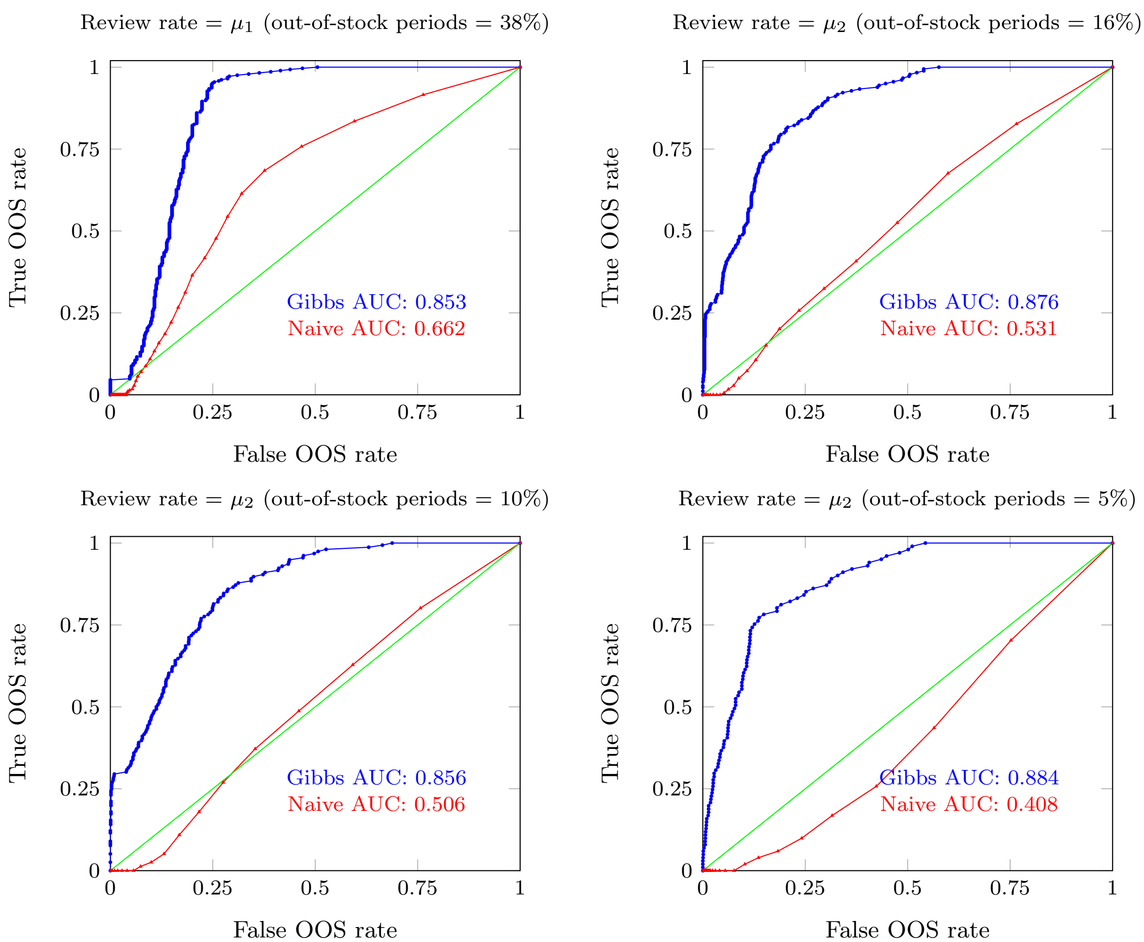

Figure 4.9: **Out-of-sample results for** $(\mathbf{s}, \mathbf{S}) = (\mathbf{5}, \mathbf{5})$: ROC curves with their corresponding Area Under the Curve (AUC) for the Gibbs Sampler and the Naive method.

Figures 4.9 and 4.10 provide results for only one simulated instance in each plot. In order to have a better picture of the predictive quality of the algorithms we ran 75 instances for each case, and computed the AUCs of the algorithms' output. The results are summarized in Tables 4.2 and 4.3. Both tables confirm our prior conclusions: the Gibbs Sampler

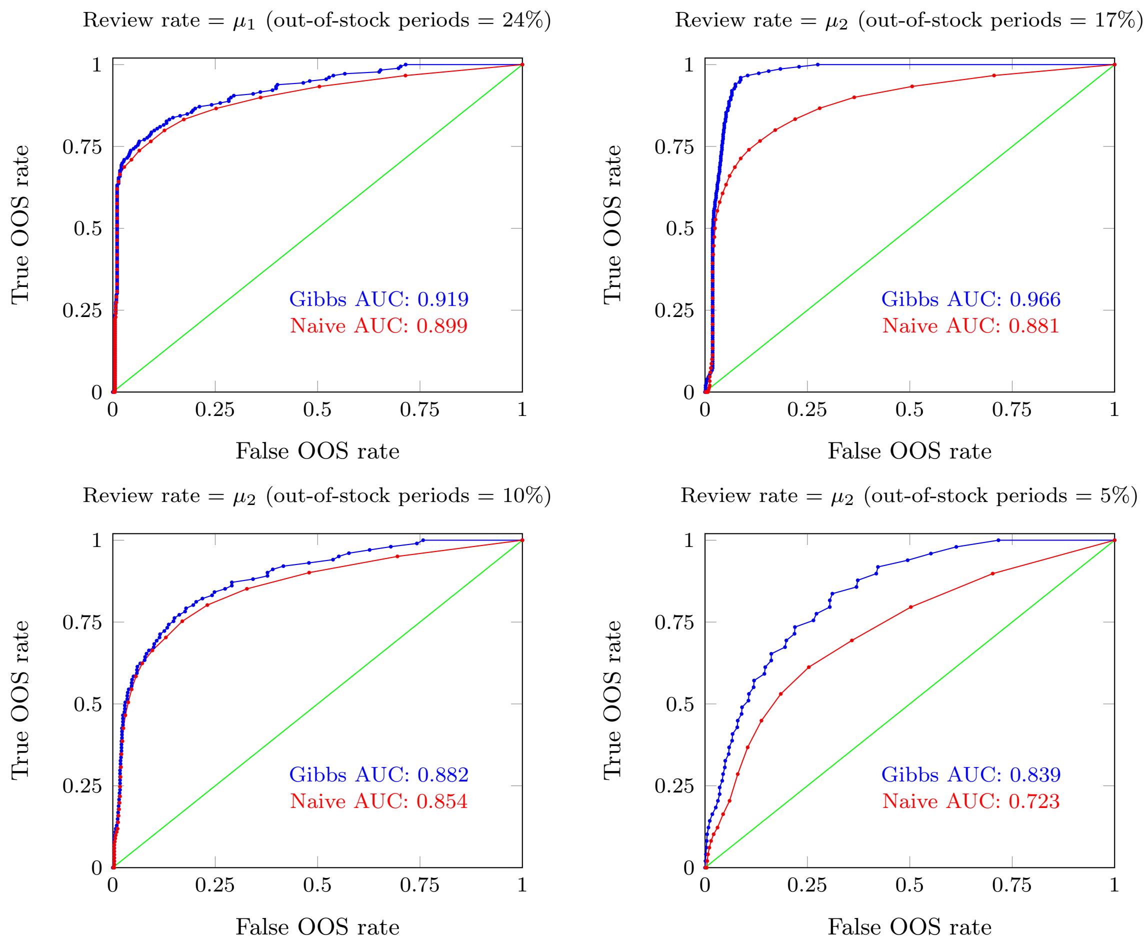

Figure 4.10: **Out-of-sample results for $(\mathbf{s}, \mathbf{S}) = (\mathbf{15}, \mathbf{30})$:** ROC curves with their corresponding Area Under the Curve (AUC) for the Gibbs Sampler and the Naive method.

outperforms the Naive method in all cases, with more substantial differences in the low stock examples. It is also interesting to note that the overall performance of our proposed methodology seems to be relatively uniform across all examples, providing very good results for different maximum stock levels and replenish frequencies, and strongly suggesting that it is indeed a very suitable option to be applied in practice.

4.3.2 Retailing data examples:

We conclude our experimental results with retailing data experiments. The data used for the experiments of this section was collected by Frogtek, a company devoted to pro-

μ_k	Mean Gibbs AUC	Mean Naive AUC	Mean rel. difference	Avg % OOS periods
μ_1	0.874 (±0.004)	0.655 (±0.006)	34 (±1)%	36%
μ_2	0.864 (±0.005)	0.567 (±0.005)	53 (±1)%	19%
μ_3	0.857 (±0.006)	0.486 (±0.006)	77 (±2)%	10%
μ_4	0.835 (±0.016)	0.448 (±0.010)	88 (±5)%	6%

Table 4.2: **Out-of-sample results for** $(\mathbf{s}, \mathbf{S}) = (\mathbf{5}, \mathbf{5})$: Estimated mean AUC (with a 95% confidence interval) of the Gibbs Sampler and Naive methods on 75 instances.

μ_k	Mean Gibbs AUC	Mean Naive AUC	Mean rel. difference	Avg % OOS periods
μ_1	0.908 (±0.006)	0.882 (±0.006)	2.9 (±0.4)%	23%
μ_2	0.904 (±0.010)	0.849 (±0.008)	6.6 (±1.3)%	14%
μ_3	0.858 (±0.018)	0.794 (±0.019)	8.6 (±1.9)%	9%
μ_4	0.862 (±0.017)	0.781 (±0.016)	10.8 (±2.2)%	6%

Table 4.3: **Out-of-sample results for** $(\mathbf{s}, \mathbf{S}) = (\mathbf{15}, \mathbf{30})$:Estimated mean AUC (with a 95% confidence interval) of the Gibbs Sampler and Naive methods on 75 instances.

viding information technology solutions for small retailers. Most of Frogtek's clients are micro-retailers who lack the infrastructure and resources to reliably track the stock of their products. However, using Frogtek's technology, it is possible to collect their POS data: for every transaction the system stores, among other information, the retailer or point of sale where the transaction took place, the products that where purchased and the transaction timestamp. Note that this data is sufficient to generate the input to our procedure: using the timestamps one can order the transaction events (recall from section 4.1.2 that **1** events are given by purchases of the product under consideration and **0** events are given by purchases of other products in the store) and obtain the input sequence d for the algorithm.

Since the data does not contain inventory information, it is not directly possible to evaluate the predictive power of our methodology in this setting. Therefore, with the help of Frogtek we selected a set of products and asked a retailer to *survey* the stock status of those products daily for a period of 44 days. Based on Frogtek's and the retailer's expertise, we focused on high demand products, more likely to be subject to stockouts. The retailer also reported an approximate maximum shelf space, which we used to set S while accounting for some uncertainty (by setting $\underline{\mu} < S$). Table 4.4 below contains a summary of the result of the inventory survey for the products under consideration.

	Description	Approximate maximum stock	Observed OOS periods	Total stock inspections
Product 1	Soda can	12	17	44
Product 2	Heavy cream	6	8	44
Product 3	Mineral water	12	25	44
Product 4	Cigarettes	5	16	44

Table 4.4: Summary of the inventory survey.

Using the output of the survey we were able to test the predictive power of our methodology, in a similar manner to the out-of-sample experiments of section 4.3.1.2: for each product we ran our methodology using data up to the start date of the survey and then computed predictions in an out-of-sample manner, using the forward filtering mechanism. The resulting ROC curves are depicted in Figure 4.11.

The results show that our methodology gives very good prediction results, outperforming the Naive method in all cases. It is important to note that many of the out-of-stock observations in these cases stem from unexpected supply chain events, such us problems on the supplier side or the retailer not having enough funds to pay for the replenishment, rather than a missmatch between supply and demand. This is reflected on the data as unusually frequent and long sequences of out-of-stock periods, which makes our model less able to fit the data and the naive method more effective. It is interesting to note then, that even under these adverse conditions, our methodology is able to produce useful predictive results.

