Manage Inventories with Learning on Demands and Buy-up Substitution Probability

Zhenwei Luo

Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong, zhen-wei.luo@polyu.edu.hk

Pengfei Guo*

College of Business, City University of Hong Kong, Kowloon, Hong Kong, penguo@cityu.edu.hk

Yulan Wang

Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong, yulan.wang@polyu.edu.hk

Abstract: ***Problem Definition:*** This paper considers a setting in which an airline company sells seats periodically, and each period consists of two selling phases, an early-bird-discount phase and a regular-price phase. In each period, when the early-bird-discount seat is stocked-out, an early-bird customer who comes for the discounted seat either purchases the regular-price seat as a substitute (called *buy-up substitution*) or simply leaves. ***Methodology/Results:*** The optimal inventory level of the discounted seats reserved for early-bird sale is a critical decision for the airline company to maximize its revenue. The airline company learns about the demands for both discounted and regular-price seats and the buy-up substitution probability from historical sales data, which, in turn, are affected by past inventory allocation decisions. In this paper, we investigate two information scenarios based on whether or not lost sales are observable, and we provide the corresponding Bayesian updating mechanism for learning about demand parameters and substitution probability. We then construct a dynamic programming model to derive the Bayesian optimal inventory level decisions in a multi-period setting. The literature finds that the unobservability of lost sales drives the inventory manager to stock more (i.e., the Bayesian optimal inventory level should be kept higher than the myopic inventory level) to observe and learn more about demand distributions. Here, we show that when the buy-up substitution probability is known, one may stock less, as one can infer some information about the primary demand for the discounted seat from the customer substitution behavior. We also find that to learn about the unknown buy-up substitution probability drives the inventory manager to stock less so as to induce more substitution trials. Finally, we develop a SoftMax algorithm to solve our dynamic programming problem. We show that the obtained stock more (less) result can be utilized to speed up the convergence of the algorithm to the optimal solution. ***Managerial Implications:*** Our results shed light on the airline seat protection level decision with learning about demand parameters and buy-up substitution probability. Compared with myopic optimization, Bayesian inventory decisions that consider the exploration–exploitation trade-off can avoid getting stuck in local optima and improve the profit. We also identify new driving forces behind the stock more (less) result that complement the Bayesian inventory

*Corresponding Author.

management literature.

Keywords: Airline Seat Allocation; Early Bird Discount; Bayesian Inventory Management; Newsvendor Model; SoftMax Algorithm

History: Received: November 2019; Accepted: October 2022 by Kamalini Ramdas, after 3 revisions.

---

1. Introduction

Airline companies often offer early-bird booking discounts to passengers. Customers who book tickets early can get a cheaper rate, and those who purchase near the departure date may have to pay a much higher price. This practice is based on market segmentation—customers have different price sensitivities and time sensitivities—and offering early-bird booking discounts can help to stimulate more demand to fill otherwise vacant seats. According to Shaw (1982), some customers, such as business travelers, are time-sensitive but price-insensitive. They generally purchase late because of tight schedules and are willing to pay a high fare. For convenience, we call these customers *regular customers*. Others, like leisure or vacation travelers, are time-insensitive but price-sensitive. They prefer to purchase early for a lower fare. We call these customers *early-bird customers*. To achieve market segmentation, an advanced purchase deadline is usually set, and the discount fare is available only before such a deadline. For example, Delta Air Lines states that "Most of our deeply discounted products require advanced purchases of 3, 7, 14 or 21 days" [1]. Japan Airlines provides a SAKITOKU ticket scheme, which offers four options for early-bird customers according to whether their purchase date is 75 days/ 55 days/ 45 days/ 28 days prior to departure; see Figure 1 for an example from the Japan Airlines website[2]. In this study, we only consider two different fare classes for simplicity. Such a two-fare setting is also commonly adopted in the airline revenue management literature; see, for example, Littlewood (1972), Brumelle et al. (1990), and Cooper et al. (2006). Under the two-fare setting, there is only one advanced purchase deadline, and we call the phase before the deadline the *early-bird-discount phase* and the remaining phase the *regular-price phase*. Hotle et al. (2015)

---

[1] For the details, please see `https://urldefense.com/v3/__https://www.delta.com/us/en/booking-information/fare-classes-and-tickets/ticket-rules-restrictions__;!!KjDnqvtInNPT!jmo8BFjDP9G9rdMHwFosBlAzXUB_j4eSNxTJAB1Z3HbK-DWRefl3YV4FbqkDHbIXe-LVF4KMqlJhByWuP9AR59AEUCka5Wo$`.

[2] For the details, please see `https://urldefense.com/v3/__https://www.jal.co.jp/jp/en/dom/waribiki/super_sakitoku.html__;!!KjDnqvtInNPT!jmo8BFjDP9G9rdMHwFosBlAzXUB_j4eSNxTJAB1Z3HbK-DWRefl3YV4FbqkDHbIXe-LVF4KMqlJhByWuP9AR59AEY7gys-o$`. The data in Figure 1 was retrieved from this website on Sep 17, 2022.

show that the advanced purchase deadline can be used to induce price-sensitive consumers to purchase in the early-bird-discount phase while leaving price-insensitive consumers in the market. We thus assume that early-bird customers arrive only in the early-bird-discount phase and that regular customers arrive only in the regular-price phase.

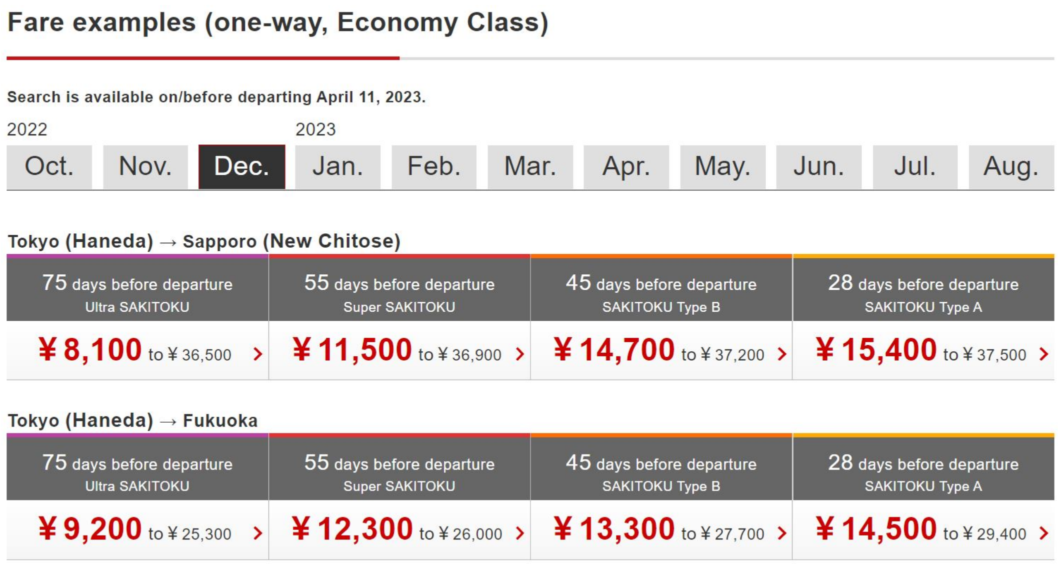


Figure 1: Illustration of Advanced Purchase Deadline: Japan Airlines

Because of fixed seat capacity, the allocation of seats among multiple fares becomes a critical decision when airline companies aim to maximize their expected revenues.[3] Such a decision is also called a booking limit decision or seat protection level decision for seats with different fares. Reserving too few seats for low-fare tickets may result in revenue loss by losing some price-sensitive passengers, while reserving too many can lead to fewer passengers purchasing high-fare seats.

To make the appropriate seat allocation decision, demand learning is essential. As Jack Bovey, a revenue optimization manager at British Airways, has pointed out, "it was the first step towards what will hopefully become an important part of how we forecast, and hence price, flights" (The Alan Turing Institute, 2020). When conducting demand forecasting for airline companies, one cannot ignore an important customer purchasing behavior: buy-up substitution, which is customers' substitution of regular-price seats for discounted seats when the latter are sold out. In practice, discounted seats often sell out very quickly (U.S. Department of Transportation, 2019). When customers come for a discounted seat but find that it is out of stock, they may choose to buy the regular-price seat as a substitute. Buy-up

[3]Another revenue-maximizing approach is dynamic pricing. However, according to Belobaba (1987), dynamic price change may lead to an irrational price war between competitors. Unlike price adjustment, seat inventory control aims to properly allocate seats among multiple fares, which is easy to manipulate and hidden from competitors, and it hence becomes a practically feasible strategy in revenue management.

substitution is an important issue in the airline seat allocation problem (Belobaba, 1987), given that buy-up substitution can lead to up to a 9% increase in airline revenue (Gallego et al., 2009) and improve airline seat management (Cooper and Li, 2012). Ja et al. (2001) show that the substitution rate ranges from 15% to 55%, and considering the substitution issue can improve the accuracy of demand estimation by $9\% - 20\%$ (Ratliff et al. 2008). Ignoring buy-up substitution could result in a severe *spiral-down effect* (Cooper et al., 2006): setting a low protection level for high-fare seats (or reserving too many low-fare tickets) results in a low estimation of the demand for high-fare seats, which causes an even lower protection level for the high-fare seats in the following periods. As the early-bird customers arrive only in the early-bird-discount phase, buy-up substitution happens only in this phase after the discounted seats sell out; see Figure 2 for an illustration.

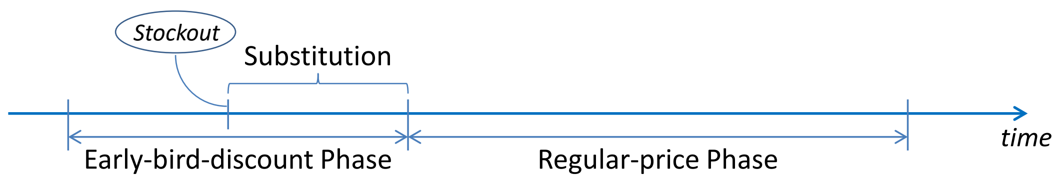


Figure 2: Illustration of Substitution Demands in the Early-bird-discount Phase

The seat allocation problem can be formulated as a newsvendor-type model to derive the optimal seat protection levels for seats at different fares; see, e.g., Littlewood (1972) and Belobaba (1987). However, the information about demand and buy-up substitution probability has to be learned from historical sales data, which are affected by past inventory allocation decisions. Thus, how to allocate seats among different fares dynamically over multiple time periods with learning about both demand distributions and buy-up substitution probability becomes an important decision problem for the airline manager. In this study, we investigate this problem and provide a solution.

For the sake of analytical tractability, we consider the following simplified but representative setting. An airline company has a fixed number of seats on a flight. Seats are sold in two phases, an early-bird-discount phase followed by a regular-price phase. Early-bird customers who prefer discounted seats arrive only in the early-bird-discount phase, while regular customers who prefer regular-price seats arrive only in the regular-price phase. If discounted seats stock out, unsatisfied early-bird customers either simply leave or purchase regular-price seats as a substitute. We first develop a baseline *single-period* inventory management problem with buy-up substitution. By utilizing this baseline model, we show that if the decision maker repeatedly makes the myopic optimization decision that maximizes only the current-period expected profit, the learning-and-optimization process can be stuck in local optima, leading to a severe profit loss.

Next, we construct a dynamic programming model to investigate the optimal multi-period inventory decisions by taking into account the associated exploration–exploitation tradeoff. Depending on whether or not lost sales are observable, we have two information scenarios, denoted as $\mathcal{O}$ and $\mathcal{U}$, respectively. When lost sales are unobservable ($\mathcal{U}$ scenario), the sales data provide only partial (censored) information on the actual demand. For example, if all of the early-bird tickets, say 50 discounted seats, are sold out, then the airline company can only know that the early-bird demand is at least 50. For each information scenario, we derive the corresponding Bayesian updating formula and the Bellman equation of our dynamic programming model to find the optimal inventory allocation decisions. However, directly solving the dynamic programming of Bayesian inventory management suffers from the curse of dimensionality. To facilitate insight generation and algorithm development, we thus focus on the comparison between the Bayesian optimal inventory level and the myopic inventory level.

First, we consider a setting in which the substitution probability is known but demand parameters need to be estimated, and we obtain the following results:

(1) In the complete observation $\mathcal{O}$ scenario, the inventory manager does not need to increase the inventory level to obtain more demand information. Thus, the Bayesian optimal inventory level is equal to the myopic inventory level.

(2) In the partial observation $\mathcal{U}$ scenario, lost sales are unobservable. When lost sales are not observed, there is a famous stock more result in the Bayesian inventory management literature—the Bayesian optimal inventory level should be set larger than the myopic inventory level to better learn about demand distribution—without considering substitution (Lariviere and Porteus, 1999) or by taking substitution into account but assuming that the substitute product is always available (Chen and Plambeck, 2008). Here, the buy-up substitution occurs when discounted seats stock out, but because of limited seat capacity, the complete observation of the substitute demand cannot be guaranteed. We show that under certain conditions, the unobservability of lost sales becomes a stock less driving force, in that the Bayesian optimal inventory level should be set lower than the myopic inventory level. The information about the primary demand can be obtained not only directly through sales data but also indirectly from the observations of buy-up substitution. Because the demand parameters can be inferred from the substitution behavior of unsatisfied customers, discounted seats can be stocked less to induce more substitution trials. Our numerical study reveals that such a stock less result can happen under more general settings.

Next, we consider the setting in which both the demand parameters and substitution probability need to be estimated. To better learn about the substitution probability requires stocking less discounted seats so that substitution can happen more frequently. Such a stock less driving force is similar to the one in Chen and Plambeck (2008). Nonetheless, to observe and learn more about the primary demand of early-bird customers may require stocking more discounted seats. Because of the interplay between multiple driving forces that may work in opposite directions, the result regarding whether to stock more or stock less is generally ambiguous. Our extensive numerical results suggest that the final comparison result depends on factors such as seat prices and the prior beliefs of the inventory manager.

As there are two types of discrete demands and buy-up substitution in our dynamic programming model, the heuristic algorithms developed in papers such as Chen (2010) cannot be simply applied here. Instead, we adopt the *SoftMax* algorithm (Goodfellow et al., 2016) to find the heuristic solution to our dynamic programming model. Through testing with simulated data sets, we find that the SoftMax algorithm performs very well and can converge to the true parameters without being stuck in local optima. Our numerical experiments indicate that compared with myopic optimization and the commonly used *Thompson sampling* algorithm, the SoftMax algorithm is the most efficient in balancing the exploration and exploitation of demand and buy-up substitution information and achieves the highest convergence rate. We further demonstrate that utilizing the obtained stock more (less) result can further speed up the convergence of the SoftMax algorithm to the optimal solution.

Our main contributions are threefold. First, our results can help airline companies to optimally determine the booking limit for discounted seats through Bayesian inventory management. Second, we identify conflicting driving forces behind the Bayesian optimal inventory level decision and show that the classic stock more result may not hold anymore, thus enriching the Bayesian inventory management literature. Third, we introduce a SoftMax algorithm to find the heuristic solution of our dynamic programming problem. We demonstrate that it outperforms the myopic optimization and the widely used Thompson sampling algorithm, and that the stock more (less) result can be utilized to speed up algorithm convergence.

The remainder of this paper is organized as follows. The related literature is reviewed in Section 2. In Section 3, we present a baseline single-period inventory model for optimizing the booking limit of discounted seats in an airline setting. In Section 4, we develop a multi-period inventory model with learning about demand parameters and the buy-up substitution probability in the context of Bayesian inventory management. We then compare the Bayesian optimal inventory level with the corresponding myopic one under various settings. The heuristic analysis is conducted in Section 5, and concluding remarks are provided in Section 6. All of the proofs are relegated to the Online Appendix.

2. Literature Review

Our study is closely related to the literature on Bayesian inventory management. In the early stage of this research stream, researchers mainly consider settings with observable lost sales. Scarf (1959) formulates a Bayesian inventory dynamic programming model with two state variables (inventory level and demand parameter). Scarf (1960) shows that the problem can be reduced to one state variable with a gamma demand distribution. Successive studies such as Azoury (1985) and Miller (1986) extend Scarf's method to other demand distributions. Lovejoy (1990) provides myopic policies by reducing the single state to zero-dimensional state space, i.e., a static optimization problem. Later researchers consider settings in which lost sales are unobservable. Such demand censoring introduces difficulty to demand estimation. According to Braden and Freimer (1991), only special types of distribution—the so-called newsvendor distributions—allow parsimonious information updating. By utilizing the newsvendor distribution and Scarf's method of state-space reduction, Lariviere and Porteus (1999) obtain analytical results for the optimal inventory decision of a multi-period newsvendor problem with unobservable lost sales, and identify a stock more result. Ding et al. (2002) further consider general demand distributions in newsvendor inventory models and show that the stock more result still holds when lost sales are unobservable. The proof of this conclusion is rectified by Lu et al. (2005) and further simplified by Bensoussan et al. (2009). Chen (2010) develops heuristics for a finite-horizon periodic-review inventory control problem with unobservable lost sales. In Chen (2010), there only exists one kind of demand, and simple inventory decisions associated with the observable lost sales scenario can be used to approximate the Bayesian optimal solution. In contrast, we consider two types of discrete demands together with buy-up substitution, and thus the heuristics in Chen (2010) cannot be applied. Instead, we introduce a randomized policy, *SoftMax* (Goodfellow et al., 2016), which can effectively utilize our stock more (less) result, as the heuristic solution of our problem. Jain et al. (2015) and Bensoussan and Guo (2015) utilize stockout times to estimate demand distribution for perishable and nonperishable products, respectively. Bensoussan et al. (2016) consider the incomplete inventory and demand information caused by invisible demand, such as spoilage, damage, pilferage, and returns. They study the inventory management problem with only sales information and develop an iterative algorithm to approximate the solution to the problem. fChen et al. (2017) investigate the allocation of limited inventory among multiple stores in the merchandise testing period, which aims to optimize the learning about the demand parameter prior to the main selling period. Chen and Wu (2019) consider a finite-horizon dynamic pricing problem with a fixed amount of inventory where the demand is price-dependent and needs to be learned via Bayesian updating. In addition to the aforementioned studies that consider Bayesian inventory management for the multi-period inventory control problem, we note that there are studies that consider profit maximization in a single-period setting with Bayesian learning and focus on issues such as shrinkage and pricing; see, e.g., Li and Ryan (2011), Harrison et al. (2012), and Li et al. (2021).

Among the studies of Bayesian inventory management, Chen and Plambeck (2008) is the first to consider the substitution issue when stockout occurs. Our paper also considers the substitution issue, but it differs greatly from Chen and Plambeck (2008). First, Chen and Plambeck (2008) assume that the substitute product is always available when the customers' desired product is sold out. However, in our setting, because of the fixed seat capacity, the substitute product—the regular-price seat—is limited and thus not always available. Second, the demand for the substitute product in Chen and Plambeck (2008) comes from stockout-based substitution only, whereas in our study, such a demand comes from two sources, buy-up substitution and the primary demand for the regular-price product.

Our work is also related to studies of inventory management with demand estimation from censored observations. Some of these studies utilize the expectation-maximization algorithm to estimate the demand and substitution parameters, including Anupindi et al. (1998), Kök and Fisher (2007), Ulu et al. (2012), Vulcano et al. (2012), and Chen and Chao (2019); some employ non-parametric demand learning, including Huh and Rusmevichientong (2009), Feng and Shanthikumar (2017), Chen and Chao (2020), and Yuan et al. (2021); and some develop operational statistics to integrate demand estimation and inventory optimization, including Liyanage and Shanthikumar (2005) and Chu et al. (2008).

Our study is related to studies of the seat allocation problem in the airline revenue management literature. According to McGill and van Ryzin (1999), the early-bird discount selling strategy was first adopted by airline companies such as BOAC (now British Airways) in the early 1970s. By using this strategy, airline companies can gain extra revenue from selling seats that would otherwise go empty without offering discounts. Littlewood (1972) provides an optimal rule for optimal seat inventory allocation from the perspective of benefit maximization. Belobaba (1987) further extends this rule to multiple fare classes by using the expected marginal seat revenue method. Pfeifer (1989) obtains a similar result with a different approach. Using marginal analysis as in Belobaba (1987), Brumelle et al. (1990) formally prove that a variant of Littlewood's rule could be optimal under a general model of the seat allocation problem. van Ryzin and McGill (2000) provide a simple adaptive approach to optimize seat protection levels. Cooper et al. (2006) show that simply following Littlewood's rule without considering buy-up substitution can cause a serious spiral-down effect, resulting in severe revenue loss. Cooper and Li (2012) further demonstrate the benefit of incorporating buy-up substitution into airline seat management. However, all of these studies either do not consider demand learning or consider it but ignore the impact of the inventory decision in the current period on the following periods (i.e., they do not consider the exploration–exploitation tradeoff in demand learning). Our study is the first to incorporate such a tradeoff in our solution to the airline seat allocation problem.

In our two-phase selling model, the first-phase price is lower than the second-phase price. We note that in business practice, there also exist markdown situations in which the first-phase price is higher than the second-phase price. Hu et al. (2015) study such a markdown inventory management problem. In their paper, there are also two selling phases in each period: a clearance phase (modeled as the first phase) with a markdown price and a regular-sales phase (modeled as the second phase) with a full price. Customers who do not get the product in the clearance phase can choose to buy it in the following regular-sales phase. The key difference between the markdown model in Hu et al. (2015) and our early-bird discount model is that the inventory used in their clearance phase is part of the unsold products from the previous period; that is, they are not newly produced, and the leftover products from the clearance phase cannot be sold in the following regular-sales phase. Thus, the selling periods in their markdown model are inter-correlated, whereas the selling periods in our early-bird discount model are independent. Another key difference is that Hu et al. (2015) consider a static model in which there is no learning about demand distribution and substitution parameters. In contrast, we consider learning about these parameters.

3. Single-Period Model

In this section, we first review a baseline single-period inventory management problem with two selling phases. The optimal inventory decision can be expressed in a similar way as a newsvendor-problem solution. Next, we consider a myopic optimization decision where the company repeatedly makes the current-period optimal inventory decision along the timeline, with demand parameters and substitution probability updated according to Bayes' rule. We show with a numerical example that the myopic optimization can get stuck in local optima.

3.1 Model Description

Previous studies on airline seat allocation problem mainly focus on a single period model; see, e.g., Brumelle et al. (1990). For completeness, we briefly review this model and state it with our notations. Consider a single selling period with two selling phases: an early-bird-discount phase and a regular-price phase. The corresponding selling prices are denoted as p_1 and p_2, respectively, with $p_1 < p_2$. The primary demand from early-bird customers for the discounted seat and the primary demand from regular customers for the regular-price seat are denoted by D_1 and D_2, respectively, which are discrete random variables. Let $f_i(\cdot|\theta_i)$ be the probability mass function of D_i, where θ_i is an unknown parameter with $\theta_i \in \Theta_i$ $(i = 1, 2)$.[4] Let M be the total number of available seats on the flight, which is a fixed number. The firm's objective is to determine the number of discounted seats, denoted as y, to maximize its total expected profit over the two selling phases.

There exists a tradeoff associated with the inventory decision y. When the firm allocates too few seats for early-bird-discount sales (i.e., y is very small), the primary demand for discounted seats may not be fully satisfied. Some unsatisfied early-bird customers may simply leave. Hence, the firm loses the opportunity to sell more. However, if the firm allocates too many seats for early-bird-discount sales (i.e., y is very large), the firm may lose the chance to force some unsatisfied early-bird customers to purchase regular-price seats as substitution.

As the primary demand for the discounted seat is D_1, the realized sales of the discounted seat can be expressed as $D_1 \wedge y$, where $a \wedge b = \min(a, b)$. If there are leftover discounted seats at the end of the early-bird-discount phase, they are sold in the regular-price phase as well. Thus, the amount of inventory available for the regular-price sales is $(M - y \wedge D_1)$. Note that the demand for regular-price seats comes from two sources: the substitution demand from unsatisfied early-bird customers who buy regular-price seats as a substitute, denoted by a random variable K, and the primary demand from regular customers (i.e., D_2). We assume that each unsatisfied customer's substitution decision is a Bernoulli trial with probability α, which is called the *buy-up substitution probability*. The random variable K then follows a binomial distribution with parameters $((D_1 - y)^+, \alpha)$, where $x^+ = \max(0, x)$. Under given values of θ_1, θ_2, and α, the firm makes the inventory-level decision to maximize its total expected profit $\pi(y|\theta_1, \theta_2, \alpha)$ as follows:

$$\max_y \pi(y|\theta_1, \theta_2, \alpha) = p_1 E[y \wedge D_1|\theta_1] + p_2 E[(K + D_2) \wedge (M - y \wedge D_1)|\theta_1, \theta_2, \alpha] \quad (1)$$
$$\text{s.t.} \quad 0 < y \leq M.$$

We obtain the following result regarding $\pi(y|\theta_1, \theta_2, \alpha)$.

Proposition 1 *The profit function $\pi(y|\theta_1, \theta_2, \alpha)$ is unimodal in y.*

According to Brumelle et al. (1990), the optimal inventory level of the discounted seat y^* can be expressed as

$$y^* = \max\left\{0 < y \leq M : Pr(K + D_2 > M - y|D_1 \geq y, \theta_1, \theta_2, \alpha) < \frac{p_1 - \alpha p_2}{(1 - \alpha)p_2}\right\}. \quad (2)$$

[4] In this study, we assume that D_1 and D_2 are independent given θ_1 and θ_2. However, θ_1 and θ_2 themselves along with the buy-up substitution probability can be correlated in the inventory manager's uncertain beliefs. Such a model setup is consistent with the one in Brumelle et al. (1990).

Note that when there are multiple optimal solutions, the above equation (2) yields the one with the smallest value. It is worth pointing out that if the substitution probability α is 0 (i.e., $K = 0$ with probability 1), the optimal inventory level of the discounted seat y^* is equal to the one under Littlewood's rule (Littlewood 1972).

We then conduct a sensitivity analysis of the optimal inventory level y^* with respect to the substitution probability α and obtain the following:

Proposition 2 *The profit function $\pi(y|\theta_1, \theta_2, \alpha)$ is submodular in (y, α); that is, $\partial[\pi(y+1|\theta_1, \theta_2, \alpha) - \pi(y|\theta_1, \theta_2, \alpha)]/\partial\alpha < 0$. Hence, y^*, the optimal inventory level of discounted seats, decreases with the substitution probability α.*

Proposition 2 shows that the inventory manager should set a lower inventory level for the discounted seat as the substitution probability increases. The underlying reason is that with a larger substitution probability, it becomes more likely that unsatisfied early-bird customers will buy a regular-price seat.

3.2 Myopic Optimization

The above single-period model assumes that the demand parameters and substitution probability are known. In practice, such information is often unknown to the inventory manager. When decisions are made repeatedly over multiple periods, unknown parameters can be learned from past sales. The myopic optimization does not consider the future effects of the current-period decision and maximizes only the expected profit in the current period, along with learning about demand parameters and substitution probability according to Bayes' rule. We now provide a numerical example to show the drawback of the myopic optimization: it can get stuck in a local optimum.