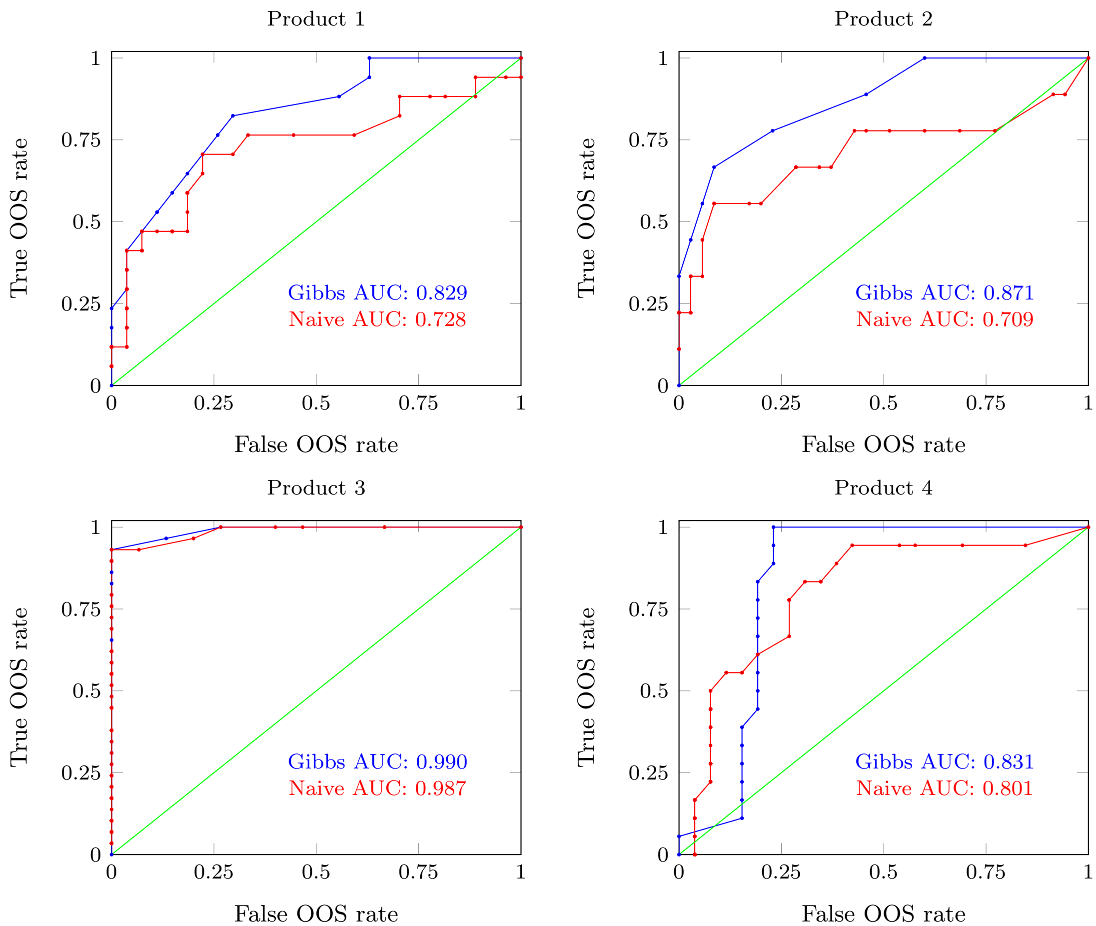

Figure 4.11: **Experimental results for the retailing data experiments.** Each plots depicts ROC curves for the Gibbs Sampler and Naive methods

4.4 Concluding Remarks

In the present chapter we study the problem of using transactional data to estimate stock levels when information about these is unavailable or unreliable. We developed a methodology where we model the stock evolution as an unobservable quantity, and the only observable data is given by purchase transactions. Because predicting stock levels is also predicting the drivers of stock evolution, we also regard our method as means to estimate demand with unreliable stock information. Through extensive experimental work that included simulated and actual retailing data we were able to verify the effectiveness of our method even in the presence of sever misspecifications.

Our model was focused on first order out-of-stock signals, without accounting for cross-product dependencies, which might provide useful information: a variation in the observed purchase frequency of a product might indicate the out-of-stock status of a related product. This is a worthwhile avenue of future research. Indeed, it could potentially not only improve the accuracy of the stock prediction, but also correspondingly estimate more complex demand models, with product demand inter-dependencies, in the presence of unreliable stock information.

Chapter 5

Conclusions

In this thesis we explore the effects of uncertainty and learning on well-known operational problems. We primarily focus on studying the challenges that arise when decisions have to be made in the presence of uncertainty, and in particular when the outcome of those decisions affects the way in which information is collected.

In chapters 2 and 3 we study two well known problems: the sequential newsvendor problem and sequential auctions with incomplete private information. We focus on understanding the impact of the coupling between decisions and information collection and in particular on quantifying the exploration-exploitation trade-off that arises from this coupling. Though different in nature, both problems have common properties that allow studying them in similar frameworks. We find that in both cases the value of exploration is limited, and a policy that ignores the information collection component of the problem can achieve near-optimal performance in wide ranges of settings.

In our study of the newsvendor problem we find that the exploration-exploitation trade-off is negligible in almost every case, with the exception of a very specific class of instances where demand is roughly deterministic but unknown. In such setting, which can be essentially regarded as degenerate, the value of exploring becomes non-trivial. Aside from those cases, the general conclusion is that, for all practical purposes, the exploration-exploitation trade-off can be disregarded.

Similarly, in the sequential auctions problem we find that the value of accounting for the exploration-exploitation trade-off is limited, though this conclusion does not hold as exhaustively as in the newsvendor case. Indeed, in this problem is not hard to identify examples in which being myopic with respect to information collection can be markedly suboptimal. However, we are able to characterize a broad class of problem settings where the myopic policy is nearly optimal, and we further show with examples that it includes many practical cases, commonly studied in the literature.

Finally, in chapter 4 we study the problem of reducing uncertainty in inventory, when access to stock information is limited or unreliable. We propose a model purely based on Point of Sale data, where inventory is regarded as a virtually unobservable quantity that can be inferred by observing sale patterns of the products. We find that our proposed methodology is able to accurately estimate and predict stock status, even under severe data misspecifiacations.

Bibliography

[Abramowitz and others, 1964] M Abramowitz et al. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Dover Publications, 1964.

[Agrawal and Smith, 1996] Narendra Agrawal and Stephen A Smith. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics (NRL)*, 43(6):839–861, 1996.

[Azoury, 1985] Katy S. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31(9), 1985.

[Babaioff *et al.*, 2009] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 79–88. ACM, 2009.

[Bensoussan *et al.*, 2007] Alain Bensoussan, Metin Çakanyildirim, and Suresh P Sethi. Partially observed inventory systems: The case of zero-balance walk. *SIAM Journal on Control and Optimization*, 46(1):176–209, 2007.

[Bensoussan *et al.*, 2009a] A. Bensoussan, M. Çakanyildirim, and Suresh P. Sethi. Censored newsvendor model revisited with unnormalized probabilities. *Journal Of Industrial And Management Optimization*, 5(2), 2009.

[Bensoussan *et al.*, 2009b] A. Bensoussan, A. Royal, and Suresh P. Sethi. Bayesian and adaptive controls for a newsvendor facing exponential demand. *Risk Decision Anal. 1*, 37(11), 2009.

[Bergemann and Valimaki, 2006] Dirk Bergemann and Juuso Valimaki. Bandit problems. 2006.

[Besbes and Muharremoglu, 2013] Omar Besbes and Alp Muharremoglu. On implications of demand censoring in the newsvendor problem. *Management Science*, 59(6):1407–1424, 2013.

[Bisi *et al.*, 2011] Arnab Bisi, Maqbool Dada, and Surya Tokdar. A censored-data multi-period inventory problem with newsvendor demand distributions. *Manufacturing Service Oper. Management*, 13(4), 2011.

[Braden and Freimer, 1991] David J. Braden and Marshall Freimer. Informational dynamics of censored observations. *Management Science*, 37(11), 1991.

[Bubeck and Cesa-Bianchi, 2012] Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122, 2012.

[Chen and Han, 2013] Fangruo Chen and Yunru Han. Rethinking the value of demand learning in a supply chain with censored demand. *Conference Presentation, M&SOM Conference 2013*, 2013.

[Chen and Mersereau, 2013] L. Chen and A.J. Mersereau. Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy. *working paper*, 2013.

[Chen, 2010] Li Chen. Bounds and heuristics for optimal bayesian inventory control with unobserved lost sales. *Operations research*, 58(2):396–413, 2010.

[Chen, 2013] Li Chen. Fixing phantom stockouts: Optimal data-driven shelf inspection policies. Technical report, Duke University. Working Paper, 2013.

[Conlon and Mortimer, 2012] Christopher T Conlon and Julie Holland Mortimer. Demand estimation under incomplete product availability. *American Economic Journal- Microeconomics. Forthcoming*, 2012.

[Dai and Jerath, 2013] Tinglong Dai and Kinshuk Jerath. Salesforce compensation with inventory considerations. *forthcoming in Management Science*, 2013.

[DeHoratius and Raman, 2008] Nicole DeHoratius and Ananth Raman. Inventory record inaccuracy: an empirical analysis. *Management Science*, 54(4):627–641, 2008.

[DeHoratius *et al.*, 2008] Nicole DeHoratius, Adam J Mersereau, and Linus Schrage. Retail inventory management when records are inaccurate. *Manufacturing & Service Operations Management*, 10(2):257–277, 2008.

[Devanur and Kakade, 2009] Nikhil R Devanur and Sham M Kakade. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 99–106. ACM, 2009.

[Ding *et al.*, 2002] Xiaomei Ding, Martin L. Puterman, and Arnab Bisi. The censored newsvendor and the optimal acquisition of information. *Operations Research*, 50(3), 2002.

[Edelman *et al.*, 2007] Benjamin Edelman, Michael Ostrovsky, and Michael SCHWARZ. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American economic review*, 97(1):242–259, 2007.

[Heese and Swaminathan, 2010] Sebastian Heese and Jayashankar M. Swaminathan. Inventory and sales effort management with unobservable lost sales. *European Journal of Operational Research*, 207(3):1263–1268, 2010.

[Huh and Rusmevichientong, 2009] W. T. Huh and P. Rusmevichientong. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research*, 34:103–123, 2009.

[Hummel and McAfee, 2014] Patrick Hummel and Preston McAfee. Machine learning in an auction environment. In *Proceedings of the 23rd international conference on World wide web*, pages 7–18. International World Wide Web Conferences Steering Committee, 2014.

[Iyer *et al.*, 2014] Krishnamurthy Iyer, Ramesh Johari, and Mukund Sundararajan. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, 2014.

[Jain *et al.*, 2014] Aditya Jain, Nils Rudi, and Tong Wang. Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, 2014.

[Kanoria and Nazerzadeh, 2014] Yash Kanoria and Hamid Nazerzadeh. Dynamic reserve prices for repeated auctions: learning from bids. Technical report, Working paper, 2014.

[Kök and Shang, 2007] A Gürhan Kök and Kevin H Shang. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing & Service Operations Management*, 9(2):185–205, 2007.

[Kunnumkal and Topaloglu, 2008] S. Kunnumkal and H. Topaloglu. Using stochastic approximation methods to compute optimal base-stock levels in inventory inventory control problems. *Operations Research*, 56:646–664, 2008.

[Lariviere and Porteus, 1999] Martin Lariviere and Evan L Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45(3):346–363, 1999.

[Li *et al.*, 2010] Sai-Ming Li, Mohammad Mahdian, and R Preston McAfee. Value of learning in sponsored search auctions. In *Internet and Network Economics*, pages 294–305. Springer, 2010.

[Lu *et al.*, 2005a] Xiangwen Lu, Jing-Sheng Song, and Kaijie Zhu. Inventory control with unobservable lost sales and bayesian updates. Technical report, Working paper, 2005.

[Lu *et al.*, 2005b] Xiangwen Lu, Jing-Sheng Song, and Kaijie Zhu. On "the censored newsvendor and the optimal acquisition of information". *Operations Research*, 53(6):1024–1026, 2005.

[Mersereau, 2014] A.J. Mersereau. Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management. Forthcoming*, 2014.

[Musalem *et al.*, 2010] Andrés Musalem, Marcelo Olivares, Eric T Bradlow, Christian Terwiesch, and Daniel Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

[Qi and Losonczi, 2010] Feng Qi and László Losonczi. Bounds for the ratio of two gamma functions. *Journal of Inequalities and Applications*, 2010:204, 2010.

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

[Scott, 2002] Steven L Scott. Bayesian methods for hidden markov models. *Journal of the American Statistical Association*, 97(457), 2002.

[Varian, 2007] Hal R Varian. Position auctions. *international Journal of industrial Organization*, 25(6):1163–1178, 2007.

[Vulcano *et al.*, 2012] Gustavo Vulcano, Garrett Van Ryzin, and Richard Ratliff. Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, 60(2):313–334, 2012.

[Wendel, 1948] JG Wendel. Note on the gamma function. *The American Mathematical Monthly*, 55(9):563–564, 1948.

Appendix

Appendix A

Appendix to Chapter 2

A1 Proofs

Preliminaries. For the proofs in this appendix, we will use the following notation conventions:

- $\mathbb{E}\big[\,\cdot\,\big]$ will denote expectation with respect to the pair (θ, D), where $\theta \sim \Gamma(a, S)$ is the unknown parameter of demand, and $D \sim F(\cdot|\theta)$. In some cases, in order to avoid ambiguities, we will specify the hyperparameters a and S in the expectation as $\mathbb{E}_{a,S}\big[\cdot\big]$.

- $C(a, S) := \min_{y\geq 0} \mathbb{E}\big[L(y, D)\big]$ denotes the single period optimal cost.

- $C^o_t(a, S)$ represents the future expected one period cost, $t + 1$ periods in the future, when demands are observable. That is, $C^o_t(a, S) := \mathbb{E}\left[C(a + t, S + D^\ell_1 + \ldots + D^\ell_t)\right]$.