Example 1 ***(Local Optimality of Myopic Optimization)*** *Consider an airline company that offers an early-bird discount sale of a flight with a large-sized jet. The total number of seats is $M = 220$. The regular price is set at $p_2 = 1200$, and the early-bird-discount price is $p_1 = 650$. Both the primary demands for the discounted and regular-price seats, D_1 and D_2, follow two-point distributions, which are known to the inventory manager. Specifically, D_1 takes the value 30 or 100 with equal probability 0.5, and D_2 takes the value 60 or 120 also with equal probability 0.5. The substitution probability α is unknown and takes either a low value 0.2 or a high value 0.8. The inventory manager holds a prior belief that $(Pr(\alpha = 0.2), Pr(\alpha = 0.8)) = (0.5, 0.5)$. Suppose that the true substitution probability is $\alpha = 0.8$.*

Now, suppose that the inventory manager adopts the myopic optimization for the seat allocation problem that only maximizes the current-period expected profit. Then, based on *the single-period optimal inventory decision stated in (2), the optimal inventory level in period 1 can be calculated. In the following period 2, the inventory manager first updates the belief about the substitution probability based on the observed sales data according to Bayes' rule and then makes the optimal inventory allocation decision according to (2) again. Such a procedure repeats itself in the remaining periods. Denote the optimal inventory level decision in period i* $(i \geq 1)$ *as* y_i^m.

In the first period, the optimal inventory decision for the discounted seats is $y_1^m = 100$. *In this case, all primary demands for the discounted seats are satisfied. As such, no substitution happens. The inventory manager's belief remains the same as in the prior period. Hence, the decision for the following periods is stuck at* $y_i^m = 100$, $i = 2, \ldots$, *and the corresponding expected profit per period is* $150,250$. *However, based on the true substitution probability* $\alpha = 0.8$, *the optimal inventory level is* $y_i^m = 1$ $(i \geq 1)$ *and the corresponding expected profit per period is* $170,090$, *a* 13.2% *increase over the one under the myopic optimization.*

The foregoing example reveals a major drawback of the myopic optimization: learning about unknown system parameters can be stuck in a local optimum and thus cannot progress at all. In the following section, we present a dynamic programming model that considers not only the current-period profit but also the efficiency of learning about demand parameters and substitution probability.

4. Multi-Period Model

We now consider a multi-period seat allocation problem with learning about demand distribution parameters (i.e., θ_1 and θ_2) and substitution probability α. The firm's objective is to maximize the total discounted expected profit over N periods, where the discount factor is denoted by δ $(0 < \delta \leq 1)$. The setting in each period remains the same as that in the single-period model; see §3.

Below, we first introduce two information scenarios and derive the Bayesian learning formula for unknown parameters in each scenario. We then specifically consider two settings. In the first setting, the inventory manager cares most about demand distributions and the substitution probability α can be known from the prior knowledge and experience. Thus, only the demand distribution parameters need to be learned. In the second setting, both the demand distribution parameters and substitution probability need to be learned. For each setting, we formulate the corresponding dynamic programming model for the optimal inventory decisions. Recall that the decision variable is the inventory level of the discounted seat y. We then conduct a comparison between the Bayesian optimal inventory level and the myopic inventory level.

4.1 Bayesian Updating under Two Information Scenarios

In our study, we consider two information scenarios based on whether or not lost sales are observable. When lost sales are observable (denoted as the $\mathcal{O}$ scenario), we have the complete observations: the realized primary demand for the discounted seat x_1, the substitution demand from unsatisfied early-bird customers for the regular-price seat x_{21}, and the realized primary demand for the regular-price seat x_{22} are all observable. Such complete observations are feasible in the current e-commerce and big-data era, in which a firm can easily track customers' purchase behavior.

Given the discounted seat inventory level y, the demand parameter θ_i $(i = 1, 2)$, and the substitution probability α, the likelihood of observing demand realizations x_1, x_{21}, and x_{22} can be written as

$$f^y_{\mathcal{O}}(x_1, x_{21}, x_{22}|\theta_1, \theta_2, \alpha) = \begin{cases} f_1(x_1|\theta_1)f_2(x_{22}|\theta_2)\binom{x_1-y}{x_{21}}\alpha^{x_{21}}(1-\alpha)^{x_1-y-x_{21}}, & \text{if } x_1 > y; \\ f_1(x_1|\theta_1)f_2(x_{22}|\theta_2), & \text{if } x_1 \leq y. \end{cases} \tag{3}$$

When lost sales are unobservable (denoted as the $\mathcal{U}$ scenario), sales of the discounted seat s_1, sales of the regular-price seat from the substitution demand s_{21}, and sales of the regular-price seat from the primary demand s_{22} are all observable. They are censored sales data of demand realizations x_1, x_{21}, and x_{22}, respectively. If $s_1 < y$, no buy-up substitution occurs and the sales of the regular-price seat are all from the primary demand. Otherwise, some unsatisfied early-bird customers may choose to buy the regular-price seat as a substitute. For example, if $s_1 = y$, $s_{21} < M - y$, and $s_{22} = M - y - s_{21}$, then only s_{21} unsatisfied early-bird customers choose to buy the regular-price seat as a substitute. In such a case, sales observations of the primary demands for both the discounted and regular-price seats are censored. Thus, one needs to sum up all likelihoods over all possible values of D_1 and D_2 that satisfy $D_1 \geq y + s_{21}$ and $D_2 \geq M - y - s_{21}$. Consequently, the resulting likelihood is

$$\left[\sum_{i=y+s_{21}}^{+\infty} f_1(i|\theta_1)\binom{i-y}{s_{21}}\alpha^{s_{21}}(1-\alpha)^{i-y-s_{21}}\right] \cdot \left[\sum_{j=M-y-s_{21}}^{+\infty} f_2(j|\theta_2)\right].$$

Similarly, we can analyze other cases. In summary, under the $\mathcal{U}$ scenario, the likelihood of observing sales quantities (s_1, s_{21}, s_{22}) is

$$f^y_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1, \theta_2, \alpha)$$

$$
= \begin{cases}
f_1(s_1|\theta_1) f_2(s_{22}|\theta_2), & \text{if } s_1 < y,\ s_{21} = 0 \text{ and } s_{22} < M - s_1; \\
f_1(s_1|\theta_1) \left[\sum\limits_{j=M-s_1}^{+\infty} f_2(j|\theta_2) \right], & \text{if } s_1 < y,\ s_{21} = 0 \text{ and } s_{22} = M - s_1; \\
\left[\sum\limits_{i=y+s_{21}}^{+\infty} f_1(i|\theta_1) \binom{i-y}{s_{21}} \alpha^{s_{21}} (1-\alpha)^{i-y-s_{21}} \right] f_2(s_{22}|\theta_2), & \\
 & \text{if } s_1 = y,\ s_{21} < M - y \text{ and } s_{22} < M - y - s_{21}; \\
\left[\sum\limits_{i=y+s_{21}}^{+\infty} f_1(i|\theta_1) \binom{i-y}{s_{21}} \alpha^{s_{21}} (1-\alpha)^{i-y-s_{21}} \right] \cdot \left[\sum\limits_{j=M-y-s_{21}}^{+\infty} f_2(j|\theta_2) \right], & \\
 & \text{if } s_1 = y,\ s_{21} < M - y \text{ and } s_{22} = M - y - s_{21}; \\
\sum\limits_{i=M}^{+\infty} \sum\limits_{k=M-y}^{i-y} f_1(i|\theta_1) \binom{i-y}{k} \alpha^k (1-\alpha)^{i-y-k}, & \\
 & \text{if } s_1 = y,\ s_{21} = M - y \text{ and } s_{22} = 0.
\end{cases} \tag{4}
$$

Let I^y_{scen} denote the information set that contains all of the available information for a given inventory level y under the information scenario *scen*, where $scen \in \{\mathcal{O}, \mathcal{U}\}$. Specifically, under the $\mathcal{O}$ scenario, $I^y_{\mathcal{O}} = \{(x_1, x_{21}, x_{22}) : 0 \le x_{21} \le (x_1 - y)^+, x_1, x_{21}, x_{22} \in N_+\}$, where N_+ is the set of all nonnegative integers; that is, the information set contains all possibilities of both the realized primary demands in two phases and the substitution demand. Similarly, under the $\mathcal{U}$ scenario, $I^y_{\mathcal{U}} = \{(s_1, s_{21}, s_{22}) : 0 \le s_1 \le y, 0 \le s_{21} \le (M - y) \cdot \mathcal{I}_{\{s_1 = y\}}, 0 \le s_{22} \le M - s_1 - s_{21}, s_1, s_{21}, s_{22} \in N_+\}$, where $\mathcal{I}_{\{.\}}$ is the indicator function; that is, the information set contains all possibilities of observed sales quantities.

Denote the joint prior distribution of θ_1, θ_2 and α in period i ($i = 1, 2, \ldots, N$) as $\phi_i(\theta_1, \theta_2, \alpha)$. Given $\phi_i(\theta_1, \theta_2, \alpha)$, the posterior distribution $\phi_{i+1}(\theta_1, \theta_2, \alpha)$ derived based on the information observed in period i serves as the prior in the following period $i + 1$. Under each information scenario $scen \in \{\mathcal{O}, \mathcal{U}\}$, given the observation in period i, $\xi \in I^y_{scen}$, the posterior distribution $\phi_{i+1}(\theta_1, \theta_2, \alpha)$ can be derived by using the corresponding likelihood function according to Bayes' rule, as follows:

$$
\phi_{i+1}(\theta_1, \theta_2, \alpha | \xi, y, \phi_i) = \frac{f^y_{scen}(\xi | \theta_1, \theta_2, \alpha) \phi_i(\theta_1, \theta_2, \alpha)}{\int_0^1 \int_{\Theta_1} \int_{\Theta_2} f^y_{scen}(\xi | \theta_1', \theta_2', \alpha') \phi_i(\theta_1', \theta_2', \alpha') d\theta_2' d\theta_1' d\alpha'}. \tag{5}
$$

Let $v_i^{scen}(\phi_i)$ denote the firm's maximum total discounted expected profit from period i to N under the information scenario *scen* when the prior distribution in period i is ϕ_i, where $i = 1, 2, \ldots, N$ and $scen \in \{\mathcal{O}, \mathcal{U}\}$. Then, for $i = 1, \cdots, N - 1$, we can write the Bayesian dynamic optimality equations as

$$
v_i^{scen}(\phi_i) = \max_{0 < y \le M} E_{\phi_i(\theta_1, \theta_2, \alpha)} \left\{ \pi(y | \theta_1, \theta_2, \alpha) + \delta \sum_{\xi \in I^y_{scen}} v_{i+1}^{scen}(\phi_{i+1}) f^y_{scen}(\xi | \theta_1, \theta_2, \alpha) \right\}, \tag{6}
$$

and for $i = N$, we have

$$v_N^{scen}(\phi_N) = \max_{0<y\leq M} E_{\phi_N(\theta_1,\theta_2,\alpha)} \left\{\pi(y|\theta_1,\theta_2,\alpha)\right\}. \quad (7)$$

For ease of exposition, we use $G_i^{scen}(y,\phi_i)$ to denote the corresponding objective function of $v_i^{scen}(\phi_i)$ $(i = 1,\cdots,N)$, which means that for $i = 1,\cdots,N-1$,

$$G_i^{scen}(y,\phi_i) = E_{\phi_i(\theta_1,\theta_2,\alpha)} \left\{\pi(y|\theta_1,\theta_2,\alpha) + \delta \sum_{\xi\in I_{scen}^y} v_{i+1}^{scen}(\phi_{i+1}) f_{scen}^y(\xi|\theta_1,\theta_2,\alpha)\right\}, \quad (8)$$

and for $i = N$,

$$G_N^{scen}(y,\phi_N) = E_{\phi_N(\theta_1,\theta_2,\alpha)} \left\{\pi(y|\theta_1,\theta_2,\alpha)\right\}. \quad (9)$$

The myopic inventory level in period i $(i = 1,\cdots,N)$ maximizes only that period's expected profit and is denoted as y_i^m. Hence, it is the optimal solution of the corresponding single-period model with a prior belief $\phi_i(\theta_1,\theta_2,\alpha)$. For ease of exposition, we use $G_i^m(y,\phi_i)$ and $v_i^m(\phi_i)$ to denote the firm's objective function and the corresponding optimal value function in period i under the myopic setting, respectively.

4.2 Only Demand Parameters Unknown

In this subsection, we consider the case in which the substitution probability α is known but the demand parameters θ_1 and θ_2 are unknown and need to be estimated. Such a setting allows us to have a better understanding of the driving forces that lead to a better estimation of unknown demand parameters. We are particularly interested in whether the Bayesian optimal inventory level should be kept larger than the myopic inventory level to obtain better parameter estimates. Because α is known, $\phi_i(\theta_1,\theta_2,\alpha)$, the prior joint distribution in period i $(i = 1,\cdots,N)$ reduces to a two-variable distribution. Let $\phi_i'(\theta_1,\theta_2)$ denote the marginal prior distribution of θ_1 and θ_2 in period i, where $\phi_i'(\theta_1,\theta_2) = \frac{\phi_i(\theta_1,\theta_2,\alpha)}{\int_{\Theta_1}\int_{\Theta_2}\phi_i(\theta_1',\theta_2',\alpha)d\theta_2' d\theta_1'}$.

Below, we first consider the $\mathcal{O}$ scenario in which we have complete observations. For period i $(i = 1,\cdots,N)$, the impact of increasing the inventory level of the discounted seat y by one unit satisfies

$$\begin{aligned} G_i^{\mathcal{O}}(y+1,\phi_i') - G_i^{\mathcal{O}}(y,\phi_i') &= E_{\phi_i'}\pi(y+1|\theta_1,\theta_2,\alpha) - E_{\phi_i'}\pi(y|\theta_1,\theta_2,\alpha) \\ &= G_i^m(y+1,\phi_i') - G_i^m(y,\phi_i'). \end{aligned} \quad (10)$$

Equation (10) implies that the marginal impact of increasing the inventory level y on the objective function under the Bayesian inventory decision remains the same as that under the myopic decision. It then follows that the Bayesian optimal inventory level is equal to the myopic inventory level, which is formally stated in the following proposition.

Proposition 3 *When lost sales are observable and the substitution probability α is known, for any period i ($i = 1, \cdots, N$), given the same prior distribution $\phi'_i(\theta_1, \theta_2)$, the Bayesian optimal inventory level is equal to the myopic inventory level; that is, $y_i^{\mathcal{O}} = y_i^m$.*

The underlying reason is that when we have complete observations, there is no need to manipulate the inventory level to observe more demand information. Hence, the decision maker only needs to maximize the current-period expected profit. Such an equality between $y_i^{\mathcal{O}}$ and y_i^m serves as a benchmark for the following comparisons in other scenarios.

Next, we study the $\mathcal{U}$ scenario in which lost sales are unobservable. Studies that consider only one type of primary demand (e.g., Lariviere and Porteus 1999, Ding et al. 2002) show that the inventory manager will stock more to learn about the demand distribution; that is, the Bayesian optimal inventory level is larger than the myopic inventory level. Here, we have two types of primary demands, one for discounted seats and the other for regular-price seats. A change in the inventory level y affects the observations of both primary demands D_1 and D_2. Does the unobservability of lost sales still drive the inventory manager to stock more? To separate out the driving forces caused by learning about each type of primary demand, we consider the case in which only the parameter of one type of primary demands needs to be estimated while that of the other type is known. Denote the marginal distribution of θ_1 (θ_2) given the value of θ_2 (θ_1) as $\phi'_{i,1}(\theta_1) := \frac{\phi'_i(\theta_1,\theta_2)}{\int_{\Theta_1} \phi'_i(\theta'_1,\theta_2)d\theta'_1}$ ($\phi'_{i,2}(\theta_2) := \frac{\phi'_i(\theta_1,\theta_2)}{\int_{\Theta_2} \phi'_i(\theta_1,\theta'_2)d\theta'_2}$). The following proposition shows that the stock more result holds.

Proposition 4 *Consider that lost sales are unobservable and the substitution probability $\alpha = 0$. When the demand parameter θ_1 (θ_2) is unknown but θ_2 (θ_1) is known, for any period i ($i = 1, \cdots, N$), given the same prior distribution $\phi'_{i,1}(\theta_1)$ ($\phi'_{i,2}(\theta_2)$), the Bayesian optimal inventory level is no less (no larger) than the corresponding myopic inventory level; that is, $y_i^{\mathcal{U}} \geq y_i^m$ ($y_i^{\mathcal{U}} \leq y_i^m$).*

Proposition 4 indicates that the unobservability of lost sales is still a driving force for the inventory manager to stock more considering either type of primary demand. Intuitively, when only one type of primary demand needs to be learned while the other type is known, we should stock more seats for the unknown demand to observe and learn more about its distribution. In this specific sense, the well-known stock more result of other studies that only consider one type of primary demand is generalized to the case of two types of primary demand.

The result in Proposition 4 is obtained through assuming no buy-up substitution ($\alpha = 0$). Will the unobservable lost sales still drive the inventory manager to stock more when there is buy-up substitution ($\alpha > 0$)? The answer is no. We show that under certain circumstances, the inventory manager will stock less, in that the Bayesian optimal inventory level is lower than the myopic inventory level, as illustrated below.

Proposition 5 *Consider that both the demand parameter of D_2 and substitution probability $\alpha(>0)$ are known but the demand parameter of D_1, θ_1, is unknown. The total number of available seats $M = 2$, and the value of D_1 cannot be 1, i.e., $f_1(1|\theta_1) = 0$ for all $\theta_1 \in \Theta_1$. Then, for any period i ($i = 1, \cdots, N$), given the same prior distribution $\phi'_{i,1}(\theta_1)$, the Bayesian optimal inventory level is no larger than the corresponding myopic inventory level; that is, $y_i^{\mathcal{U}} \leq y_i^m$.*

In Proposition 5, no matter whether the inventory level y is set as 1 or 2, given the same realized primary demand for the discounted seat, the observed early-bird sales convey the same demand information. For example, if the realized D_1 is 0, the early-bird sales are 0 in both the $y = 1$ and $y = 2$ cases. If the realized D_1 is 2 or larger, then the early-bird sales $s_1 = 1$ in the case $y = 1$ and $s_1 = 2$ in the case $y = 2$. Both $s_1 = 1$ and $s_1 = 2$ convey the same demand information that $D_1 \geq 2$ because the demand cannot be 1. However, the inventory manager can infer some information about the primary demand D_1 from the customer substitution behavior when $y = 1$. In other words, stocking fewer discounted seats can induce more observations and learning about early-bird demand. Will such a stock less phenomenon still occur under more general demand distributions with a large seat capacity M? We now provide an example to show that the stock less result still occurs under more general settings.

Example 2 ***(Stocking Less Caused by Unobservability of Lost Sales)*** *Consider the case in which the early-bird-discount price $p_1 = 700$ and the regular price $p_2 = 1200$. The primary demand for the discounted seat D_1 follows a truncated Poisson distribution ($0 \leq D_1 \leq 300$) with an unknown parameter θ_1, where θ_1 takes the value 160 or 270. The primary demand for the regular-price seat D_2 also follows a truncated Poisson distribution ($0 \leq D_2 \leq 100$) with parameter $\theta_2 = 5$. The buy-up substitution probability α is known and $\alpha = 0.2$. At the beginning of the first period, the inventory manager holds a prior belief that θ_1 equals 160 with probability 0.8 and equals 270 with probability 0.2. The manager aims to determine the optimal number of discounted seats for the first period to maximize the total expected profit over two periods (with a discount factor $\delta = 1$). Denote the optimal inventory levels of the discounted seat in the first period under the Bayesian inventory management and the myopic optimization as $y_1^{\mathcal{U}}$ and y_1^m, respectively.*

We then vary M, the total number of seats, between 80 and 200 with a step length of 20. Table 1 lists the corresponding optimal inventory levels $y_1^{\mathcal{U}}$ and y_1^m. It shows that the stock

Table 1: Optimal Inventory Levels ($y_1^{\mathcal{U}}$ and y_1^m) under Different Values of M

M	80	90	100	110	120	130	140	150	160	170	180	190
$y_1^{\mathcal{U}}$	48	62	74	87	99	113	126	138	150	162	166	169
y_1^m	52	65	77	90	102	115	128	140	151	162	166	169
$y_1^{\mathcal{U}} - y_1^m$	-4	-3	-3	-3	-3	-2	-2	-2	-1	0	0	0

less result occurs when M is no greater than 160. Table 1 also shows that the difference between the two optimal inventory levels (i.e., $y_1^{\mathcal{U}} - y_1^m$) decreases as M increases. This implies that an increase in the total number of seats weakens the manager's incentive to stock less in this example. Intuitively, as the total number of seats M increases, the manager shall allocate more seats for early-bird-discount sale, i.e., the myopic optimal inventory level y_1^m becomes larger[5]. This results in more observations of the primary demand D_1 and thus reduces the manager's exploration incentive. In this example, when M becomes sufficiently large ($M \geq 170$), the difference between the two optimal inventory levels becomes zero.

The above example shows that the stock less result can happen under general settings when lost sales are unobservable and only the distribution of the primary demand for the discounted seat is unknown and needs to be learned. It is worth pointing out that besides the observed sales of the discounted seat s_1, the observed substitution demand s_{21} also conveys some information about the primary demand for the discounted seat D_1, which can drive the inventory manager to stock less. This is in sharp contrast to the stock more conclusion drawn in the study of Chen and Plambeck (2008), which assumes that the substitute product is always available. In Chen and Plambeck (2008), the substitution demand can be fully observed and only the lost sales are unobservable, which pushes the inventory manager to stock more to learn about the demand distribution. Here, the substitution demand cannot be fully observed as the substitute product, the regular-price seat, is not always available because of fixed seat capacity. Stocking fewer discounted seats provides more observations and learning about the substitution demand. Hence, the stock more result may no longer hold. The above results imply that the unobservable lost sales contain counter driving forces, which make the relationship between the Bayesian optimal inventory level $y_i^{\mathcal{U}}$ and the myopic inventory level y_i^m generally uncertain.

[5] This can be easily verified by checking the expression of $\pi(y+1|\theta_1, \theta_2, \alpha) - \pi(y|\theta_1, \theta_2, \alpha)$ presented in the proof of Proposition 1 in the online Appendix.

4.3 Demand and Substitution Parameters All Unknown

In this subsection, we consider the case in which the demand parameters θ_i $(i = 1, 2)$ and substitution probability α are all unknown. Let us first consider the $\mathcal{O}$ scenario in which lost sales are observable. In this scenario, one can have complete observations of realized demands to estimate demand parameters. However, given the need to estimate the substitution probability, inventory levels should be set such that we can obtain more observations about customer substitution behavior. In the $\mathcal{O}$ scenario, the marginal impact of increasing the inventory level y by one unit under Bayesian inventory management and that under myopic optimization have the following relationship, where the proof of inequality (11) can be found in online Appendix:

$$G_i^{\mathcal{O}}(y+1, \phi_i) - G_i^{\mathcal{O}}(y, \phi_i) \leq G_i^m(y+1, \phi_i) - G_i^m(y, \phi_i), \tag{11}$$

Based on (11), we obtain the following result.

Proposition 6 *When lost sales are observable and both the demand parameters and substitution probability are unknown, for any period i $(i = 1, \cdots, N)$, given the same prior $\phi_i(\theta_1, \theta_2, \alpha)$, the Bayesian optimal inventory level is no larger than the corresponding myopic inventory level; that is, $y_i^{\mathcal{O}} \leq y_i^m$.*

In comparison with the result stated in Proposition 3 where the substitution probability is known, Proposition 6 reveals that learning about the unknown substitution probability drives the inventory manager to stock less. Note that when lost sales are observable, reducing y, the inventory level of the discounted seat, can induce more observations of substitution demand while having no impact on the observations of the primary demands. Such a stock less result is similar to the one obtained in Chen and Plambeck (2008). In this sense, our study generalizes the stock less result from their setting of an unlimited capacity of the substitute product with a single source of demand to a setting of a limited capacity of the substitute product with two sources of demand.

Next, we consider the $\mathcal{U}$ scenario where lost sales are unobservable. Here, the driving forces identified in §4.2 still play their roles. In addition, one may need to stock less to observe and learn more about the substitution probability. The overall effect of those driving forces on the optimal inventory decision becomes unclear. The relationship between the Bayesian optimal inventory level and the myopic inventory level now depends on multiple factors, including seat prices and prior beliefs of the inventory manager, as shown in the following subsection.

4.4 A Numerical Study

We now examine how the system parameters affect the occurrence of the stock more (less) result when lost sales are unobservable. For the sake of comparison, we consider a two-period setting, under which we can conveniently solve the dynamic programming model to derive the Bayesian optimal inventory levels. Such a setting also allows us to capture the key exploration–exploitation tradeoff.

Consider the case in which an airline company offers an early-bird discount sale of a flight with a medium-sized jet. The total number of seats $M = 120$ and the regular price $p_2 = 1200$. The inventory manager aims to determine the optimal number of discounted seats for the first period to maximize the total expected profit over the two periods (with a discount factor $\delta = 1$). The primary demands for the discounted and regular-price seats, D_1 and D_2, follow the truncated Poisson distributions ($0 \le D_1 \le 300$, $0 \le D_2 \le 100$) with parameters λ_1 and λ_2, respectively. The two parameters λ_1 and λ_2 are correlated, and their relationship is indicated by a parameter θ, which takes the value 1, 2, or 3. When $\theta = 1$, $\lambda_1 = 160$ and $\lambda_2 = 5$; when $\theta = 2$, $\lambda_1 = 160$ and $\lambda_2 = 20$; and when $\theta = 3$, $\lambda_1 = 270$ and $\lambda_2 = 5$. The buy-up substitution probability α takes either a low value 0.2 or a high value 0.7. At the beginning of the first period, the inventory manager holds the prior beliefs $\tilde{u} = (Pr(\theta = 1), Pr(\theta = 2), Pr(\theta = 3))$ and $\tilde{w} = (Pr(\alpha = 0.2), Pr(\alpha = 0.7))$. Denote the optimal inventory levels of the discounted seat in the first period under Bayesian inventory management and myopic optimization as $y_1^{\mathcal{U}}$ and y_1^m, respectively.

We first fix the early-bird-discount price at $p_1 = 700$ to investigate how varying prior beliefs $\tilde{u}$ and $\tilde{w}$ affect the occurrence of the stock more (less) result.

(i) $\tilde{u} = (u_1, 0, 1 - u_1)$ ($0 \le u_1 \le 1$) and $\tilde{w} = (1, 0)$ under which the distribution of D_1 is unknown.