- $C(\theta) := \min_{y\geq 0} \mathbb{E}\big[L(y, D)\big|\theta\big]$ denotes the single period optimal cost when θ is known. We also will denote $C^o_\infty(a, S) := \mathbb{E}\left[C(\theta)\right]$.

- The myopic order quantity $y^m(a, S)$ is defined as $y^m(a, S) := M^{-1}(r|a, S)$. The myopic order quantity when θ is known is denoted as $y^m(\theta) := F^{-1}(r|\theta)$.
  In the Weibull case these can be explicitly written as
  $y^m(a, S) := S^{1/\ell}\left((1 - r)^{-1/a} - 1\right)^{1/\ell}$ and $y^m(\theta) := \theta^{-1/\ell}(-\log(1 - r))^{1/\ell}$ respectively.

The proofs rely on a set of shorter technical lemmas presented in Appendix A2.

Proof of Lemma 2.1. We will establish the result of Lemma 2.1 for general newsvendor distributions. Define, for this proof only,

$$\begin{aligned} F(x|\theta) &:= 1 - e^{-\theta d(x)}, \\ \pi(\theta|a,S) &:= \frac{S^a \theta^{a-1} e^{-S\theta}}{\Gamma(a)}, \end{aligned}$$

where $d\colon [0,\infty) \to [0,\infty)$ is a differentiable, increasing and unbounded function.[1]

In the general newsvendor distribution case, one has

$$V_T^m(a,S) = C(a,S) + \mathbb{E}_{a,S}[V_{T-1}^m(a+1, S+d(D))] + \Gamma_{T-1}(a,S), \tag{A.1}$$

where

$$\Gamma_{T-1}(a,S) := (1-r)V_{T-1}^m(a, S+d(y^m)) - \mathbb{E}_{a,S}[V_{T-1}^m(a+1, S+d(D))\mathbb{I}_{\{D \geq y^m\}}],$$

and y^m represents the myopic order quantity. Here, with some slight abuse of notation, we keep the same notation as in the Weibull case in the main text. To lighten notation, we omit the dependency of y^m on a and S; in what follows, y^m always represents the myopic order quantity with respect to a and S, that is $y^m = d^{-1}(S[(1-r)^{-1/a} - 1])$.

First note that, by Lemma A2.3,

$$\mathbb{E}_{a,S}\left[V_T^m(a, S+d(D)) | D \geq y^m\right] = \mathbb{E}_{a,S+d(y^m)}\left[V_T^m\big(a, S+d(y^m)+d(D)\big)\right]. \tag{A.2}$$

Next, note that

$$\begin{aligned} &\Gamma_T(a,S) && \text{(A.3)} \\ &= (1-r)V_T^m(a, S+d(y^m)) - (1-r)\mathbb{E}_{a,S}[V_T^m(a+1, S+d(D))|D \geq y^m] \\ &= (1-r)\Big[C(a, S+d(y^m)) + \mathbb{E}_{a,S+d(y^m)}[V_{T-1}^m(a+1, S+d(y^m)+d(D))] + \\ &\qquad \Gamma_{T-1}(a, S+d(y^m)) - \mathbb{E}_{a,S}[V_T^m(a+1, S+d(D))|D \geq y^m]\Big] \\ &= (1-r)\Big[C(a, S+d(y^m)) + \mathbb{E}_{a,S+d(y^m)}[V_{T-1}^m(a+1, S+d(y^m)+d(D))] + \\ &\qquad \Gamma_{T-1}(a, S+d(y^m)) - \mathbb{E}_{a,S+d(y^m)}[V_T^m\big(a+1, S+d(y^m)+d(D)\big)]\Big], && \text{(A.4)} \end{aligned}$$

[1]Note that $d(\cdot)$ is defined to be *nondecreasing* in page 20 of the paper. To make the exposition more clear in this proof, and that of Lemma A2.3, we assume $d(\cdot)$ to be strictly increasing, though the proof can be extended to the more general case by defining, for example, $d^{-1}(z) := \inf\{y|d(z) = y\}$.

where the second equality follows from expanding $V_T^m(a, S+d(y^m))$ according to (A.1) and the third inequality follows from applying equation (A.2) to the last term in the second equation. Define

$$C_t^m(a,S) := V_{t+1}^m(a,S) - V_t^m(a,S), \quad \text{for} \;\; t = 0,\ldots,T.$$

$C_t^m(a,S)$ represents the future expected cost $t+1$ periods in the future, if the myopic policy is applied, and we start with hyperparamters a and S. One can then rewrite (A.4) as

$$\begin{aligned}
\Gamma_T(a,S) &= (1-r)\Big[C(a, S+d(y^m)) + \Gamma_{T-1}(a, S+d(y^m)) \\
&\qquad - \mathbb{E}_{a,S+d(y^m)}[C_{T-1}^m(a+1, S+d(y^m)+d(D))]\Big] \\
&= (1-r)\Big[C(a, S(1-r)^{-1/a}) + \Gamma_{T-1}(a, S(1-r)^{-1/a}) \\
&\qquad - \mathbb{E}_{a,S(1-r)^{-1/a}}[C_{T-1}^m(a+1, S(1-r)^{-1/a}+d(D))]\Big],
\end{aligned}$$

where the last equality comes from the fact that $S + d(y^m) = S(1-r)^{-1/a}$. By repeating the argument, one obtains

$$\begin{aligned}
\Gamma_T(a,S) &= \sum_{k=1}^{T}(1-r)^k\Big[C\left(a, S(1-r)^{-k/a}\right) - \\
&\qquad \mathbb{E}_{a,S(1-r)^{-k/a}}\left[C_{T-k}^m\left(a+1, S(1-r)^{-k/a}+d(D)\right)\right]\Big] \\
&\leq \sum_{k=1}^{T}(1-r)^k\Big[C\left(a, S(1-r)^{-k/a}\right) - \\
&\qquad \mathbb{E}_{a,S(1-r)^{-k/a}}\left[C_{T-k}^o\left(a+1, S(1-r)^{-k/a}+d(D)\right)\right]\Big] \\
&= \sum_{k=1}^{T}(1-r)^k\left[C\left(a, S(1-r)^{-k/a}\right) - C_{T-k+1}^o\left(a, S(1-r)^{-k/a}\right)\right],
\end{aligned}$$

The inequality is due to the fact that for any $t \geq 1$, $C_t^o(a,S) \leq C_t^m(a,S)$, a fact we formally prove in Lemma A2.4. This completes the proof.

□

■ □ ■

Proof of Theorem 2.1. The proof is based on first bounding the difference $V_T^m(a,S) - V_T^o(a,S)$ by the difference $V_{T-1}^o(a,S) - (T-1)C_\infty^o(a,S)$ properly scaled. We then derive a bound on the latter difference.

Specializing the bound on $\Gamma_T(a,S)$ given in Lemma 2.1 for the Weibull case, one obtains

$$\begin{aligned}\Gamma_T(a,S) &\le \sum_{k=1}^{T}(1-r)^k\left[C(a,S(1-r)^{-k/a}) - C^o_{T-k+1}(a,S(1-r)^{-k/a})\right]\\ &\overset{(a)}{=} S^{1/\ell}\sum_{k=1}^{T}(1-r)^{k\frac{a\ell-1}{a\ell}}\left[C(a,1)-C^o_{T-k+1}(a,1)\right]\\ &\overset{(b)}{\le} \left[\sum_{k=1}^{T}(1-r)^{k\frac{a\ell-1}{a\ell}}\right]\left[C(a,S)-C^o_\infty(a,S)\right], \end{aligned} \tag{A.5}$$

where (a) follows from the scalability property and (b) follows from the fact that $C^o_t(a,S) \ge C^o_\infty(a,S)$ for any $t = 0,\ldots,T$.[2]

By equation (A.1), page 96, specialized to the Weibull case, we have

$$V^m_T(a,S) - V^o_T(a,S) \;=\; \mathbb{E}_{a,S}\left[V^m_{T-1}(a+1,S+D^\ell) - V^o_{T-1}(a+1,S+D^\ell)\right] + \Gamma_{T-1}(a,S).$$

If we denote $\widehat{D}^\ell_t := D^\ell_1 + \ldots + D^\ell_t$ and proceed recursively, we obtain

$$\begin{aligned} & V^m_T(a,S) - V^o_T(a,S) \qquad\qquad (A.6)\\ &= \sum_{t=0}^{T-2}\mathbb{E}_{a,S}\left[\Gamma_{T-1-t}\left(a+t,S+\widehat{D}^\ell_t\right)\right]\\ &\le \left[\sum_{k=1}^{T}(1-r)^{k\frac{a\ell-1}{a\ell}}\right]\sum_{t=0}^{T-2}\mathbb{E}_{a,S}\left[C(a+t,S+\widehat{D}^\ell_t) - C^o_\infty(a+t,S+\widehat{D}^\ell_t)\right]\\ &= \left[\sum_{k=1}^{T}(1-r)^{k\frac{a\ell-1}{a\ell}}\right]\sum_{t=0}^{T-2}\left(C^o_t(a,S)-C^o_\infty(a,S)\right)\\ &= \left[\sum_{k=1}^{T}(1-r)^{k\frac{a\ell-1}{a\ell}}\right]\left(V^o_{T-1}(a,S)-(T-1)C^o_\infty(a,S)\right)\\ &= \frac{\lambda-\lambda^{T+1}}{1-\lambda}\left(V^o_{T-1}(a,S)-(T-1)C^o_\infty(a,S)\right), \qquad\qquad (A.7)\end{aligned}$$

where the inequality follows by (A.5).

We now bound the difference $V^o_{T-1}(a,S)-(T-1)C^o_\infty(a,S)$. For that we rely on lemma

[2]It is not hard to derive this fact from the definition of $C^o_t(a,S)$ and $C^o_\infty(a,S)$. Intuitively, if θ is known there is always less demand uncertainty than if θ is unknown (even after any number of demand observations) and hence $C^o_\infty(a,S)$ is smaller than $C^o_t(a,S)$.

A2.5 that bounds the difference $C^o_t(a,S) - C^o_\infty(a,S)$ of costs in the observable case.

$$
\begin{aligned}
V^o_{T-1}(a,S) - (T-1)C^o_\infty(a,S) &= \sum_{t=0}^{T-2}\left[C^o_t(a,S) - C^o_\infty(a,S)\right] \\
\leq \quad S^{1/\ell}K(r,\ell)\left[\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{\frac{1}{\ell}} \sum_{t=0}^{T-2}\frac{1}{a+t-1/\ell} \\
\leq \quad S^{1/\ell}K(r,\ell)\left[\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{\frac{1}{\ell}} \left[\log(a-1/\ell+T-2) - \log(a-1/\ell) + (a-1/l)^{-1}\right] \\
\leq \quad S^{1/\ell}K(r,\ell)\left[\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{\frac{1}{\ell}} \left[\log(1+\frac{T}{a-1/\ell}) + (a-1/l)^{-1}\right]. \qquad \text{(A.8)}
\end{aligned}
$$

Let us now define

$$
Q(a,r,\ell) := [K(r,\ell)]^\ell\,\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}.
$$

Note that, because $\exp\{x\} \leq 1+(e-1)x$ for any $x \in (0,1)$, we have that

$$
\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell} = O\left(\frac{1}{a-1/\ell} + \frac{e-1}{(a-1/\ell)^2}\right) = O\left(\frac{1}{a}\right) \qquad \text{as } a \to \infty.
$$

Therefore, $Q(a,r,\ell) = O(1/a)$ when r and ℓ are fixed.

By substituting $Q(a,r,\ell)$ in (A.8) and combining the inequality with (A.7), we obtain the result of the theorem and the proof is complete.

□

■ □ ■

Proof of Proposition 2.1. As stated in the proof of Theorem 2.1, it is not hard to show that $C^o_t(a,S) \geq C^o_\infty(a,S)$ for all t. Therefore,

$$
V^o_T(a,S) := \sum_{t=0}^{T-1} C^o_t(a,S) \geq TC^o_\infty(a,S) := T\mathbb{E}_{a,S}\left[C(\theta)\right].
$$

Combining the latter with the result of Theorem 2.1, one obtains

$$
\frac{V^m_T(a,S) - V^o_T(a,S)}{V^o_T(a,S)} \leq S^{1/\ell}\left[Q(a,r,\ell)\right]^{1/\ell}\frac{\lambda-\lambda^{T+1}}{1-\lambda}\,\frac{\log\left(1+T/(a-1/\ell)\right) + (a-1/l)^{-1}}{T\mathbb{E}_{a,S}\left[C(\theta)\right]}.
$$

Since the right side is $O(T^{-1}\log(T))$ as $T \uparrow \infty$, the proof is complete. □

■ □ ■

Proof of Proposition 2.2. Note that, since the myopic value of information is independent of S (cf. the scalability property), the statement is equivalent to

$$\frac{V_T^m(a,a) - V_T^o(a,a)}{V_T^o(a,a)} = O\left(\frac{1}{a}\right) \text{ as } a \to \infty.$$

We start by showing that $V_T^o(a,a)$ is lower bounded by a positive constant for any a. Since $V_T^o(a,S) := \sum_{t=0}^{T-1} C_t^o(a,S)$, it suffices to show that each term $C_t^o(a,a)$ is lower bounded by a positive constant itself,

$$\begin{aligned} C_t^o(a,a) &\geq C_\infty^o(a,a) \\ &= \mathbb{E}_{a,a}\left[\frac{1}{1-r}\int_{y^m(\theta)}^\infty xf(x|\theta)dx - \int_0^\infty \overline{F}(x|\theta)dx\right] \\ &= \mathbb{E}_{a,a}\left[\theta^{-1/\ell}\right]\left[\frac{1}{1-r}\int_{y^m(1)}^\infty xf(x|1)dx - \int_0^\infty \overline{F}(x|1)dx\right] \\ &\geq \mathbb{E}_{a,a}\left[\theta\right]^{-1/\ell}\left[\frac{1}{1-r}\int_{y^m(1)}^\infty xf(x|1)dx - \int_0^\infty \overline{F}(x|1)dx\right] \\ &= \frac{1}{1-r}\int_{y^m(1)}^\infty xf(x|1)dx - \int_0^\infty \overline{F}(x|1)dx > 0, \end{aligned}$$

where the first equality follows from Lemma A2.1 and the second inequality is a result of Jensen's inequality applied to the expectation term. The last equality follows from the fact that $\mathbb{E}_{a,a}\left[\theta\right] = 1$. We then have shown that

$$V_T^o(a,a) \geq TC_\infty^o(a,a) \geq T\left[\frac{1}{1-r}\int_{y^m(1)}^\infty xf(x|1)dx - \int_0^\infty \overline{F}(x|1)dx\right] =: \underline{m} > 0.$$

Using Theorem 2.1 one obtains

$$\frac{V_T^m(a,a) - V_T^o(a,a)}{V_T^o(a,a)} \leq \frac{\lambda - \lambda^{T+1}}{\underline{m}(1-\lambda)}\left[aQ(a,r,\ell)\right]^{1/\ell}\left[\log(1 + \frac{T}{a-1/\ell}) + \frac{1}{a-1/l}\right],$$

and since $aQ(a,r,\ell) = O(1)$, the right side is $O(a^{-1})$ as $a \uparrow \infty$ and the result is established.

□

■ □ ■

Proof of Proposition 2.3. Following the inequality given in (A.7), in the proof of Theorem 2.1 (page 98), it suffices to show that

$$\frac{\lambda-\lambda^{T+1}}{1-\lambda}\frac{V^o_{T-1}(a,S)-(T-1)\mathbb{E}\big[C(\theta)\big]}{V^o_{T-1}(a,S)} = \begin{cases} O\left((1-r)^{1-1/a\ell}\right) & \text{as} \quad r\to 1^-, \\ O\left(r^{1/\ell}\right) & \text{as} \quad r\to 0^+. \end{cases} \tag{A.9}$$

where $\lambda := (1-r)^{1-\frac{1}{a\ell}}$.

Note: for this proof, all expectations are taken with respect to hyperparameters a and S, that is, $\mathbb{E}[\cdot]\equiv\mathbb{E}_{a,S}[\cdot]$.

i.) We start with the case $r\to 1^-$. Let us start by noting that $(\lambda-\lambda^{T+1})/(1-\lambda) = O((1-r)^{1-1/a\ell})$ as $r\to 1^-$ and hence, it suffices, for example, to show that the ratio involving the value functions in (A.9) converges to a constant. In particular we will show that

$$\frac{(T-1)\mathbb{E}\big[C(\theta)\big]}{V^o_{T-1}(a,S)}\to 0, \quad \text{as} \quad r\to 1^-.$$

It is not hard to see that both numerator and denominator converge to infinity as $r\to 1^-$. We can therefore apply L'Hôpital's rule and differentiate both terms with respect to r. Using the expression for $C(\theta)$ given in Lemma A2.1 one has

$$\begin{aligned}
&\mathbb{E}\big[C(\theta)\big]\\
&= \mathbb{E}\left[\frac{1}{1-r}\int_{y^m(\theta)}^{\infty} xf(x|\theta)dx-\int_0^{\infty}\overline{F}(x|\theta)dx\right]\\
&= \mathbb{E}\left[y^m(\theta)+\frac{1}{1-r}\int_{y^m(\theta)}^{\infty}\overline{F}(z|\theta)dz-\int_0^{\infty}\overline{F}(z|\theta)dz\right]\\
&= \mathbb{E}\left[\theta^{-1/\ell}\right](-\log(1-r))^{1/\ell}+\frac{\mathbb{E}\left[\theta^{-1/\ell}\right]}{1-r}\int_{(-\log(1-r))^{1/\ell}}^{\infty} e^{-z^\ell}dz-\mathbb{E}\left[\int_0^{\infty}e^{-\theta z^\ell}dz\right].
\end{aligned}$$

By differentiating with respect to r one obtains

$$\begin{aligned}
\frac{\partial}{\partial r}\mathbb{E}\big[C(\theta)\big] &= \mathbb{E}\left[\theta^{-1/\ell}\right]\frac{\partial}{\partial r}\left[(-\log(1-r))^{1/\ell}+\frac{1}{1-r}\int_{(-\log(1-r))^{1/\ell}}^{\infty}e^{-z^\ell}dz\right]\\
&= \mathbb{E}\left[\theta^{-1/\ell}\right]\frac{1}{(1-r)^2}\int_{(-\log(1-r))^{1/\ell}}^{\infty}e^{-z^\ell}dz\\
&= \mathbb{E}\left[\frac{1}{(1-r)^2}\int_{y^m(\theta)}^{\infty}e^{-\theta z^\ell}dz\right],
\end{aligned}$$

where the second equality is a result of the chain rule and the fundamental theorem of calculus applied to the integral term.

A similar argument applied to $C_t^o(a,S)$ yields that

$$\frac{\partial}{\partial r}\mathbb{E}\big[C_t^o(a,S)\big] \;=\; \mathbb{E}\left[\frac{1}{(1-r)^2}\int_{y^m(a_t,S_t)}^{\infty}\overline{M}(z|a_t,S_t)dz\right],$$

where $S_t := S + D_1^\ell + \ldots + D_t^\ell$ and $a_t := a + t$. Then one has that

$$\begin{aligned}\lim_{r\to 1^-}\frac{(T-1)\mathbb{E}\big[C(\theta)\big]}{V_{T-1}^o(a,S)} &\overset{(a)}{=} \lim_{r\to 1^-}\frac{(T-1)\mathbb{E}\left[\int_{y^m(\theta)}^{\infty}e^{-\theta z^\ell}dz\right]}{\sum_{t=0}^{T-2}\mathbb{E}\left[\int_{y^m(a_t,S_t)}^{\infty}\overline{M}(z|a_t,S_t)dz\right]}\\ &\overset{(b)}{=} \lim_{r\to 1^-}\frac{(T-1)(1-r)\frac{\partial}{\partial r}\mathbb{E}\left[y^m(\theta)\right]}{\sum_{t=0}^{T-2}(1-r)\frac{\partial}{\partial r}\mathbb{E}\left[y^m(a_t,S_t)\right]}\\ &\overset{(c)}{=} \lim_{r\to 1^-}\frac{(T-1)\mathbb{E}\left[\theta^{-1/\ell}\right](-\log(1-r))^{1/\ell-1}(1-r)^{-1}}{\sum_{t=0}^{T-2}\mathbb{E}\left[S_t^{1/\ell}\right]\left((1-r)^{-1/a_t}-1\right)^{1/\ell-1}a_t^{-1}(1-r)^{-1/a_t-1}}\end{aligned}$$

where (a) follows from applying L'Hôpital's rule, (b) follows from applying L'Hôpital's rule and interchanging differentiation and expectation[3] and (c) is a result of applying L'Hôpital's to the right side of (b) . Elementary calculus (in particular, repetitively applying L'Hôpital's rule to the ratio) yields that the last limit is equal to 0, and hence the $r \to 1^-$ case is complete.

ii.) We now analyze the case $r \to 0^+$. Let us start by noting that, following (A.9) and noting that $(\lambda - \lambda^{T+1})/(1-\lambda) \to T$ as $r \to 0^+$, it suffices to show that

$$\lim_{r\to 0^+}\frac{V_{T-1}^o(a,S) - (T-1)\mathbb{E}\big[C(\theta)\big]}{r^{1/\ell}V_{T-1}^o(a,S)}$$

exists and is finite.

Defining the following notation

$$f(r) \;:=\; V_{T-1}^o(a,S), \qquad g(r) \;:=\; (T-1)\mathbb{E}\big[C(\theta)\big],$$

we aim to establish

$$\lim_{r\to 0^+}\frac{f(r)-g(r)}{r^{1/\ell}f(r)} < \infty. \tag{A.10}$$

By following similar arguments as the ones used in *i.*) one can show that

[3]The interchange in the numerator is justified by the Dominated Convergence Theorem: $\int_{y^m(\theta)}^{\infty}e^{-\theta z^\ell}dz \leq \int_0^{\infty}e^{-\theta z^\ell}dz$ for any $r \in (0,1)$ and $\theta > 0$, and $\mathbb{E}\left[\int_0^{\infty}e^{-\theta z^\ell}dz\right] = \mathbb{E}\left[\mathbb{E}[D|\theta]\right] = \mathbb{E}\left[D\right] < \infty$. A similar argument justifies the interchange in the denominator.

a) Both numerator and denominator in (A.10) converge to 0 as $r \to 0^+$.

b) $\lim_{r\to 0^+} f'(r) = \lim_{r\to 0^+} g'(r) = (T-1)\mathbb{E}[D]$.

c) The second derivatives of $f(\cdot)$ and $g(\cdot)$ are given by

$$\begin{aligned} f''(r) &= \frac{2}{(1-r)^3}\widetilde{f}(r) + \frac{1}{(1-r)^2}\widetilde{f}'(r), \\ g''(r) &= \frac{2}{(1-r)^3}\widetilde{g}(r) + \frac{1}{(1-r)^2}\widetilde{g}'(r), \end{aligned}$$

where

$$\begin{aligned} \widetilde{f}(r) &:= \sum_{t=0}^{T-2} \mathbb{E}\left[\int_{y^m(a_t,S_t)}^{\infty} \overline{M}(z|a_t,S_t)dz\right], \\ \widetilde{g}(r) &:= (T-1)\mathbb{E}\left[\int_{y^m(\theta)}^{\infty} e^{-\theta z^\ell} dz\right], \end{aligned}$$

and

$$\begin{aligned} \widetilde{f}'(r) &:= -\frac{1}{\ell}\sum_{t=0}^{T-2} \frac{\mathbb{E}\left[S_t^{1/\ell}\right]}{a_t}\left((1-r)^{-1/a_t}-1\right)^{1/\ell-1}(1-r)^{-1/a_t}, \\ \widetilde{g}'(r) &:= -\frac{(T-1)}{\ell}\mathbb{E}\left[\theta^{-1/\ell}\right](-\log(1-r))^{1/\ell-1}. \end{aligned}$$

By item *b*) above and L'Hôpital's rule, one has that

$$\lim_{r\to 0^+} r^{-1} f(r) = \lim_{r\to 0^+} r^{-1} g(r) = (T-1)\mathbb{E}[D]. \tag{A.11}$$

This implies that, if we take derivatives in (A.10) both the numerator and denominator converge to 0. By taking second derivatives one obtains

$$\lim_{r\to 0^+} \frac{f''(r)-g''(r)}{r^{1/\ell-1}\left[(\frac{1}{\ell}-1)\frac{1}{\ell}r^{-1}f(r) + \frac{2}{l}f'(r) + rf''(r)\right]} \tag{A.12}$$

Next, we establish that

$$\lim_{r\to 0^+} \frac{f''(r)-g''(r)}{r^{1/\ell-1}} \in \mathbb{R}. \tag{A.13}$$

Let us start by noting that, by elementary calculus,

$$\begin{aligned} \lim_{r\to 0^+} \frac{(-\log(1-r))^{1/\ell-1}}{r^{1/\ell-1}} &= 1, \\ \lim_{r\to 0^+} \frac{\left((1-r)^{-1/a_t}-1\right)^{1/\ell-1}(1-r)^{-1/a_t}}{r^{1/\ell-1}} &= 1/a_t^{1/\ell-1}, \end{aligned} \tag{A.14}$$

and therefore

$$\lim_{r\to 0^+} \frac{\widetilde{f}'(r)-\widetilde{g}'(r)}{r^{1/\ell-1}} = \frac{T-1}{\ell}\mathbb{E}\left[\theta^{-1/\ell}\right] - \sum_{t=0}^{T-2}\mathbb{E}\left[\frac{S_t^{1/\ell}}{a_t^{1/\ell}}\right] =: K \in \mathbb{R}_{++},$$

where $K>0$ follows from Jensen's inequality and the law of iterated expectations (see page 113 for a detailed derivation of this fact). This implies that

$$\lim_{r\to 0^+} \frac{f''(r)-g''(r)}{r^{1/\ell-1}} \quad = \quad \lim_{r\to 0^+} 2\frac{\widetilde{f}(r)-\widetilde{g}(r)}{r^{1/\ell-1}} + K.$$

Note that if $\ell \geq 1$, and because $\left[\widetilde{f}(r)-\widetilde{g}(r)\right] \to 0$, the proof is complete. If $\ell < 1$, we can apply L'Hôpital's rule to obtain

$$\lim_{r\to 0^+} \frac{f''(r)-g''(r)}{r^{1/\ell-1}} \quad = \quad \lim_{r\to 0^+} \frac{2}{\frac{1}{\ell}-1}\frac{\widetilde{f}'(r)-\widetilde{g}'(r)}{r^{1/\ell-2}} + K = K,$$

where the second equality follows from applying (A.14) and the expressions for $\widetilde{f}'(r)$ and $\widetilde{g}'(r)$. Hence (A.13) is established.