In this case, λ_2, the parameter of the primary demand for the regular-price seat D_2, is known ($\lambda_2 = 5$). The buy-up substitution probability $\alpha = 0.2$. λ_1, the parameter of the primary demand for the discounted seat D_1, however, is unknown and needs to be estimated. We then vary the prior belief $\tilde{u}$ by changing the value of u_1. Note that a prior belief with a larger u_1 leads to a lower expectation of D_1. Table 2(a) shows that the optimal inventory levels under both Bayesian inventory management and myopic optimization increase as u_1 increases. When $u_1 = 0$, the inventory manager does not need to learn about the demand distribution, as λ_1 is also known ($\lambda_1 = 270$). In such a situation, the Bayesian optimal inventory level $y_1^{\mathcal{U}}$ is equal to the myopic inventory level y_1^m, as shown in Table 2(a). As u_1 increases from 0 to 0.1, the parameter of D_1 becomes unknown, and the inventory manager needs to increase the inventory level of the discounted seat to better learn about D_1. Hence, a stock more result occurs. However, when u_1 takes larger values from 0.2 to 0.5, the myopic

Table 2: Impact of System Parameters on Optimal Inventory Levels ($y_1^{\mathcal{U}}$ and y_1^m)

(a) Unknown D_1: $\tilde{u} = (u_1, 0, 1 - u_1)$ and $\tilde{w} = (1, 0)$											
u_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$y_1^{\mathcal{U}}$	77	78	79	80	84	90	97	99	99	101	104
y_1^m	77	77	79	80	84	90	98	100	102	103	104
$y_1^{\mathcal{U}} - y_1^m$	0	+1	0	0	0	0	−1	−1	−3	−2	0

(b) Unknown D_2: $\tilde{u} = (u_2, 1 - u_2, 0)$ and $\tilde{w} = (1, 0)$											
u_2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$y_1^{\mathcal{U}}$	84	86	87	88	90	93	95	98	100	102	104
y_1^m	84	86	87	89	91	95	98	100	102	103	104
$y_1^{\mathcal{U}} - y_1^m$	0	0	0	−1	−1	−2	−3	−2	−2	−1	0

(c) Unknown α: $\tilde{u} = (1, 0, 0)$ and $\tilde{w} = (w, 1 - w)$											
w	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$y_1^{\mathcal{U}}$	1	1	1	1	8	52	95	96	97	97	104
y_1^m	1	1	1	1	8	52	98	100	102	103	104
$y_1^{\mathcal{U}} - y_1^m$	0	0	0	0	0	0	−3	−4	−5	−6	0

(d) Varying p_1: $\tilde{u} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\tilde{w} = (\frac{1}{2}, \frac{1}{2})$													
p_1	600	650	700	750	800	850	900	950	1000	1050	1100	1150	1200
$y_1^{\mathcal{U}}$	1	1	25	60	65	71	72	72	73	74	98	102	120
y_1^m	1	1	29	69	75	79	83	86	91	97	103	107	120
$y_1^{\mathcal{U}} - y_1^m$	0	0	−4	−9	−10	−8	−11	−14	−18	−23	−5	−5	0

inventory level y_1^m becomes larger such that the benefit brought by stocking more is less than the loss caused by it, and thus the inventory manager has no incentive to stock more. At the same time, because the myopic inventory level is not that large, the inventory manager has no incentive to stock less as well. Consequently, the inventory manager does not manipulate the inventory level (i.e., $y_1^{\mathcal{U}} = y_1^m$). When u_1 further increases to 0.6, the myopic inventory level y_1^m becomes large enough. In this situation, reducing the inventory level of the discounted seat can induce more substitution trials, from which the inventory manager can infer more demand information, leading to a stock less result. Such a stock less result remains when u_1 takes the larger values from 0.7 to 0.9. The difference between the two inventory levels $y_1^{\mathcal{U}}$ and y_1^m reaches the maximum at $u_1 = 0.8$. However, as u_1 continues to increase to 0.9, the uncertainty on D_1 weakens, and so does the incentive to stock less, shortening the gap between $y_1^{\mathcal{U}}$ and y_1^m. Lastly, when $u_1 = 1$, under which λ_1 is known ($\lambda_1 = 160$), the two inventory levels $y_1^{\mathcal{U}}$ and y_1^m become equal again.

(ii) $\tilde{u} = (u_2, 1 - u_2, 0)$ ($0 \leq u_2 \leq 1$) and $\tilde{w} = (1, 0)$ under which the distribution of D_2 is unknown.

In this case, λ_1, the parameter of D_1 is known ($\lambda_1 = 160$) and the buy-up substitution probability $\alpha = 0.2$. λ_2, the parameter of D_2, however, is unknown and needs to be estimated. We then vary the prior belief $\tilde{u}$ by changing the value of u_2. Note that a prior belief with a larger u_2 leads to a lower expectation on D_2. Table 2(b) shows that the optimal inventory levels under both the Bayesian inventory management and the myopic optimization increase as u_2 increases. Again, when u_2 takes the value 0 or 1, λ_2 becomes known, and thus the Bayesian optimal inventory level $y_1^{\mathcal{U}}$ is equal to the myopic inventory level y_1^m, as the inventory manager does not need to learn about the demand distribution. For other values of u_2, λ_2 is unknown. Table 2(b) reveals that to learn more about the unknown demand parameter λ_2 drives the inventory manager to stock less. Also, we observe that as u_2 increases, the difference between the two inventory levels $y_1^{\mathcal{U}}$ and y_1^m first increases, reaches the maximum at $u_2 = 0.6$, and then decreases.

(iii) $\tilde{u} = (1, 0, 0)$ and $\tilde{w} = (w, 1 - w)$ ($0 \leq w \leq 1$) under which α is unknown.

In this case, both demand parameters, λ_1 and λ_2, are known ($\lambda_1 = 160$, $\lambda_2 = 5$), but the buy-up substitution probability α is unknown. We then vary $\tilde{w}$, the prior belief about α, by changing the value of w. Note that a prior belief with a larger w leads to a lower expectation on the buy-up substitution probability α. Table 2(c) shows that the optimal inventory levels under both the Bayesian inventory management and the myopic optimization increase as w increases. This also reconfirms the result stated in Proposition 2 that a larger substitution probability incentivizes the inventory manager to reserve more regular-price seats. Again, the Bayesian optimal inventory level $y_1^{\mathcal{U}}$ is equal to the myopic inventory level y_1^m when

w takes the value 0 or 1, under which α is known. Recall from §4.3 that when lost sales are observable, learning about the unknown substitution probability drives the inventory manager to stock less so as to observe more substitution trials. Here, Table 2(c) reveals that when lost sales are unobservable, the stock less result still prevails in our numeric setting. We note that when $w \leq 0.5$, the myopic inventory level y_1^m is already very small. Under this situation, the benefit brought by stocking less cannot compensate the loss caused by it. Thus, the Bayesian optimal inventory level $y_1^{\mathcal{U}}$ is equal to y_1^m. However, when y_1^m is large (i.e, when w takes the value greater than 0.5), the value of stocking less to better observe customers' substitution behavior becomes sufficiently high, resulting in $y_1^{\mathcal{U}} < y_1^m$. Such an inventory level difference increases as w increases from 0.6 to 0.9.

A close look at the above three cases reveals that the prior beliefs about the demand parameters and the substitution probability play a critical role in whether to stock more or stock less. Besides the prior beliefs, the difference between the early-bird-discount price p_1 and the regular price p_2 is another important factor that affects the optimal inventory decisions. Recall that the regular price $p_2 = 1200$. We now fix the prior beliefs to be $\tilde{u} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\tilde{w} = (\frac{1}{2}, \frac{1}{2})$ and then vary the early-bird-discount price p_1 between 600 and 1200 with a step length of 50. Table 2(d) shows that the optimal inventory levels under both the Bayesian inventory management and the myopic optimization increase as p_1 increases. This is because the marginal profit from selling a discounted seat increases, leading to more discounted seats reserved for the early-bird sale. Table 2(d) also reveals that under our numeric setting, the stock less result is prevalent when both demand parameters and substitution probability are unknown. However, there is no monotonic relationship between the degree of stocking less $(y_1^{\mathcal{U}} - y_1^m)$ and the profit margin difference in two selling phases $(p_2 - p_1)$.

Our numerical experiments demonstrate that the system parameters—the inventory manager's prior beliefs and seat prices—critically affect the inventory manager's incentive to stock more (less). For more general settings with multiple periods, we can utilize those driving forces behind stocking more (less) to design heuristic algorithms to find solutions.

4.5 Comparison of Expected Profits under Two Information Scenarios

We now compare the system performances under the two information scenarios, which helps us to understand the value of lost sales information. Note that the total discounted expected profit from period i to N without any demand learning can be expressed as $\sum_{n=0}^{N-i} \delta^n \cdot \pi(y, \phi_i)$. We then have the following:

Proposition 7 *For period i ($i = 1, \cdots, N$), given the same prior distribution ϕ_i,*

(a) the objective functions under the two information scenarios satisfy the following relationship: $\sum_{n=0}^{N-i} \delta^n \cdot \pi(y, \phi_i) \leq G_i^{\mathcal{U}}(y, \phi_i) \leq G_i^{\mathcal{O}}(y, \phi_i)$;

(b) the optimality value functions under the two information scenarios satisfy the following relationship: $\sum_{n=0}^{N-i} \delta^n \cdot \max_{0<y\leq M} \pi(y, \phi_i) \leq v_i^{\mathcal{U}}(\phi_i) \leq v_i^{\mathcal{O}}(\phi_i)$.

Proposition 7 shows that the inventory management with Bayesian learning can always achieve a profit no less than that without any learning. Moreover, the profit is higher under the more informative $\mathcal{O}$ scenario than under the less informative $\mathcal{U}$ scenario.

5. Solution Algorithm and Performance

Solving the dynamic programming problem with demand learning is subject to the curse of dimensionality. Most studies that consider a general demand distribution focus on deriving the comparison result regarding stock more and stock less (e.g., Ding et al. 2002, Chen and Plambeck 2008). Few papers consider heuristics. We note that Chen (2010) uses the decisions under the scenario where lost sales are observable to approximate the Bayesian optimal solutions. In Chen (2010), there is only one type of demand, and the objective function of the Bayesian dynamic programming is convex; hence, the state-space reduction technique (see, e.g., Scarf 1959, Azoury 1985) can be used for some conjugate priors. However, in our setting, there are two types of discrete demands along with the buy-up substitution. The properties shown in Chen (2010) no longer hold, and thus, we cannot apply the heuristics of Chen (2010).

To find solutions to our dynamic programming problem, we adopt the *SoftMax* algorithm (see Goodfellow et al. 2016). Below, we first introduce our SoftMax algorithm and then test it on a large data set (generated via simulation). We show that the SoftMax heuristic can effectively avoid local optima traps that the myopic optimization may face. We also illustrate that the SoftMax algorithm can effectively utilize the exploitation-exploration tradeoff by comparing its performance with those associated with the myopic optimization and the commonly used Thompson sampling algorithm. We then apply the obtained stock more (less) result to speed up the convergence rate of the SoftMax algorithm.

5.1 SoftMax Algorithm

For a Bayesian inventory management problem, a good heuristic algorithm should properly capture the tradeoff between exploitation and exploration. That is, it should not only consider the profit maximization in the short run but also consider effective demand learning via stocking more (less) than the myopic decision level. In this sense, the SoftMax policy is quite suitable. For a vector of expected profits corresponding to all feasible inventory levels, the SoftMax policy converts it into a normalized probability distribution consisting of M probabilities, which are proportional to the exponentials of the expected profits. Then, the higher the expected profit of an inventory level is, the more likely this inventory level will be generated. Furthermore, as the profit function $\pi(y|\theta_1, \theta_2, \alpha)$ in (1) is unimodal in y (see Proposition 1), those inventory levels close to the myopic inventory level yield relatively high expected profits, and thus they are more likely to be generated. This unimodality property enables the SoftMax policy to well control the deviation of the generated solution from the myopic inventory level for proper exploration. Under our model setting, the corresponding SoftMax algorithm contains the following four steps.

SoftMax Algorithm

Step 1. In period i ($i = 1, \cdots, N$), based on the prior belief $\phi_i(\theta_1, \theta_2, \alpha)$, calculate $V_i(y) \triangleq E_{\phi_i}\pi(y|\theta_1, \theta_2, \alpha)$ for all $y = 1, \cdots, M$.

Step 2. Convert $V_i(y)$ into a normalized probability $P^S(y) = \frac{exp\left(\frac{V_i(y)}{\tau_i}\right)}{\sum_k exp\left(\frac{V_i(k)}{\tau_i}\right)}$, where τ_i is a system parameter[6], it decreases with i such that the chance for exploration decreases as periods move on, and it approaches very close to 0 when $\phi_i(\theta_1, \theta_2, \alpha) = 1$ for some $(\theta_1, \theta_2, \alpha)$ (i.e., we stop exploration after we identify the true parameters).

Step 3. Generate the Bayesian optimal inventory level using $P^S(\cdot)$.

Step 4. Observe demand realizations and update the belief.

A critical issue in algorithm efficiency is to avoid being trapped in local optima. We now test whether the proposed SoftMax algorithm can avoid getting stuck in local optima. Recall Example 1 in §3.2, where the myopic optimization gets stuck in a local optimum. We now apply our SoftMax algorithm to this example and assume that lost sales are unobservable. We first simulate $1,000,000$ sample paths based on the underlying true buy-up substitution probability $\alpha = 0.8$. Denote the average profit in period i, i.e., the average of the realized profits in period i over all the sample paths, under the true buy-up substitution probability (referred to as the *clairvoyant optimum* hereafter) as A_i^{True}, and the corresponding average profits under the SoftMax algorithm and myopic optimization as $A_i^{SoftMax}$ and A_i^{Myopic}, respectively. For the SoftMax algorithm, we set the system parameter $\tau_i = \frac{170090}{30+50i}$. Note that the inventory level of the discounted seat y should not exceed the upper bound of the primary demand for the discounted seat (i.e., 100). We thus generate the normalized

[6] In practice, the inventory manager can try different values of τ_i and test their performances using the existing data or datasets from the similar flights to decide which parameter value to use.

probability associated with an inventory level y as follows:

$$P^S(y) = \frac{exp\left(\frac{V_i(y)}{\tau_i}\right)}{\sum_{k=1}^{100} exp\left(\frac{V_i(k)}{\tau_i}\right)}, y = 1, 2 \cdots, 100.$$

Figure 3 depicts the average profits in each period associated with the clairvoyant optimum, the SoftMax algorithm, and the myopic optimization. It shows that compared with the myopic optimization, the SoftMax algorithm can avoid being stuck in local optima and converge to the clairvoyant optimum quickly.