Now, by combining (A.13) and (A.11) above, one obtains that the limit in (A.12) is finite. This completes the proof. □

■ □ ■

Proof of Theorem 2.2. Suppose demands are exponential and $a>1$. The proof relies on an alternative informational system. In the *random rejection* system, at each step the decision maker obtains either a full demand observation, independently of the order quantity y and with probability r, or no observation at all, with probability $1-r$. Because in this case the information collection is independent of the decision process, a myopic policy is optimal, and the optimal cost is given by the solution to the Bellman equation

$$V_T^r(a,S) := C(a,S) + r\mathbb{E}\left[V_{T-1}^r(a+1,S+D)\right] + (1-r)V_{T-1}^r(a,S).$$

It is not hard to show (by, for example, following similar steps to those in the proof of Proposition 2.4 for $\ell=1$) that the cost function for the random rejection system also satisfies the scalability property, that is,

$$V_T^r(a,S) = SV_T^r(a,1), \quad \text{for any} \ \ T\geq 1, S>0, a>1.$$

By using this fact and the fact that, for the exponential case, $\mathbb{E}_{a,S}[D] = S/(a-1)$ one can rewrite the recursive formula for $V_T^r(a,S)$ as

$$V_T^r(a,S) = C(a,S) + r\frac{a}{a-1}V_{T-1}^r(a+1,S) + (1-r)V_{T-1}^r(a,S). \tag{A.15}$$

To establish Theorem 2.2, we will establish the two following inequalities.

a) $V_T^m(a,S) \geq V_T^r(a,S) \geq V_T^o(a,S)$.

b) $V_T^r(a,S) - V_T^o(a,S) \geq (1-r)(-\log(1-r))^2\frac{S}{a-1}\left[\log\left(1+\frac{T-2}{a}\right) - \frac{T-2}{a+T-2}\right]$.

a) The second inequality follows directly from the fact that full observations is a more informative system than random rejections. We prove the first inequality, $V_T^m(a,S) \geq V_T^r(a,S)$, in two steps.

Step 1. We first establish the following cost relationship in the rejection system.

$$V_T^r(a,S) \geq \frac{a}{a-1}V_T^r(a+1,S).$$

We proceed by induction in T. The case $T=1$ follows from the fact that $V_1^r(a,S) = C(a,S)$ and [Bisi *et al.*, 2011, Lemma 1]. Suppose $V_{T-1}^r(a,S) \geq \frac{a}{a-1}V_{T-1}^r(a+1,S)$ for all $a > 1$. Using the recursive equation for $V_T^r(a,S)$ one can write

$$\begin{aligned} V_T^r(a,S) &= C(a,S) + r\frac{a}{a-1}V_{T-1}^r(a+1,S) + (1-r)V_{T-1}^r(a,S) \\ \frac{a}{a-1}V_T^r(a+1,S) &= \frac{a}{a-1}C(a+1,S) + r\frac{a}{a-1}\frac{(a+1)}{(a+1)-1}V_{T-1}^r(a+2,S) \\ &\quad + \frac{a}{a-1}(1-r)V_{T-1}^r(a+1,S) \end{aligned}$$

The base case implies that the first term on the right side of the first equation dominates the first term in the right side of the second equation. Similarly, by the induction hypothesis, the two last terms in the right side of the first equation dominate the corresponding terms in the right side of the second equation and hence the inequality is established.

Step 2. We now prove the inequality we are after, $V_T^m(a,S) \geq V_T^r(a,S)$, by induction on T. If $T=1$ the result follows from the fact that $V_1^m(a,S) = V_1^r(a,S) = C(a,S)$. Suppose that $V_{T-1}^m(a,S) \geq V_{T-1}^r(a,S)$. Then

$$\begin{aligned} V_T^m(a,S) &= C(a,S) + \frac{a}{a-1}\left(1-(1-r)^{1-\frac{1}{a}}\right)V_{T-1}^m(a+1,S) + (1-r)^{1-\frac{1}{a}}V_{T-1}^m(a,S) \\ &\geq C(a,S) + \frac{a}{a-1}\left(1-(1-r)^{1-\frac{1}{a}}\right)V_{T-1}^r(a+1,S) + (1-r)^{1-\frac{1}{a}}V_{T-1}^r(a,S), \end{aligned}$$

where the first equality follows from Proposition 2.4, page 29, applied to $\ell = 1$.
Using the inequality above and Bellman's recursion for $V_T^r(a,S)$ in (A.15)

$$\begin{aligned} V_T^m(a,S) - V_T^r(a,S) &\geq (1-r)\frac{a}{a-1}\left(1-(1-r)^{-\frac{1}{a}}\right)V_{T-1}^r(a+1,S) + \\ &\quad (1-r)\left((1-r)^{-\frac{1}{a}}-1\right)V_{T-1}^r(a,S) \\ &\geq (1-r)\frac{a}{a-1}\left(1-(1-r)^{-\frac{1}{a}}\right)V_{T-1}^r(a+1,S) + \\ &\quad (1-r)\frac{a}{a-1}\left((1-r)^{-\frac{1}{a}}-1\right)V_{T-1}^r(a+1,S) \\ &= 0 \end{aligned}$$

where the second inequality follows from step 1 above. This completes the induction step and a) is established.

b) We first rewrite the full observation and random rejection value functions for the exponential case:

$$\begin{aligned} V_T^o(a,S) &= C(a,S) + \frac{a}{a-1}V_{T-1}^o(a+1,S) \\ V_T^r(a,S) &= C(a,S) + r\frac{a}{a-1}V_{T-1}^r(a+1,S) + (1-r)V_{T-1}^r(a,S) \\ &= C(a,S) + \Gamma_{T-1}^r(a,S) + \frac{a}{a-1}V_{T-1}^r(a+1,S), \end{aligned}$$

where $\Gamma_{T-1}^r(a,S) := (1-r)\left[V_{T-1}^r(a,S) - \frac{a}{a-1}V_{T-1}^r(a+1,S)\right]$.
Therefore,

$$V_T^r(a,S) - V_T^o(a,S) = \Gamma_{T-1}^r(a,S) + \frac{a}{a-1}\left[V_{T-1}^r(a+1,S) - V_{T-1}^o(a+1,S)\right].$$

By proceeding recursively one obtains

$$V_T^r(a,S) - V_T^o(a,S) = \sum_{k=0}^{T-2}\frac{a+k-1}{a-1}\Gamma_{T-k-1}^r(a+k,S). \tag{A.16}$$

Lemma A2.8, presented in Appendix A2, states that

$$\Gamma_T^r(a,S) \geq \frac{1-r}{a-1}\left[(a-1)C(a,S) - (a+T-1)C(a+T,S)\right]. \tag{A.17}$$

By combining (A.16) and (A.17) one obtains

$$\begin{aligned}
&V_T^r(a,S) - V_T^o(a,S) \\
&\geq \frac{1-r}{a-1}\sum_{k=0}^{T-2}\left[(a+k-1)C(a+k,S) - (a+T-2)C(a+T-1,S)\right] \\
&\overset{(a)}{\geq} (1-r)\frac{\log^2(1-r)S}{2(a-1)}\sum_{k=0}^{T-2}\left[\frac{1}{a+k} - \frac{1}{a+T-1}\right] \\
&\geq (1-r)\frac{\log^2(1-r)S}{2(a-1)}\left[\sum_{k=0}^{T-2}\frac{1}{a+k} - \frac{T-1}{a+T-1}\right] \\
&\geq (1-r)\frac{\log^2(1-r)S}{2(a-1)}\left[\log(a+T-2) - \log(a) + \frac{1}{a} - \frac{T-1}{a+T-1}\right],
\end{aligned}$$

where (a) follows from Lemma A2.9.c) in appendix A2. This completes the proof. □

■ □ ■

Proof of Proposition 2.4. We proceed by induction. For $T = 1$ we have $V_1^m(a,S) = C(a,S)$ and the result holds by Lemma A2.2.

Suppose that the result holds for $T-1$. Then

$$\begin{aligned}
V_T^m(a,S) &:= C(a,S) + \int_0^{y^m(a,S)} V_{T-1}^m(a+1, S+z^\ell) m(z|a,S)dz + \\
&\quad (1-r)V_{T-1}^m\left(a, S + [y^m(a,S)]^\ell\right) \\
&= C(a,S) + \int_0^{y^m(a,S)} (S+z^\ell)^{1/\ell} m(z|a,S)dz \; V_{T-1}^m(a+1,1) + \\
&\quad (1-r)\left[S + [y^m(a,S)]^\ell\right]^{1/\ell} V_{T-1}^m(a,1) \\
&= C(a,S) + \int_0^{y^m(a,S)} \frac{aS^a \ell z^{\ell-1}}{(S+z^\ell)^{a+1-\frac{1}{\ell}}} dz \; V_{T-1}^m(a+1,1) + \\
&\quad (1-r)S^{1/\ell}(1-r)^{-1/a\ell} \; V_{T-1}^m(a,1) \\
&= S^{1/\ell}C(a,1) - a\frac{S^a}{(a-\frac{1}{\ell})}\frac{1}{(S+z^\ell)^{a-\frac{1}{\ell}}}\Bigg|_0^{y^m(a,S)} V_{T-1}^m(a+1,1) + \\
&\quad S^{1/\ell}(1-r)^{1-1/a\ell} \; V_{T-1}^m(a,1) \\
&= S^{1/\ell}C(a,1) - a\frac{S^{1/\ell}}{(a-\frac{1}{\ell})}\left[1-(1-r)^{1-1/a\ell}\right] \; V_{T-1}^m(a+1,1) + \\
&\quad S^{1/\ell}(1-r)^{1-1/a\ell} \; V_{T-1}^m(a,1) \\
&= S^{1/\ell}\Bigg[C(a,1) - \frac{a\ell}{a\ell-1}\left[1-(1-r)^{1-1/a\ell}\right] \; V_{T-1}^m(a+1,1) + \\
&\quad (1-r)^{1-1/a\ell} \; V_{T-1}^m(a,1)\Bigg],
\end{aligned}$$

where the second equality follows from the inductive hypothesis applied to $V_{T-1}^m(a+1, S+z^\ell)$ and $V_{T-1}^m(a, S+[y^m(a,S)]^\ell)$, and the third equality follows from the definition of $y^m(a,S)$. We deduce that the result holds for T; this completes the induction argument and the proof. □

A2 Technical Lemmas

Lemma A2.1.

$$\begin{aligned}
C(a,S) &= \frac{1}{1-r}\mathbb{E}[D\mathbb{I}_{\{D\geq y^m(a,S)\}}] - E[D], \\
C(\theta) &= \frac{1}{1-r}\mathbb{E}[D\mathbb{I}_{\{D\geq y^m(\theta)\}}|\theta] - E[D|\theta].
\end{aligned}$$

Proof.

$$\begin{aligned}
C(a,S) &= \mathbb{E}[L(y^m(a,S),D)] \\
&= \mathbb{E}[(y^m(a,S)-D)^+] + \frac{r}{1-r}\mathbb{E}[(D-y^m(a,S))^+] \\
&= (y^m(a,S)-\mathbb{E}[D]) + \left(1+\frac{r}{1-r}\right)\mathbb{E}\left[(D-y^m(a,S))^+\right] \\
&= (y^m(a,S)-\mathbb{E}[D]) + \frac{1}{1-r}\Big[\mathbb{E}\left[D\mathbb{I}_{\{D\geq y^m(a,S)\}}\right] - (1-r)y^m(a,S)\Big] \\
&= \frac{1}{1-r}\mathbb{E}[D\mathbb{I}_{\{D\geq y^m(a,S)\}}] - E[D].
\end{aligned}$$

This completes the proof for $C(a,S)$. The proof for $C(\theta)$ is analogous. □

■ □ ■

The following Lemma has been proven in the literature (see, for example, [Azoury, 1985]), but we include a proof here for completeness:

Lemma A2.2 (Scalability)**.** *Suppose demands are Weibull and* $a\ell > 1$*, then*

$$C(a,S) = S^{1/\ell}C(a,1).$$

Proof. By Lemma A2.1,

$$\begin{aligned}
C(a,S) &= \frac{1}{1-r}\mathbb{E}[D\mathbb{I}_{\{D\geq y^m(a,S)\}}] - E[D] \\
&= \frac{1}{1-r}\int_{y^m(a,S)}^{\infty} zm(z|a,S)dz - \int_0^{\infty}\overline{M}(z|a,S)dz.
\end{aligned}$$

The myopic order quantity is given by $y^m(a,S) = S^{1/\ell}\left((1-r)^{-1/a}-1\right)^{1/\ell}$ and hence $y^m(a,S) = S^{1/\ell}y^m(a,1)$. Also, one has that

$$\begin{aligned}
\int_{y^m(a,S)}^{\infty} zm(z|a,S)dz &= \int_{y^m(a,S)}^{\infty}\frac{aS^a\ell z^\ell}{(S+z^\ell)^{a+1}}dz \\
&= \int_{S^{1/\ell}y^m(a,1)}^{\infty}\frac{az^\ell}{S(1+\frac{z^\ell}{S})^{a+1}}dz \\
&= S^{1/\ell}\int_{y^m(a,1)}^{\infty}\frac{az^\ell}{(1+z^\ell)^a}dz \\
&= S^{1/\ell}\int_{y^m(a,1)}^{\infty} zm(z|a,1)dz.
\end{aligned}$$

Similarly, it is not hard to show that

$$\int_0^\infty \overline{M}(z|a,S)dz = S^{1/\ell} \int_0^\infty \overline{M}(z|a,1)dz,$$

and hence

$$C(a,S) = S^{1/\ell} \left[\frac{1}{1-r} \int_{y^m(a,1)}^\infty zm(z|a,1)dz - \int_0^\infty \overline{M}(z|a,1)dz \right] = S^{1/\ell} C(a,1).$$

This completes the proof. □

■ □ ■

Lemma A2.3. *Let (D,θ) follow a Newsvendor distribution with parameters (a,S), and (D',θ) a Newsvendor distribution with parameters $(a,S+d(y))$. For any $x,y \geq 0$,*

$$\mathbb{P}\big(d(D) \geq x | D \geq y\big) = \mathbb{P}\big(d(y) + d(D') \geq x\big).$$

In words, this lemma states that the conditional distribution of $d(D)$, given that D is greater than y, is equivalent to the unconditional distribution of $d(y) + d(D')$, where D' follows the resulting distribution after having one censored observation at y.

Proof. We start by showing the following equivalence:

$$\mathbb{P}\big(d(D) \geq x | D \geq y\big) = \mathbb{P}\big(d(y) + d(D') \geq x\big) \quad \forall x,y \geq 0 \tag{A.18}$$

$$\Leftrightarrow \quad \mathbb{P}\big(d(D) \geq d(x') | D \geq y'\big) = \mathbb{P}\big(d(y') + d(D') \geq d(x')\big) \quad \forall x',y' \geq 0. \tag{A.19}$$

To show necessity note that, given x, if there exists x' such that $d(x') = x$ the result follows immediately by setting $x' := d^{-1}(x)$ and $y' := y$. If no such x' exists, this means, since $d(\cdot)$ is increasing and unbounded, that $x < d(0)$. But this implies that the events $\{d(D) \geq x\}$ and $\{d(y) + d(D') \geq x\}$ equal the entire sample space, and hence both sides of equation (A.18) equal 1. This completes the proof for the necessity implication; the sufficiency can be proven by analogous arguments.

We now give a proof of equation (A.19). If $x' < y'$ both sides of equation (A.19) are equal to 1 and the result holds. Suppose that $x' \geq y'$. Then

$$\begin{aligned}
\mathbb{P}\big(d(D) \geq d(x')|D \geq y'\big) &= \frac{\mathbb{P}(d(D) \geq d(x'))}{\mathbb{P}(D \geq y')} \\
&= \int_\Theta \frac{\mathbb{P}(d(D) \geq d(x')|\theta)}{\mathbb{P}(D \geq y')} \pi(\theta|a,S)d\theta \\
&= \int_\Theta \frac{\mathbb{P}(d(D) \geq d(x')|\theta)}{\mathbb{P}(D \geq y'|\theta)} \frac{\mathbb{P}(D \geq y'|\theta)\pi(\theta|a,S)}{\mathbb{P}(D \geq y')} d\theta, \quad \text{(A.20)}
\end{aligned}$$

where the first equality comes from the fact that the events $\{D \geq y'\}$ and $\{d(D) \geq d(y')\}$ are equivalent, and $x' \geq y'$ implies $d(x') \geq d(y')$.

Note that the second fraction inside the integral in (A.20) is equivalent to the posterior distribution of θ, given a censored observation of $D \geq y$. That is,

$$\frac{\mathbb{P}(D \geq y'|\theta)\pi(\theta|a,S)}{\mathbb{P}(D \geq y')} = \pi(\theta|a, S + d(y')),$$

where the equality follows from the Bayes update rule of a censored observation in the Newsvendor family. By replacing this expression in (A.20) one obtains

$$\begin{aligned}
&\int_\Theta \frac{\mathbb{P}(d(D) \geq d(x')|\theta)}{\mathbb{P}(D \geq y'|\theta)} \frac{\mathbb{P}(D \geq y'|\theta)\pi(\theta|a,S)}{\mathbb{P}(D \geq y')} d\theta \\
&= \int_\Theta \frac{\mathbb{P}(d(D) \geq d(x')|\theta)}{\mathbb{P}(D \geq y'|\theta)} \pi(\theta|a, S + d(y'))d\theta \\
&= \int_\Theta e^{-\theta(d(x') - d(y'))} \pi(\theta|a, S + d(y'))d\theta \\
&= \int_\Theta \mathbb{P}\big(d(D) \geq d(x') - d(y')|\theta\big) \pi(\theta|a, S + d(y'))d\theta \\
&= \mathbb{P}\big(d(y') + d(D') \geq d(x')\big),
\end{aligned}$$

where the third equality comes from the fact that the distribution of $d(D)$ conditional on θ is exponential with parameter θ. This completes the proof. □

■ □ ■

Lemma A2.4. *For any $t \geq 0$ we have*

$$C_t^m(a,S) \geq C_t^o(a,S).$$

Proof. To simplify the exposition, we will write the proof for $t = 1$, that is, one period in the future. The extension to general t follows with similar reasonings.

Suppose D_1 represents the demand realization in the first period. One can write the future costs as

$$\begin{aligned} C_1^m(a,S) &= \mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1 \wedge y^m\right]\right] \\ C_1^o(a,S) &= \mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1\right]\right]. \end{aligned} \tag{A.21}$$

Note that if we condition the outer expectation in (A.21) on the event $\{D_1 < y^m\}$ we obtain

$$\mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1\right]\bigg| D_1 < y^m\right] = \mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1 \wedge y^m\right]\bigg| D_1 < y^m\right]. \tag{A.22}$$

Conditioning on the complementary event in (A.21) one obtains

$$\begin{aligned} \mathbb{E}\left[\min_y \mathbb{E}[L(y,D)\big|D_1]\bigg|D_1 \geq y^m\right] &\leq \min_y \mathbb{E}\left[\mathbb{E}[L(y,D)\big|D_1]\Big|D_1 \geq y^m\right] \\ &= \min_y \mathbb{E}[L(y,D)|D_1 \geq y^m] \\ &= \min_y \mathbb{E}[L(y,D)|D_1 \wedge y^m, D_1 \geq y^m] \\ &= \mathbb{E}\left[\min_y \mathbb{E}[L(y,D)|D_1 \wedge y^m]\bigg| D_1 \geq y^m\right]. \end{aligned} \tag{A.23}$$

By combining (A.22) and (A.23) one obtains

$$C_1^o(a,S) = \mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1\right]\right] \leq \mathbb{E}\left[\min_y \mathbb{E}\left[L(y,D)\big|D_1 \wedge y^m\right]\right] = C_1^m(a,S).$$

This completes the proof. □

■ □ ■

Lemma A2.5. *Suppose Demand is Weibull, with $a\ell > 1$. For any a, S, l, r and $t = 1, \ldots, T$*

$$C_t^o(a,S) - C_\infty^o(a,S) \leq S^{1/\ell} K(r,\ell) \left[\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{1/\ell} \frac{1}{a+t-1/\ell}.$$

where $K(r,\ell)$ is a constant depending only on r and l.

Proof. Let $(D_1, \ldots, D_T)$ denote the vector of demands, over the time horizon, and let $S_t := S + D_1^\ell + \ldots + D_t^\ell$ and $a_t := a + t$.

By Lemma A2.1 one has

$$\begin{aligned} C_t^o(a,S) &= \frac{1}{1-r}\mathbb{E}_{a,S}\left[\int_{y^m(a_t,S_t)}^{\infty} x m(x|a_t,S_t)dx\right] - \mathbb{E}_{a,S}\left[\mathbb{E}_{a_t,S_t}[D]\right] \\ C_\infty^o(a,S) &= \frac{1}{1-r}\mathbb{E}_{a,S}\Big[\int_{y^m(\theta)}^{\infty} x f(x|\theta)dx\Big] - \mathbb{E}_{a,S}\left[\mathbb{E}[D|\theta]\right]. \end{aligned}$$

Note that the last terms on the right side of both equations above are equal to $\mathbb{E}_{a,S}[D]$. Therefore, by subtracting both equations one obtains

$$\begin{aligned}
&(1-r)[C_t^o(a,S) - C_\infty^o(a,S)] \\
&= \mathbb{E}_{a,S}\left[\int_{y^m(a_t,S_t)}^\infty x m(x|a_t,S_t)dx\right] - \mathbb{E}_{a,S}\left[\int_{y^m(\theta)}^\infty x f(x|\theta)dx\right] \\
&= \mathbb{E}_{a,S}\left[\int_{y^m(a_t,S_t)}^\infty x\frac{a_t S_t^{a_t}\ell x^{\ell-1}}{(S_t+x^\ell)^{a_t+1}}dx\right] - \mathbb{E}_{a,S}\left[\int_{y^m(\theta)}^\infty x\theta\ell x^{\ell-1}e^{-\theta x^\ell}dx\right] \\
&= a_t\mathbb{E}_{a,S}\left[S_t^{1/\ell}\right]\int_{(1-r)^{-\frac{1}{a_t}}-1}^\infty \frac{u^{1/\ell}}{(1+u)^{a_t+1}}du - \mathbb{E}_{a,S}\left[\theta^{-1/\ell}\right]\int_{-\log(1-r)}^\infty u^{1/\ell}e^{-u}du, \quad \text{(A.24)}
\end{aligned}$$

where the last equality follows from the change of variable $u := x^\ell/S_t$ and $u := \theta x^\ell$ in the first and second integrals, respectively.

Note the law of iterated expectations and an application of Jensen's inequality yield that

$$\mathbb{E}_{a,S}\left[\theta^{-1/\ell}\right] = \mathbb{E}_{a,S}\left[\mathbb{E}_{a_t,S_t}\left[\theta^{-1/\ell}\right]\right] \geq \mathbb{E}_{a,S}\left[\mathbb{E}_{a_t,S_t}\left[\theta\right]^{-1/\ell}\right] = \mathbb{E}_{a,S}\left[\frac{S_t^{1/\ell}}{a_t^{1/\ell}}\right].$$

Returning to (A.24), one obtains

$$\begin{aligned}
&(1-r)\left[C_t^o(a,S) - C_\infty^o(a,S)\right] \leq \\
&\quad \mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right]\left[a_t^{1/\ell+1}\int_{(1-r)^{-\frac{1}{a_t}}-1}^\infty \frac{u^{1/\ell}}{(1+u)^{a_t+1}}du - \int_{-\log(1-r)}^\infty u^{1/\ell}e^{-u}du\right] \\
&= \mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right]\left[\left(\frac{a_t}{a_t+1}\right)^{1/\ell+1}\int_{(a_t+1)\left((1-r)^{-\frac{1}{a_t}}-1\right)}^\infty \frac{u^{1/\ell}}{(1+\frac{u}{a_t+1})^{a_t+1}}du - \right. \\
&\qquad\qquad \left.\int_{-\log(1-r)}^\infty u^{1/\ell}e^{-u}du\right] \\
&\leq \mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right]\left[\int_{-\log(1-r)}^\infty \frac{u^{1/\ell}}{(1+\frac{u}{a_t+1})^{a_t+1}}du - \int_{-\log(1-r)}^\infty u^{1/\ell}e^{-u}du\right] \\
&\leq \mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right]\left[(a_t+1)^{1/\ell+1}\int_0^\infty \frac{u^{1/\ell}}{(1+u)^{a_t+1}}du - \int_0^\infty u^{1/\ell}e^{-u}du\right]. \quad \text{(A.25)}
\end{aligned}$$

The second inequality follows from the fact that $a_t/(a_t+1) < 1$ and the fact that $(a_t + 1)\left[(1-r)^{-\frac{1}{a_t}} - 1\right] > a_t\left[(1-r)^{-\frac{1}{a_t}} - 1\right] > -\log(1-r) > 0$ (see, for example, Lemma A2.9.a). The last inequality follows from the fact that the first integrand is grater than or equal to the second one.