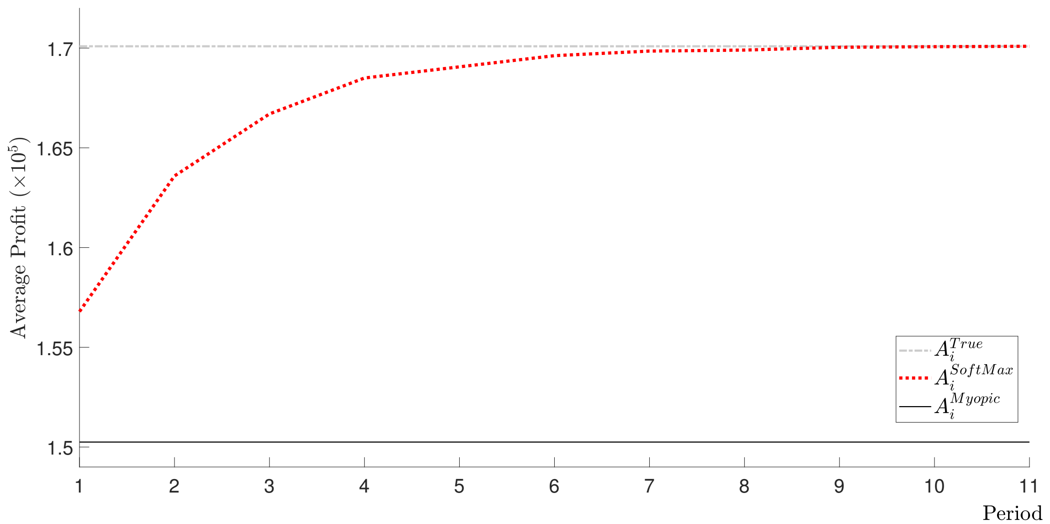


Figure 3: Average Profits under the Clairvoyant Optimum, SoftMax Algorithm and Myopic Optimization: Same Setting as Example 1 with Unobservable Lost Sales

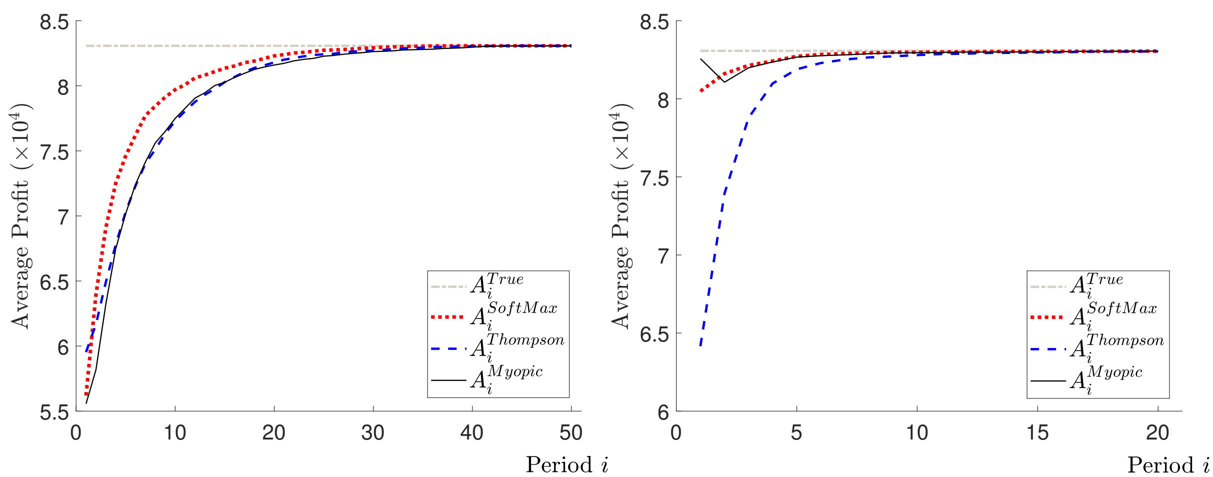
5.2 Heuristic Performance

Here, we further examine the efficiency of our proposed SoftMax algorithm. Besides the myopic optimization, we also consider the widely used Thompson sampling (also known as Bayesian posterior sampling) algorithm, which samples the demand and substitution parameters according to the prior belief.

We apply the SoftMax algorithm, myopic optimization, and Thompson sampling algorithm to the following numeric setting to examine their performances. Consider that an airline company offers an early-bird sale of a flight with a medium-sized jet and that lost sales are unobservable (i.e., the $\mathcal{U}$ scenario). The total number of seats $M = 120$, the early-bird-discount price $p_1 = 600$ and the regular price $p_2 = 1200$. The primary demands for the discounted and regular-price seats, D_1 and D_2, follow truncated Poisson distributions

$(0 \le D_1 \le 140,\ 0 \le D_2 \le 160)$ with parameters λ_1 and λ_2, respectively. The two parameters λ_1 and λ_2 are correlated, and their relationship is indicated by a parameter θ, which takes the value 1, 2 or 3. When $\theta = 1$, $\lambda_1 = 20$ and $\lambda_2 = 30$; when $\theta = 2$, $\lambda_1 = 20$ and $\lambda_2 = 100$; and when $\theta = 3$, $\lambda_1 = 80$ and $\lambda_2 = 30$. The buy-up substitution probability α takes either a low value 0.2 or a high value 0.7. At the beginning of the first period, the inventory manager holds the prior beliefs $\tilde{u} = (Pr(\theta = 1), Pr(\theta = 2), Pr(\theta = 3))$ and $\tilde{w} = (Pr(\alpha = 0.2), Pr(\alpha = 0.7))$. The underlying true parameter values are $\theta = 3$ (i.e., $\lambda_1 = 80$ and $\lambda_2 = 30$) and $\alpha = 0.2$, based on which we simulate $1,000,000$ sample paths.

Recall that when lost sales are unobservable, there is no definite result regarding whether to stock more or stock less (see §4). For such a scenario, as the inventory level is bounded above by the seat capacity M, we just randomly generate the Bayesian optimal inventory level $y_i^{\mathcal{U}}$ from 1 to M. Denote the average profit, i.e., the average of the realized profits in period i over all the sample paths under the true demand parameters and buy-up substitution probability, as A_i^{True}, and the corresponding values under the SoftMax algorithm, myopic optimization, and Thompson sampling algorithm as $A_i^{SoftMax}$, A_i^{Myopic}, and $A_i^{Thompson}$, respectively. For the SoftMax algorithm, we set the system parameter $\tau_i = \frac{\max_{\theta,\alpha} \max_{y} \pi(y|\theta,\alpha)}{500+40i}$, and the normalized probability associated with an inventory level y is $P^S(y) = \frac{exp\left(\frac{V_i(y)}{\tau_i}\right)}{\sum_{k=1}^{M} exp\left(\frac{V_i(k)}{\tau_i}\right)}$, $y = 1, \cdots, M$.



(a) $\tilde{u} = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$ and $\tilde{w} = (\frac{1}{4}, \frac{3}{4})$

(b) $\tilde{u} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\tilde{w} = (\frac{1}{2}, \frac{1}{2})$

Figure 4: Average Profits under the Clairvoyant Optimum, SoftMax Algorithm, Myopic Optimization, and Thompson Sampling: Unobservable Lost Sales

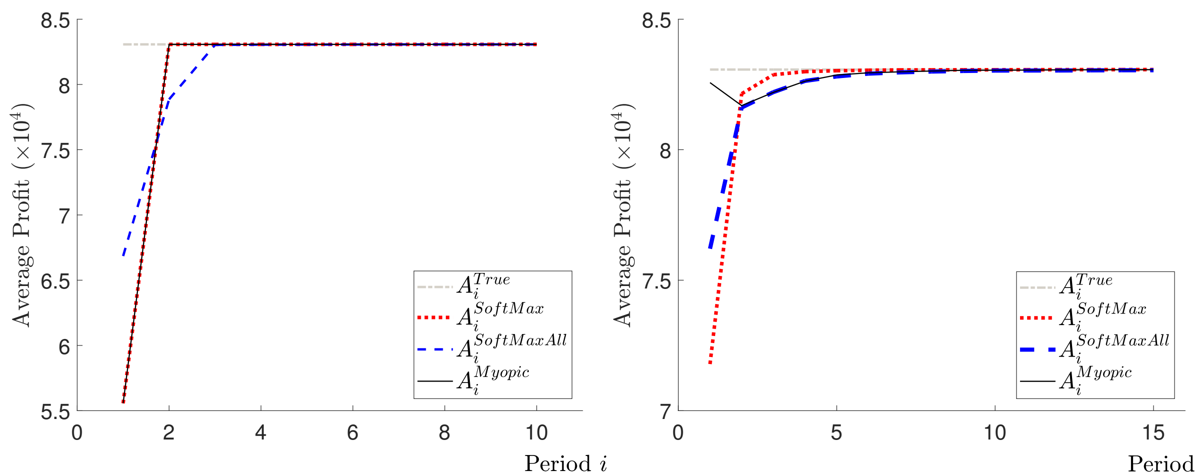
Throughout our numerical experiments with various combinations of prior beliefs $\tilde{u}$ and $\tilde{w}$, we identify two representative patterns regarding the performances of the average profits under the three optimization approaches. The first pattern occurs when the inventory manager's prior beliefs $\tilde{u}$ and $\tilde{w}$ deviate significantly from the true values. We now consider the prior beliefs $\tilde{u} = \left(\frac{3}{4}, \frac{1}{8}, \frac{1}{8}\right)$ and $\tilde{w} = \left(\frac{1}{4}, \frac{3}{4}\right)$ as an example for illustration. The corresponding average profits in each period under the clairvoyant optimum, SoftMax algorithm, myopic optimization, and Thompson sampling are depicted in Figure 4(a). Figure 4(a) shows that in such a situation, compared with myopic optimization, the SoftMax algorithm not only always yields a higher average profit but also converges to the optimal profit much faster. Although the Thompson sampling algorithm results in the highest profit in the first period, its speed of convergence to the optimal profit is slower than that of the SoftMax algorithm. This implies that the way the Thompson sampling captures the exploration–exploitation tradeoff (i.e., randomly generating demand and substitution parameters according to the prior beliefs) explores the demand information less efficiently than what the SoftMax algorithm does (i.e., randomly generating the Bayesian optimal inventory level according to a normalized probability distribution, which consists of probabilities proportional to the exponentials of the expected profits corresponding to all feasible inventory levels).

The second pattern occurs when the prior beliefs are either uninformative or very close to the true values. We now take the uninformative prior beliefs $\tilde{u} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\tilde{w} = \left(\frac{1}{2}, \frac{1}{2}\right)$ as an example for illustration. The corresponding average profits in each period under the clairvoyant optimum, SoftMax algorithm, myopic optimization, and Thompson sampling are depicted in Figure 4(b). Figure 4(b) shows that in this situation, the myopic optimization results in a higher average profit in the first period than the SoftMax and Thompson sampling algorithms, as the latter two are exploring the demand information via deviating from the myopic inventory level.[7] Again, Figure 4(b) shows that the SoftMax algorithm converges to the optimal profit at a much faster speed than both the myopic optimization and Thompson sampling algorithm do.

Next, we reconsider the above Bayesian inventory management problem by assuming that lost sales can be observed (i.e., the $\mathcal{O}$ scenario). For such a scenario, the stock less result holds (see Proposition 6). Then, the optimal inventory level is bounded above by the myopic inventory level. We now integrate the stock less result into the SoftMax algorithm and randomly generate the Bayesian optimal inventory level $y_i^{\mathcal{U}}$ from 1 to y_i^m. Then, the normalized probability associated with an inventory level y becomes $P^S(y) = \frac{exp\left(\frac{V_i(y)}{\tau_i}\right)}{\sum_{k=1}^{y_i^m} exp\left(\frac{V_i(k)}{\tau_i}\right)}$ for $y = 1, \cdots, y_i^m$ and $P^S(y) = 0$ for $y = y_i^m + 1, \cdots, M$. For comparison, we also run the SoftMax algorithm without utilizing the stock less result, under which we consider all possible inventory levels and $P^S(y) = \frac{exp\left(\frac{V_i(y)}{\tau_i}\right)}{\sum_{k=1}^{M} exp\left(\frac{V_i(k)}{\tau_i}\right)}$, $y = 1, \cdots, M$. The system parameters

[7] There is an interesting pattern in Figure 4(b) that the average profit A_i^{Myopic} decreases as we progress from period 1 to period 2. This is because the myopic inventory level may happen to be very close to the clairvoyant optimal one when the belief is quite imprecise but far away from it when the belief is relatively accurate.

τ_i under these two SoftMax algorithms are set to be the same with value $\tau_i = \frac{\max_{\theta,\alpha} \max_{y} \pi(y|\theta,\alpha)}{100+20i}$. Again, we simulate $1,000,000$ sample paths based on the true demand and substitution parameters. Denote the average realized profits in period i under the clairvoyant optimum, the myopic optimization, the SoftMax algorithm utilizing the stock less result, and the SoftMax algorithm without utilizing the stock less result as A_i^{True}, A_i^{Myopic}, $A_i^{SoftMax}$, and $A_i^{SoftMaxAll}$, respectively.