Consider now the last term in square braces. One can rewrite the integrals using the Gamma function as follows

$$
\begin{aligned}
&(a_t+1)^{1/\ell+1}\int_0^\infty \frac{u^{1/\ell}}{(1+u)^{a_t+1}}du - \int_0^\infty u^{1/\ell}e^{-u}du \\
&= (a_t+1)^{1/\ell+1}\frac{\Gamma(1/\ell+1)\Gamma(a_t-1/\ell)}{\Gamma(a_t+1)} - \Gamma(1/\ell+1) \\
&= \Gamma(1/\ell+1)\left[\frac{(a_t+1)^{1/\ell+1}\Gamma(a_t-1/\ell)}{\Gamma(a_t+1)} - 1\right] \\
&\leq \Gamma(1/\ell+1)\widehat{K}(1/\ell+1)\frac{1}{a_t-1/\ell}, \qquad \text{(A.26)}
\end{aligned}
$$

where the first equality follows from basic properties of the Gamma and Beta functions[4] (see, for example, [Abramowitz and others, 1964], Chapter 6) and the last inequality follows from Lemma A2.6. In particular, $\widehat{K}(\cdot)$ is defined in (A.28).

Therefore (A.25) and (A.26) yield

$$
C_t^o(a,S) - C_\infty^o(a,S) \;\leq\; \mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right]\frac{\Gamma(1/\ell+1)\widehat{K}(1/\ell+1)}{1-r}\frac{1}{a_t-1/\ell}. \qquad \text{(A.27)}
$$

Furthermore, it is possible to show (see Lemma A2.7) that

$$
\mathbb{E}_{a,S}\left[\left(\frac{S_t}{a_t}\right)^{1/\ell}\right] \leq \left[S\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{\frac{1}{\ell}}
$$

and hence, by setting

$$
K(r,\ell) := \frac{\Gamma(1/\ell+1)\widehat{K}(1/\ell+1)}{1-r},
$$

one obtains

$$
C_t^o(a,S) - C_\infty^o(a,S) \leq S^{1/\ell}K(r,\ell)\left[\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{1/\ell}\frac{1}{a_t-1/\ell}.
$$

This completes the proof. □

■ □ ■

[4] In particular we are using the definition of the Gamma function: $\Gamma(x) := \int_0^\infty u^{x-1}e^{-u}du$, the following properties of the beta function: $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^\infty \frac{u^{x-1}}{(u+1)^{x+y}}du$, and letting $x := 1/\ell + 1$ and $y := a_t - 1/\ell$.

Lemma A2.6. *Let a, b be positive real numbers such that $a > b$. Then*

$$\frac{a^b \Gamma(a-b)}{\Gamma(a)} - 1 \leq \frac{\widehat{K}(b)}{a-b},$$

where $\widehat{K}(\cdot)$ is given by

$$\widehat{K}(x) = \begin{cases} 0 & \text{if } x = 0, \\ x\left(\widehat{K}(x-1)+1\right) & \text{if } x \in \mathbb{N}, \\ \widehat{K}(\lfloor x \rfloor) + x - \lfloor x \rfloor & \text{if } x \in \mathbb{R} \setminus \mathbb{N}. \end{cases} \tag{A.28}$$

Proof. i.) If $0 \leq b < 1$ the results follows directly from a traditional bound on the gamma function, first developed by [Wendel, 1948] (a proof can be found in [Qi and Losonczi, 2010]).[5]

ii.) Suppose $b = n \in \mathbb{N}$, the result follows from the inequality (A.29) we establish next and the factorial property of the gamma function: $\Gamma(x) = (x-1)\Gamma(x-1) \;\; \forall x \geq 1$.

Let $n \in \mathbb{N}$ and $a \in \mathbb{R}_+$ such that $a > n$. We establish the following inequality by induction on n:

$$\frac{a^n}{(a-1)(a-2)\ldots(a-n)} - 1 \leq \frac{\widehat{K}(n)}{a-n}, \tag{A.29}$$

where $\widehat{K}(n)$ was defined in (A.28). The base case, $n = 1$, is trivial. Suppose the inequality holds for $n-1$, then

$$\begin{aligned} \frac{a^n}{(a-1)(a-2)\ldots(a-n)} - 1 &= \frac{a^{n-1}}{(a-1)(a-2)\ldots(a-(n-1))}\frac{a}{a-n} - 1 \\ &\overset{(a)}{\leq} \left(\frac{\widehat{K}(n-1)}{a-n+1} + 1\right)\frac{a}{a-n} - 1 \\ &= \frac{\left(\widehat{K}(n-1) + a - n + 1\right)a - (a-n+1)(a-n)}{(a-n+1)(a-n)} \\ &= \frac{a\widehat{K}(n-1) + (a-n+1)n}{(a-n+1)(a-n)} \\ &\overset{(b)}{=} \frac{1}{a-n}\left[\frac{a/n}{a-n+1}(\widehat{K}(n)-n) + n\right] \\ &\overset{(c)}{\leq} \frac{1}{a-n}\widehat{K}(n), \end{aligned}$$

[5] The original bound is written as: $\left(\frac{x}{x+s}\right)^{(1-s)} \leq \frac{\Gamma(x+s)}{x^s\Gamma(x)}$ for any $0 < s < 1$ and $x > 0$. Case *i.*) can be derived from this bound by letting $x := a - b$ and $s := b$.

where (a) is a consequence of the induction hypothesis, (b) follows by the recursive definition of $\widehat{K}(n)$ and (c) is a result of the fact that $a/n \le a-n+1$ for all $a \ge n \ge 0$. This concludes the induction argument.

iii.) Finally, suppose $b \in \mathbb{R} \setminus \mathbb{N}$ and $b > 1$. Let $n := \lfloor b \rfloor$ and $\epsilon := b - \lfloor b \rfloor$. Then

$$\begin{aligned}
\frac{a^b \Gamma(a-b)}{\Gamma(a)} &= \frac{a^{n+\epsilon}\Gamma(a-n-\epsilon)}{\Gamma(a)} \\
&= \frac{a^n}{(a-1)(a-2)\dots(a-n)} \frac{a^\epsilon \Gamma(a-n-\epsilon)}{\Gamma(a-n)} \\
&\le \left(\frac{\widehat{K}(n)}{a-n} + 1\right) \frac{a^\epsilon \Gamma(a-n-\epsilon)}{\Gamma(a-n)} \\
&\le \left(\frac{\widehat{K}(n)}{a-n} + 1\right) \left(1 + \frac{\epsilon}{a-n-\epsilon}\right) \\
&= \frac{\widehat{K}(n)}{a-n-\epsilon} + \frac{a-n}{a-n-\epsilon} \\
&= \frac{\widehat{K}(n)+\epsilon}{a-n-\epsilon} + 1 \\
&= \frac{\widehat{K}(n) + b - \lfloor b \rfloor}{a-b} + 1 \\
&= \frac{\widehat{K}(b)}{a-b} + 1
\end{aligned}$$

where in the first inequality, we have use (A.29) and in the second inequality, we have used the fact that $\epsilon < 1$. This completes the proof. □

■ □ ■

Lemma A2.7. *Fix $S > 0$ and $a > 1/\ell$. Let $D_1, \dots, D_t$ be iid random variables, Weibull distributed with parameter $\theta \sim Gamma(a, S)$. Then*

$$\mathbb{E}\left[\left(\frac{S + D_1^\ell + \ldots + D_t^\ell}{a+t}\right)^{1/\ell}\right] \le \left[S \frac{\exp\{1/(a-1/\ell)\}}{a - 1/\ell}\right]^{\frac{1}{\ell}}.$$

Proof. First, recall that

$$\mathbb{E}_D[(S + D^\ell)] = \int_0^\infty (S + z^\ell) m(z|a, S) dz = S^{1/\ell} \int_0^\infty (1 + z^\ell) m(z|a, 1) dz = S^{1/\ell} \frac{a\ell}{a\ell - 1}.$$

Therefore, one has that

$$\begin{aligned}
&\mathbb{E}\left[\left(S+D_1^\ell+\ldots+D_t^\ell\right)^{1/\ell}\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\left(S+D_1^\ell+\ldots+D_t^\ell\right)^{1/\ell}\Big|D_1,\ldots,D_{t-1}\right]\right]\\
&= \mathbb{E}\left[\left(S+D_1^\ell+\ldots+D_{t-1}^\ell\right)^{1/\ell}\right]\int_0^\infty(1+z^\ell)m(z|a+t-1,1)dz\\
&= \mathbb{E}\left[\left(S+D_1^\ell+\ldots+D_{t-1}^\ell\right)^{1/\ell}\right]\frac{(a+t-1)\ell}{(a+t-1)\ell-1}.
\end{aligned}$$

Continuing recursively one obtains

$$\begin{aligned}
&\mathbb{E}\left[\left(S+D_1^\ell+\ldots+D_t^\ell\right)^{1/\ell}\right]\\
&= S^{\frac{1}{\ell}}\frac{a\ell}{a\ell-1}\frac{(a+1)\ell}{(a+1)\ell-1}\cdots\frac{(a+t-1)\ell}{(a+t-1)\ell-1}\\
&= S^{\frac{1}{\ell}}\frac{a}{a-\frac{1}{\ell}}\frac{a+1}{a+1-\frac{1}{\ell}}\cdots\frac{a+t-1}{a+t-1-\frac{1}{\ell}}\\
&= S^{\frac{1}{\ell}}\prod_{k=0}^{t-1}\left(1+\frac{1}{\ell a-1+\ell k}\right)\\
&\leq S^{\frac{1}{\ell}}\exp\left\{\sum_{k=0}^{t-1}\frac{1}{\ell a-1+\ell k}\right\}\\
&\leq S^{\frac{1}{\ell}}\exp\left\{\frac{1}{\ell}\left[\log(\ell a-1+\ell(t-1))-\log(\ell a-1)+\frac{1}{a-1/\ell}\right]\right\}\\
&= S^{\frac{1}{\ell}}\left[1+\frac{(t-1)}{a-1/\ell}\right]^{\frac{1}{\ell}}\left[e^{\frac{1}{a-1/\ell}}\right]^{\frac{1}{\ell}}.
\end{aligned}$$

And hence

$$\begin{aligned}
\mathbb{E}_{\hat{D}}\left[\left(\frac{S+D_1^\ell+\ldots+D_t^\ell}{a+t}\right)^{1/\ell}\right] &\leq S^{\frac{1}{\ell}}\left[\frac{a+t-1-1/\ell}{(a-1/\ell)(a+t)}\right]^{\frac{1}{\ell}}\left[e^{\frac{1}{a-1/\ell}}\right]^{\frac{1}{\ell}}\\
&\leq \left[S\frac{\exp\{1/(a-1/\ell)\}}{a-1/\ell}\right]^{\frac{1}{\ell}},
\end{aligned}$$

which concludes the proof. □

■ □ ■

Lemma A2.8. *Suppose demand are exponential. Let*

$$\Gamma_T^r(a,S) := (1-r)\left[V_T^r(a,S)-\frac{a}{a-1}V_T^r(a+1,S)\right].$$

Then, for any $a > 1$*,* $T \geq 1$*,* $S > 0$ *and* $r \in (0,1)$

$$\Gamma_T^r(a,S) \geq \frac{1-r}{a-1}\left[(a-1)C(a,S) - (a+T-1)C(a+T,S)\right].$$

Proof. We proceed by induction on T. The base case, $T = 1$, follows directly from the definition of $\Gamma_1^r(a,S)$. Suppose the result holds for $T-1$. Then

$$\begin{aligned}
\frac{\Gamma_T^r(a,S)}{1-r} &= V_T^r(a,S) - \frac{a}{a-1}V_T^r(a+1,S) \\
&= C(a,S) + r\frac{a}{a-1}V_{T-1}^r(a+1,S) + (1-r)V_{T-1}^r(a,S) - \\
&\quad \frac{a}{a-1}\left[C(a+1,S) + r\frac{a+1}{a}V_{T-1}^r(a+2,S) + (1-r)V_{T-1}^r(a+1,S)\right] \\
&= \frac{\Gamma_1^r(a,S)}{1-r} + \frac{r}{1-r}\frac{a}{a-1}\Gamma_{T-1}^r(a+1,S) + \Gamma_{T-1}^r(a,S) \\
&\geq \frac{1}{a-1}\left[(a-1)C(a,S) - aC(a+1,S)\right] + \\
&\quad \frac{r}{a-1}\left[aC(a+1,S) - (a+T-1)C(a+T,S)\right] + \\
&\quad \frac{(1-r)}{a-1}\left[(a-1)C(a,S) - (a+T-2)C(a+T-1,S)\right] \\
&= \frac{1}{a-1}\left[(a-1)C(a,S) - (a+T-1)C(a+T,S)\right] + \\
&\quad \frac{1-r}{a-1}\Big[(a-1)C(a,S) - aC(a+1,S) + \\
&\quad (a+T-1)C(a+T,S) - (a+T-2)C(a+T-1,S)\Big] \\
&\geq \frac{1}{a-1}\left[(a-1)C(a,S) - (a+T-1)C(a+T,S)\right],
\end{aligned}$$

where the first inequality follows from the inductive hypothesis and the second inequality follows from the fact that the function $f(x) := (x-1)C(x,S)$ is a decreasing convex function (see Lemma A2.9 below). This completes the induction step. □

■ □ ■

Lemma A2.9. *Let* $f : (1,+\infty) \to \mathbb{R}$*, defined as* $f(a) := (a-1)C(a,1) = a\left[(1-r)^{-1/a} - 1\right]$*. Then, for every* $a > 1$

a) $f(a) + \log(1-r) = \sum_{k=2}^{\infty} \frac{(-\log(1-r))^k}{k!}\frac{1}{a^{k-1}}$ *for any* $a > 1$*.*

b) $f(\cdot)$ *is decreasing and convex.*

c) $f(a) - f(b) \geq \dfrac{\log^2(1-r)}{2}\left[\dfrac{1}{a} - \dfrac{1}{b}\right]$, *for any* $b \geq a$.

Proof. **a)**

$$\begin{aligned} f(a) + \log(1-r) &= a\big[(1-r)^{-1/a} - 1\big] + \log(1-r) \\ &= a\big[e^{-\log(1-r)/a} - 1 - \frac{-\log(1-r)}{a}\big] \\ &= a\sum_{k=2}^{\infty}\left[\frac{(-\log(1-r))^k}{k!}\frac{1}{a^k}\right] \\ &= \sum_{k=2}^{\infty}\left[\frac{(-\log(1-r))^k}{k!}\frac{1}{a^{k-1}}\right] \end{aligned}$$

where the third equality follows from the Taylor exp of e^x.

b) Follows directly from a), since the terms in the sum are positive, decreasing and convex functions.

c) Follows from a) by subtracting the two series and discarding all terms except $k = 2$.

□

■ □ ■

Appendix B

Appendix to Chapter 3

Lemma B.1.

$$\Gamma_T(\pi) \leq \sum_{k=0}^{T-1} (1-q(\mu(\pi)))^k \left[\mathbb{E}[Q^o(\pi^o_{T-k+1})] - Q(\pi)\right] \leq \frac{1-(1-q(\mu(\pi)))^T}{q(\mu(\pi))} \left[Q_\infty(\pi) - Q(\pi)\right]$$

Proof. We have

$$\begin{aligned}
\Gamma_T(\pi) &:= \mathbb{E}[V^m_T(\pi \oplus X)] - V^m_T(\pi) \\
&= \mathbb{E}[V^m_T(\pi \oplus X)] - \Big[Q(\pi) + q(\mu(\pi))\mathbb{E}[V^m_{T-1}(\pi \oplus X)] + (1-q(\mu(\pi)))V^m_{T-1}(\pi)\Big] \\
&= \mathbb{E}[V^m_T(\pi \oplus X)] - \mathbb{E}[V^m_{T-1}(\pi \oplus X)] - Q(\pi) + \\
&\quad (1-q(\mu(\pi)))\Big[\mathbb{E}[V^m_{T-1}(\pi \oplus X)] - V^m_{T-1}(\pi)\Big] \\
&= \mathbb{E}[V^m_T(\pi \oplus X)] - \mathbb{E}[V^m_{T-1}(\pi \oplus X)] - Q(\pi) + (1-q(\mu(\pi)))\Gamma_{T-1}(\pi) \\
&= \mathbb{E}[Q^m_T(\pi \oplus X)] - Q(\pi) + (1-q(\mu(\pi)))\Gamma_{T-1}(\pi),
\end{aligned}$$

where $Q^m_t(\pi)$ represents the future one period reward, t+1 periods in the future, when the myopic policy is applied. Proceeding recursively one obtains

$$\begin{aligned}
\Gamma_T(\pi) &= \sum_{k=0}^{T-1} (1-q(\mu(\pi)))^k \left[\mathbb{E}[Q^m_{T-k}(\pi \oplus X)] - Q(\pi)\right] \\
&\leq \sum_{k=0}^{T-1} (1-q(\mu(\pi)))^k \left[\mathbb{E}[Q^o_{T-k}(\pi \oplus X)] - Q(\pi)\right] \\
&= \sum_{k=0}^{T-1} (1-q(\mu(\pi)))^k \left[\mathbb{E}[Q(\pi^o_{T-k+1})] - Q(\pi)\right],
\end{aligned}$$

where the inequality comes from the fact that $Q^m_k(\pi) \leq \mathbb{E}[Q(\pi^o_k)]$. □

■ □ ■

Lemma B.2. *Suppose $\pi(\cdot)$, X and $R(X)$ are defined as in the Example 3.3, in Section 3.2.1, and suppose $1/q(x) := O(x^{-\alpha})$, with $a+\alpha > 2$ and $s_1 > \alpha + 1$. Then*

$$\mathbb{E}\left[q^{-1}(\mu(\pi_t^o))\,\mathrm{Var}_{\widehat{\nu}\sim\pi_t^o}(\mu_{\widehat{\nu}})\right] = O(1/t).$$

Proof. Let us start by recalling that

$$\begin{aligned}
\mu_\nu &:= \mathbb{E}[R(X)|\nu] = \nu_\epsilon\,\nu_Y^{-1} \\
\mu(\pi) &:= \mathbb{E}[R(X)] = \mathbb{E}[\epsilon]\mathbb{E}[Y] = \frac{s_1}{s_1+s_2}\frac{S}{a-1},
\end{aligned}$$

and the Bayesian updates are given by

$$\begin{aligned}
s_1' &= s_1 + \sum_{k=1}^{t}\epsilon_k \\
s_2' &= s_2 + t - \sum_{k=1}^{t}\epsilon_k \\
a' &= a + \sum_{k=1}^{t}\epsilon_k \\
S' &= S + \sum_{k=1}^{t} Y_k\epsilon_k
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Var}_{\widehat{\nu}\sim\pi}(\mu_{\widehat{\nu}}) &:= \mathbb{E}[\mu_\nu^2] - \mathbb{E}[\mu_\nu]^2 \\
&= \mathbb{E}[\nu_\epsilon^2]\mathbb{E}[\nu_Y^{-2}] - \mathbb{E}[\nu_\epsilon]^2\mathbb{E}[\nu_Y^{-1}]^2 = \mathrm{Var}(\nu_\epsilon^2)E[\nu_Y^{-2}] + \mathbb{E}[\nu_\epsilon]^2\,\mathrm{Var}(\nu_Y) \\
&\overset{(i)}{=} \frac{\mu(\pi_\epsilon)(1-\mu(\pi_\epsilon))}{s_1+s_2+1}\frac{S^2}{(a-1)^2}\frac{a-1}{a-2} + \mu(\pi_\epsilon)^2\frac{S^2}{(a-1)^2(a-2)} \\
&\leq \frac{S^2}{(a-1)^2}\left(\frac{1}{s_1+s_2+1}\frac{a-1}{a-2} + \frac{1}{a-2}\right)
\end{aligned}$$

where (i) comes from the expressions for the first and second moments of the Beta and Inverse Gamma distributions. Let $t_\epsilon := \sum_{k=1}^{t}\epsilon_k$. We have,

$$\begin{aligned}
& q^{-1}(\mu(\pi_t^o))\,\mathrm{Var}_{\widehat{\nu}\sim\pi_t^o}(\mu_{\widehat{\nu}}) \\
&\leq \mu(\pi)^{-\alpha}\left(\frac{S+\sum_{k=1}^{t}Y_k\epsilon_k}{a+t_\epsilon-1}\right)^2\left(\frac{1}{s_1+s_2+t+1}\frac{a+t_\epsilon-1}{a+t_\epsilon-2} + \frac{1}{a+t_\epsilon-2}\right) \\
&\leq \left(\frac{S+\sum_{k=1}^{t}Y_k\epsilon_k}{a+t_\epsilon-1}\right)^{2-\alpha}\left(\frac{s_1+t_\epsilon}{s_1+s_2+t}\right)^{-\alpha} O\left(\frac{1}{a+t_\epsilon-2}\right) \qquad \text{(B.1)}
\end{aligned}$$

Note that, since $\sum_{k=1}^{t} Y_k \epsilon_k \overset{\mathcal{D}}{=} \sum_{k=1}^{t_\epsilon} Y_k$, by conditioning on t_ϵ in (B.1) the first term can be show to be bounded if $\alpha + a > 2$ for any t_ϵ, following the same arguments of Example 3.2 (recall that the Exponential distribution is a member the Weibull family, with $\ell = 1$). Similarly, we have

$$\left(\frac{s_1 + t_\epsilon}{s_1 + s_2 + t}\right)^{-\alpha} O\left(\frac{1}{a + t_\epsilon - 2}\right) = O\left(\left(\frac{s_1 + t_\epsilon}{s_1 + s_2 + t_\epsilon}\right)^{-\alpha-1} \frac{1}{t}\right),$$

and therefore

$$\mathbb{E}\left[q^{-1}(\mu(\pi_t^o)) \operatorname{Var}_{\widehat{\nu} \sim \pi_t^o}(\mu_{\widehat{\nu}})\right] = O\left(\mathbb{E}\left[\left(\frac{s_1 + t_\epsilon}{s_1 + s_2 + t_\epsilon}\right)^{-\alpha-1}\right] \frac{1}{t}\right) = O(1/t),$$

where the last equality comes from applying the same arguments as in Example 3.1 to the expectation term. □

■ □ ■

Lemma B.3. *Suppose $b^m(x)$ is defined as the optimal bid in a first price auction. If $q'(x) = \Theta(x^{\alpha-1})$ as $x \to 0$. then*

$$1/q \circ b^m(x) = O(x^{-\alpha}) \text{ as } x \to 0.$$

Proof. Recall that $b^m(x)$ can be defined as the solution to

$$q'(b)(x - b) = q(b). \tag{B.2}$$

Because $q'(x) = \Theta(x^{\alpha-1})$, by integrating one has that $q(x) = \Theta(x^\alpha)$. By using these two facts on (B.2) one can conclude that $x - b^m(x) = \Theta(b^m(x))$ and hence that $b^m(x) = \Theta(x)$. Therefore

$$\begin{aligned} q(b^m(x)) &= q'(b^m(x))(x - b^m(x)) \\ &= \Theta\left((b^m(x))^{\alpha-1}\right) \Theta(b^m(x)) = \Theta(x^\alpha). \end{aligned}$$

This implies that $1/q(b^m(x)) = O(x^{-\alpha})$, which completes the proof. □

Appendix C

Appendix to Chapter 4

Posterior distributions of model parameters

Section 4.2.1 describes how the distribution of the purchase probability θ is updated by observing downwards transitions of the chain. This is a particular case of the Multinomial-Dirichlet conjugate structure, which applies to the rest of the parameters of our model: ϵ,μ, and $\{\mu_i\}$. We next define the posterior updates for all the parameters:

$$\begin{aligned}
\theta \mid h, d &\sim Beta(1 + Purchases, 1 + NonPurchases) \\
\epsilon \mid h, d &\sim Beta(1 + ObservedPurchases, 1 + UnobservedPurchases) \\
\mu \mid h, d &\sim Beta(1 + Replenishes, 1 + NonReplenishes) \\
\{\mu_j\}_{j=\underline{\mu},\ldots,S} \mid h, d &\sim Dirichlet(\{1 + ReplenishLevel_j\}_{j=\underline{\mu},\ldots,S}),
\end{aligned}$$

where,

$$\begin{aligned}
Purchases &:= |\{t : h_t = h_{t-1} - 1\}| \\
NonPurchases &:= |\{t : h_t = h_{t-1} \wedge h_{t-1} > 0\}| \\
ObservedPurchases &:= |\{t : h_t = h_{t-1} - 1 \wedge d_{t-1} = 1\}| \\
UnobservedPurchases &:= |\{t : h_t = h_{t-1} - 1 \wedge d_{t-1} = 0\}| \\
Replenishes &:= |\{t : h_t > h_{t-1}\}| \\
NonReplenishes &:= |\{t : h_t <= h_{t-1} \wedge h_{t-1} < S\}| \\
ReplenishLevel_j &:= |\{t : h_t > h_{t-1} \wedge h_t = j \wedge h_{t-1} < \underline{\mu}\}|.
\end{aligned}$$

Note how in the definition of $NonPurchases$ we only consider in-stock periods, since purchases are not observable in the out-of-stock state. Similarly, $NonReplenishes$ only considers states below the maximum stock level. Finally, note in the definition of $ReplenishLevel_j$ how we only consider replenish events with a state below $\underline{\mu}$ (see section 4.2.1 in the paper for a discussion).