(a) $\tilde{u} = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$ and $\tilde{w} = (\frac{1}{4}, \frac{3}{4})$

(b) $\tilde{u} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\tilde{w} = (\frac{1}{2}, \frac{1}{2})$

Figure 5: Average Profits under the Clairvoyant Optimum, Myopic Optimization, SoftMax Algorithm Utilizing the Stock Less Result and SoftMax Algorithm without Utilizing the Stock Less Result: Observable Lost Sales

By varying the combination of prior beliefs $\tilde{u}$ and $\tilde{w}$, our numerical experiments show that compared with the other algorithms, the SoftMax algorithm that utilizes the stock less result updates the posterior beliefs to approach the true values of unknown parameters more quickly; see the two representative patterns in Figure 5 for illustration. That is, the utilization of the stock less result can help speed up the algorithm convergence to the optimal solution.

6. Conclusion

In this study, we investigate an airline company's optimal seat allocation decision when it provides the early-bird booking discount under a multi-period setting. The capacity of seats is fixed, and within each period, the company offers a price discount in the early-bird-discount phase and charges the full price in the regular-price phase. When the discounted seats are sold out, unsatisfied early-bird customers may purchase the regular-price seat as a substitute. Setting a proper inventory level of the discounted seats is critical for the company's revenue management. However, an optimal seat inventory allocation requires the inventory manager to have knowledge about the primary demands for the discounted and regular-price seats and the buy-up substitution probability of those unsatisfied early-bird customers, which are often unknown. Incorporating demand learning into the seat inventory allocation then becomes important for the inventory manager. To that end, we develop a dynamic inventory management model with Bayesian learning about both demand parameters and buy-up substitution probability.

We examine two information scenarios based on whether or not lost sales are observable. Under each scenario, we compare the Bayesian optimal inventory level of the discounted seat with the corresponding myopic inventory level to examine whether the inventory manager needs to stock more (less) to observe and learn more about demand distributions and customer substitution behavior. When only demand parameters are unknown and there is no buy-up substitution, we show that the unobservability of lost sales drives the inventory managers to stock more to better learn about the primary demand, a result also observed in previous studies on Bayesian inventory management, such as Lariviere and Porteus (1999) and Ding et al. (2002). However, when buy-up substitution exists, unobservable lost sales may drive the inventory manager to stock less. This is in sharp contrast to the stock more result of Chen and Plambeck (2008). In our setting, the number of the substitute product, regular-price seat, is limited, whereas in Chen and Plambeck (2008), the substitute product is always available. Stocking fewer discounted seats can induce more substitution trials and thus help the inventory manager to infer some information about the primary demand from the observed substitution behavior. When both demand parameters and the buy-up substitution probability are unknown, learning about the substitution probability drives the inventory manager to stock less to induce more substitution events to occur, while learning about the primary demands may induce the inventory manager to stock more. Whether to stock more or stock less depends on multiple factors, including seat prices and the inventory manager's prior beliefs.

Last, we provide a SoftMax algorithm to find the optimal solution of our dynamic programming problem. Our numerical experiments reveal that the proposed SoftMax algorithm can effectively avoid being trapped in local optima, an issue that the myopic optimization may not be able to escape. The SoftMax algorithm outperforms both the myopic optimization and the Thompson sampling algorithm in the sense that it can converge much faster to the optimal solution. We further demonstrate that the stock more (less) result can be utilized to explore and exploit the demand information more efficiently and thus can improve the convergence speed of the algorithm.

We conclude this paper by discussing some limitations and directions for future research.

Our study considers a multi-period two-fare seat allocation problem. A direct extension is to consider multiple fares with multiple advance purchase deadlines. Also, the prices are fixed in our model. Future research can consider joint inventory and pricing decisions with learning about demand distribution and substitution parameters. Finally, our model does not consider customers' no-show behaviour. It would be interesting to embed such behaviour into the model and examine how it affects the seat allocation decisions.

Acknowledgments

The authors gratefully thank the department editor (Prof. Kamalini Ramdas), an anonymous associate editor, and two anonymous referees for their very helpful comments and suggestions. The first author Zhenwei Luo acknowledges the financial support by the Internal Start-up Fund of the Hong Kong Polytechnic University (Project ID: P0039035) and the National Natural Science Foundation of China (Grant No. 71971184). The corresponding author Pengfei Guo acknowledges the financial support from the Research Grants Council of Hong Kong (No. 15508518). And the third author Yulan Wang's work was supported by the Research Grants Council of Hong Kong (RGC Reference Number: 15505318). All authors contributed equally to the work.

References

Anupindi, R., M. Dada, S. Gupta. 1998. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* 17(4): 406-423.

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

Belobaba, P. 1987. Air travel demand and airline seat inventory management. Massachusetts Institute of Technology.

Bensoussan, A., M. Çakanyıldırım, M. Li, et al. 2016. Managing inventory with cash register information: Sales recorded but not demands. *Production and Operations Management* 25(1): 9-21.

Bensoussan, A., M. Çakanyıldırım, S. P. Sethi. 2009. A note on "The censored newsvendor and the optimal acquisition of information". *Operations Research* 57(3): 791-794.

Bensoussan, A., P. Guo. 2015. Technical note—Managing nonperishable inventories with learning about demand arrival rate through stockout times. *Operations Research* 63(3): 602-609.

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

Brumelle, S. L., J. I. McGill, T. H. Oum, et al. 1990. Allocation of airline seats between stochastically dependent demands. *Transportation Science* 24(3): 183-192.

Chen, B., X. Chao. 2019. Parametric demand learning with limited price explorations in a backlog stochastic inventory system. *IISE Transactions* 51(6): 605-613.

Chen, B., X. Chao. 2020. Dynamic inventory control with stockout substitution and demand learning. *Management Science* 66(11): 5108-5127.

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

Chen, L., A. J. Mersereau, Z. Wang. 2017. Optimal merchandise testing with limited inventory. *Operations Research* 65(4): 968-991.

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

Chen, L., C. Wu. 2019. Open-loop policies in Bayesian dynamic pricing: Some counter-intuitive observations and insights. *Operations Research Letters* 47(5): 331-338.

Chu, L. Y., J. G. Shanthikumar, Z. J. M. Shen. 2008. Solving operational statistics via a Bayesian analysis. *Operations Research Letters* 36(1): 110-116.

Cooper, W. L., T. Homem-de-Mello, A. J. Kleywegt. 2006. Models of the spiral-down effect in revenue management. *Operations Research* 54(5): 968-987.

Cooper, W. L., L. Li. 2012. On the use of buy up as a model of customer choice in revenue management. *Production and Operations Management* 21(5): 833-850.

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

Feng, Q., J. G. Shanthikumar. 2017. Supply and demand functions in inventory models. *Operations Research* 66(1): 77-91.

Gallego, G., L. Li, R. Ratliff. 2009. Choice-based EMSR methods for single-leg revenue management with demand dependencies. *Journal of Revenue and Pricing Management* 8(2/3): 207–240.

Goodfellow, I., Y. Bengio, A. Courville. 2016. Deep learning. MIT Press.

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

Hotle, S. L., M. Castillo, L. A. Garrow, M. J. Higgins. 2015. The impact of advance purchase deadlines on airline consumers' search and purchase behaviors. *Transportation Research Part A: Policy and Practice* 82: 1-16.

Hu, P., S. Stephen, Y. Man. 2015. Joint inventory and markdown management for perishable goods with strategic consumer behavior. *Operations Research* 64(1): 118-134.

Huh, W. T., P. Rusmevichientong. 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* 34(1): 103-123.

Ja, S., B. V. Rao, S. Chandler. 2001. Passenger recapture estimation in airline revenue management. *AGIFORS 41st Annual Symposium*, AGIFORS, Sydney, Australia.

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

Kök, A. G., M. L. Fisher. 2007. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6): 1001-1021.

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

Li, R., J. K. Ryan. 2011. A Bayesian inventory model using real-time condition monitoring information. *Production and Operations Management* 20(5): 754-771.

Li, R., J. S. Song, S. Sun, X. Zheng. 2022. Fight inventory shrinkage: Simultaneous learning of inventory level and shrinkage rate. *Production and Operations Management* 31(6): 2477-2491.

Littlewood, K. 1972. Forecasting and control of passenger bookings. *Airline Group International Federation of Operational Research Societies Proceedings* 12: 95-117.

Liyanage, L. H., J. G. Shanthikumar. 2005. A practical inventory control policy using operational statistics. *Operations Research Letters* 33(4): 341-348.

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

Lu, X., J. S. Song, K. Zhu. 2005. On "The censored newsvendor and the optimal acquisition of information". *Operations Research* 53(6): 1024-1026.

McGill, J. I., G. van Ryzin. 1999. Revenue management: Research overview and prospects. *Transportation Science* 33(2): 233-256.

Miller, B. L. 1986. Scarf's state reduction method, flexibility, and a dependent demand inventory model. *Operations Research* 34(1): 83-90.

Pfeifer, P. E. 1989. The airline discount fare allocation problem. *Decision Sciences* 20(1): 149-157.

Ratliff, R. M., B. V. Rao, C. P. Narayan, K. Yellepeddi. 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management* 7(2): 153-171.

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

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

Shaw, S. 1982. Air transport: A marketing perspective. Pitman Books, London.

The Alan Turing Institute. 2020. Dynamic forecasting with British Airways. Retrieved Jul 2021, from https://urldefense.com/v3/__https://www.turing.ac.uk/research/impact-stories/dynamic-forecasting-british-airways__;!!KjDnqvtInNPT!0DyKjrNlpO90wTYJKxiU4Tu5jbh595wm6x0b9E5J4GItCv9SANy4GbXcyReYEQwjBis$.

U.S. Department of Transportation. 2019. Fly rights: A consumer guide to air travel. Retrieved Jul 2021, from https://urldefense.com/v3/__https://www.transportation.gov/airconsumer/fly-rights__;!!KjDnqvtInNPT!0DyKjrNlpO90wTYJKxiU4Tu5jbh595wm6x0b9E5J4GItCv9SANy4GbXcyReY82Ksy6g$.

Ulu, C., D. Honhon, A. Alptekinoglu. 2012. Learning consumer tastes through dynamic assortments. *Operations Research* 60(4): 833-849.

van Ryzin, G., J. McGill. 2000. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science* 46(6): 760-775.

Vulcano, G., G. van Ryzin, R. Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* 60(2): 313-334.

Yuan, H., Q. Luo, C. Shi. 2021. Marrying stochastic gradient descent with bandits: Learning algorithms for inventory systems with fixed costs. *Management Science* 67(10): 6089-6115.

Online Appendix
"Manage Inventories with Learning on Demands and Buy-up Substitution Probability"

Proof of Proposition 1: Through some derivation, we can show that

$$\pi(y+1|\theta_1,\theta_2,\alpha)-\pi(y|\theta_1,\theta_2,\alpha)=(p_1-p_2)\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)$$
$$+(1-\alpha)p_2\sum_{i=y+1}^{+\infty}\left\{\sum_{j=0}^{M-y-1}\binom{i-y-1}{j}\alpha^j(1-\alpha)^{i-y-j-1}\left[\sum_{k=0}^{M-y-j-1}f_1(i|\theta_1)f_2(k|\theta_2)\right]\right\}$$
$$=(p_1-p_2)\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)+(1-\alpha)p_2\sum_{i=y+1}^{+\infty}f_1(i)Pr(K+D_2\leq M-y-1|D_1=i,\theta_1,\theta_2,\alpha)$$
$$=(p_1-\alpha p_2)\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)-(1-\alpha)p_2\sum_{i=y+1}^{+\infty}f_1(i)Pr(K+D_2>M-y-1|D_1=i,\theta_1,\theta_2,\alpha),$$

where $Pr(\cdot)$ denotes probability. If $f_1(i|\theta_1)=0$ for all $i>y$, then $\pi(y+1|\theta_1,\theta_2,\alpha)-\pi(y|\theta_1,\theta_2,\alpha)=0$. Otherwise, $\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)>0$, and we can show that

$$\pi(y+1|\theta_1,\theta_2,\alpha)-\pi(y|\theta_1,\theta_2,\alpha)$$
$$=\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)\left[(p_1-\alpha p_2)-(1-\alpha)p_2Pr(K+D_2>M-y-1|D_1>y,\theta_1,\theta_2,\alpha)\right]$$
$$\triangleq\sum_{i=y+1}^{+\infty}f_1(i|\theta_1)G(y+1|\theta_1,\theta_2,\alpha).$$

Note that the above $G(y|\theta_1,\theta_2,\alpha)$ is exactly the marginal revenue defined in Brumelle et al. (1990). According to Brumelle et al. (1990), $G(y|\theta_1,\theta_2,\alpha)$ is nonincreasing in y. Along with $\sum_{i=y+1}^{+\infty}f_1(i)>0$, we can then prove the proposition.

Proof of Proposition 2: Through some derivation, we can show that

$$\pi(y+1|\theta_1,\theta_2,\alpha)-\pi(y|\theta_1,\theta_2,\alpha)$$
$$=(p_1-p_2)\sum_{i=y+1}^{+\infty}f_1(i|\theta_1,\theta_2)$$
$$+(1-\alpha)p_2\sum_{i=y+1}^{+\infty}\left\{\sum_{j=0}^{M-y-1}\binom{i-y-1}{j}\alpha^j(1-\alpha)^{i-y-j-1}\left[\sum_{k=0}^{M-y-j-1}f_1(i|\theta_1)f_2(k|\theta_2)\right]\right\}$$
$$=(p_1-p_2)\sum_{i=y+1}^{+\infty}f_1(i|\theta_1,\theta_2)+(1-\alpha)p_2\sum_{i=y+1}^{+\infty}f_1(i)Pr(K+D_2\leq M-y-1|D_1=i,\theta,\alpha),$$

where $Pr(\cdot)$ denotes probability. Consider the two substitution probabilities α^1 and α^2 with $\alpha^1 < \alpha^2$. Denote the corresponding variables of K as K^1 and K^2, respectively. Obviously, the variable $K^2 + D_2$ is stochastically larger than $K^1 + D_2$. Thus, we have

$$Pr(K^1 + D_2 \leq M - y - 1 | D_1 = i, \theta, \alpha^1) \geq Pr(K^2 + D_2 \leq M - y - 1 | D_1 = i, \theta, \alpha^2).$$

Then, it is easy to verify that $\frac{\partial[\pi(y+1|\theta_1,\theta_2,\alpha) - \pi(y|\theta_1,\theta_2,\alpha)]}{\partial \alpha} < 0$, based on which we can obtain the result in Proposition 2.

Proof of Proposition 4: Under the $\mathcal{U}$ scenario, when the substitution probability is 0, the amount of substitution demands s_{21} is always 0. Thus, the likelihood of observing sales quantities (s_1, s_{22}) can be expressed as

$$\begin{aligned} & f^y_{\mathcal{U}}(s_1, s_{22}|\theta_1, \theta_2, 0) \\ = & \begin{cases} f_1(s_1|\theta_1) f_2(s_{22}|\theta_2), & \text{if } s_1 < y \text{ and } s_{22} < M - s_1; \\ f_1(s_1|\theta_1) \sum\limits_{j=M-s_1}^{+\infty} f_2(j|\theta_2), & \text{if } s_1 < y \text{ and } s_{22} = M - s_1; \\ \sum\limits_{i=y}^{+\infty} f_1(i|\theta_1) f_2(s_{22}|\theta_2), & \text{if } s_1 = y \text{ and } s_{22} < M - y; \\ \sum\limits_{i=y}^{+\infty} f_1(i|\theta_1) \sum\limits_{j=M-y}^{+\infty} f_2(j|\theta_2), & \text{if } s_1 = y \text{ and } s_{22} = M - y. \end{cases} \end{aligned}$$

When the demand parameter θ_2 is known but θ_1 is unknown, we can write $f^y_{\mathcal{U}}(s_1, s_{22}|\theta_1)$ as shorthand for $f^y_{\mathcal{U}}(s_1, s_{22}|\theta_1, \theta_2, \alpha)$. Below, we will first show that for $i = 1, \cdots, N-1$, $0 < y < M$, and any $\phi'_{i,1}(\theta_1)$,
(a) when $s_1 < y$ and $s_{22} < M - s_1$,

$$E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f^y_{\mathcal{U}}(s_1, s_{22}|\theta_1) \right\} = f_2(s_{22}|\theta_2) \cdot E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f_1(s_1|\theta_1) \right\}; \quad (12)$$

(b) when $s_1 < y$ and $s_{22} = M - s_1$,

$$E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f^y_{\mathcal{U}}(s_1, M - s_1|\theta_1) \right\} = \left[\sum_{j=M-s_1}^{+\infty} f_2(j|\theta_2) \right] \cdot E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f_1(s_1|\theta_1) \right\}; \quad (13)$$

(c) when $s_1 = y$ and $s_{22} < M - y$,

$$\begin{aligned} & E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \right\} = f_2(s_{22}|\theta_2) \cdot E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) \sum_{i=y}^{+\infty} f_1(i|\theta_1) \right\} \\ & \leq f_2(s_{22}|\theta_2) \cdot E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f_1(y|\theta_1) + v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1) \right\}; \quad (14) \end{aligned}$$

(d) when $s_1 = y$ and $s_{22} = M - y$,

$$E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f^y_{\mathcal{U}}(y, M - y|\theta_1) \right\} = \left[\sum_{j=M-y}^{+\infty} f_2(j|\theta_2) \right] \cdot E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) \sum_{i=y}^{+\infty} f_1(i|\theta_1) \right\}$$

$$\leq \left[\sum_{j=M-y}^{+\infty} f_2(j|\theta_2)\right] E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f_1(y|\theta_1) + v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1) \right\}. \tag{15}$$

For above four relationships, we only prove the third one, i.e., (14), for illustration purpose. The others can be proved following the same logic.

At first, the equality relationship in (14) is obvious as $f_2(s_{22}|\theta_2)$ is independent of θ_1. As for the inequality "$\leq$" relationship, according to the backward induction, we have that when $n = N-1$,

$$\begin{aligned}
&E_{\phi'_{N-1,1}(\theta_1)} \left\{ v^{\mathcal{U}}_N(\phi'_{N,1}) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \right\} \\
=&\int_{\Theta_1} \max_{0<y'\leq M} \left\{ \int_{\Theta_1} \pi(y'|\theta'_1) \phi'_{N-1,1}(\theta'_1|y, s_{22}, y, \phi'_{N-1,1}) d\theta'_1 \right\} \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1) f_2(s_{22}|\theta_2)\right] \phi'_{N-1,1}(\theta_1) d\theta_1 \\
=&f_2(s_{22}|\theta_2) \max_{0<y'\leq M} \left\{ \int_{\Theta_1} \pi(y'|\theta'_1) \frac{\left[\sum_{i=y}^{+\infty} f_1(i|\theta'_1)\right] \phi'_{N-1,1}(\theta'_1)}{\int_{\Theta_1} \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1)\right] \phi'_{N-1,1}(\theta_1) d\theta_1} d\theta'_1 \right\} \int_{\Theta_1} \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1)\right] \phi'_{N-1,1}(\theta_1) d\theta_1 \\
=&f_2(s_{22}|\theta_2) \max_{0<y'\leq M} \left\{ \int_{\Theta_1} \pi(y'|\theta'_1) \left[\sum_{i=y}^{+\infty} f_1(i|\theta'_1)\right] \phi'_{N-1,1}(\theta'_1) d\theta'_1 \right\} \\
\leq&f_2(s_{22}|\theta_2) \max_{0<y'\leq M} \left\{ \int_{\Theta_1} \pi(y'|\theta'_1) f_1(y|\theta'_1) \phi'_{N-1,1}(\theta'_1) d\theta'_1 \right\} \\
&+ f_2(s_{22}|\theta_2) \max_{0<y'\leq M} \left\{ \int_{\Theta_1} \pi(y'|\theta'_1) \left[\sum_{i=y+1}^{+\infty} f_1(i|\theta'_1)\right] \phi'_{N-1,1}(\theta'_1) d\theta'_1 \right\} \\
=&f_2(s_{22}|\theta_2) E_{\phi'_{N-1,1}(\theta_1)} \left\{ v^{\mathcal{U}}_N(\phi'_{N,1}) f_1(y|\theta_1) \right\} + f_2(s_{22}|\theta_2) E_{\phi'_{N-1,1}(\theta_1)} \left\{ v^{\mathcal{U}}_N(\phi'_{N,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1) \right\}.
\end{aligned}$$

So, the inequality relationship holds for period $N-1$. Then, assume that it holds for period $n+1$ ($n = 1, \cdots, N-2$), and let us check whether it holds for period n. We can show that

$$\begin{aligned}
&E_{\phi'_{n,1}(\theta_1)} \left\{ v^{\mathcal{U}}_{n+1}(\phi'_{n+1,1}) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \right\} \\
=&\max_{0<y'\leq M} \Bigg\{ \int_{\Theta_1} \pi(y'|\theta_1) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \phi'_{n,1}(\theta_1) d\theta_1 \\
&+ \delta \sum_{s'_1} \sum_{s'_{22}} \int_{\Theta_1} v^{\mathcal{U}}_{n+2}(\phi'_{n+2}(\theta'_1|s'_1, s'_{22}, y'|y, s_{22}, y|\phi'_{n,1})) f^{y'}_{\mathcal{U}}(s'_1, s'_{22}|\theta_1) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \phi'_{n,1}(\theta_1) d\theta_1 \Bigg\} \\
=&\max_{0<y'\leq M} \Bigg\{ \int_{\Theta_1} \pi(y'|\theta_1) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) \phi'_{n,1}(\theta_1) d\theta_1 \\
&+ \delta \sum_{s'_1} \sum_{s'_{22}} \int_{\Theta_1} v^{\mathcal{U}}_{n+2}(\phi'_{n+2}(\theta'_1|y, s_{22}, y|s'_1, s'_{22}, y'|\phi'_{n,1})) f^y_{\mathcal{U}}(y, s_{22}|\theta_1) f^{y'}_{\mathcal{U}}(s'_1, s'_{22}|\theta_1) \phi'_{n,1}(\theta_1) d\theta_1 \Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&= \max_{0<y'\leq M} \Bigg\{ \int_{\Theta_1} \pi(y'|\theta_1) f_{\mathcal{U}}^{y}(y, s_{22}|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \\
&\quad + \delta \sum_{s'_1}\sum_{s'_{22}} E_{\phi'_{n+1,1}(\theta_1|s'_1,s'_{22},y',\phi'_{n,1})} \left[v_{n+2}^{\mathcal{U}}(\phi'_{n+2,1}(\theta'_1|y, s_{22}, y, \phi'_{n+1,1})) f_{\mathcal{U}}^{y}(y, s_{22}|\theta_1)\right] \\
&\qquad\qquad\qquad \cdot \int_{\Theta_1} f_{\mathcal{U}}^{y'}(s'_1, s'_{22}|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \Bigg\} \\
&\leq \max_{0<y'\leq M} \Bigg\{ f_2(s_{22}|\theta_2) \int_{\Theta_1} \pi(y'|\theta_1) \left[f_1(y|\theta_1) + \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right] \phi'_{n,1}(\theta_1)d\theta_1 \\
&\quad + \delta \sum_{s'_1}\sum_{s'_{22}} \Bigg\{ f_2(s_{22}|\theta_2) E_{\phi'_{n+1,1}(\theta_1|s'_1,s'_{22},y',\phi'_{n,1})} \left[v_{n+2}^{\mathcal{U}}(\phi'_{n+2,1}(\theta'_1|y, s_{22}, y, \phi'_{n+1,1})) f_1(y|\theta_1)\right] \\
&\qquad + f_2(s_{22}|\theta_2) E_{\phi'_{n+1,1}(\theta_1|s'_1,s'_{22},y',\phi'_{n,1})} \left[v_{n+2}^{\mathcal{U}}(\phi'_{n+2,1}(\theta'_1|y, s_{22}, y, \phi'_{n+1,1})) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right] \Bigg\} \\
&\qquad\qquad\qquad \cdot \int_{\Theta_1} f_{\mathcal{U}}^{y'}(s'_1, s'_{22}|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \Bigg\} \\
&\leq \max_{0<y'\leq M} \Bigg\{ f_2(s_{22}|\theta_2) \int_{\Theta_1} \pi(y'|\theta_1) f_1(y|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \\
&\quad + \delta \sum_{s'_1}\sum_{s'_{22}} f_2(s_{22}|\theta_2) E_{\phi'_{n+1,1}(\theta_1|s'_1,s'_{22},y',\phi'_{n,1})} \left[v_{n+2}^{\mathcal{U}}(\phi'_{n+2,1}(\theta'_1|y, s_{22}, y, \phi'_{n+1,1})) f_1(y|\theta_1)\right] \\
&\qquad\qquad\qquad \cdot \int_{\Theta_1} f_{\mathcal{U}}^{y'}(s'_1, s'_{22}|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \Bigg\} \\
&\quad + \max_{0<y'\leq M} \Bigg\{ f_2(s_{22}|\theta_2) \int_{\Theta_1} \pi(y'|\theta_1) \left[\sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right] \phi'_{n,1}(\theta_1)d\theta_1 \\
&\quad + \delta \sum_{s'_1}\sum_{s'_{22}} f_2(s_{22}|\theta_2) E_{\phi'_{n+1,1}(\theta_1|s'_1,s'_{22},y',\phi'_{n,1})} \left[v_{n+2}^{\mathcal{U}}(\phi'_{n+2,1}(\theta'_1|y, s_{22}, y, \phi'_{n+1,1})) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right] \\
&\qquad\qquad\qquad \cdot \int_{\Theta_1} f_{\mathcal{U}}^{y'}(s'_1, s'_{22}|\theta_1)\phi'_{n,1}(\theta_1)d\theta_1 \Bigg\} \\
&= f_2(s_{22}|\theta_2) E_{\phi'_{n,1}(\theta_1)} \left\{ v_{n+1}^{\mathcal{U}}(\phi'_{n+1,1}) f_1(y|\theta_1)\right\} + f_2(s_{22}|\theta_2) E_{\phi'_{n,1}(\theta_1)} \left\{ v_{n+1}^{\mathcal{U}}(\phi'_{n+1,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right\}.
\end{aligned}
$$

Hence, we prove (14). Then, through some derivation, we can obtain that

$$
\begin{aligned}
&E_{\phi'_{i,1}(\theta_1)} \left\{ \sum_{s_1}\sum_{s_{22}} v_{i+1}^{\mathcal{U}}(\phi'_{i+1,1}) f_{\mathcal{U}}^{y}(s_1, s_{22}|\theta_1)\right\} \\
&\leq \left[\sum_{j=M-y}^{+\infty} f_2(j|\theta_2)\right] E_{\phi'_{i,1}(\theta_1)} \left\{ v_{i+1}^{\mathcal{U}}(\phi'_{i+1,1}) f_1(y|\theta_1)\right\} + \left[\sum_{j=M-y}^{+\infty} f_2(j|\theta_2)\right] E_{\phi'_{i,1}(\theta_1)} \left\{ v_{i+1}^{\mathcal{U}}(\phi'_{i+1,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{s_{22}=0}^{M-y-1} f_2(s_{22}|\theta_2) E_{\phi'_{i,1}(\theta_1)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,1}) f_1(y|\theta_1)\right\} + \sum_{s_{22}=0}^{M-y-1} f_2(s_{22}|\theta_2) E_{\phi'_{i,1}(\theta_1)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,1}) \sum_{i=y+1}^{+\infty} f_1(i|\theta_1)\right\}\\
&+\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1} f_2(s_{22}|\theta_2) E_{\phi'_{i,1}(\theta_1)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,1}) f_1(s_1|\theta_1)\right\}\\
&+\sum_{s_1=0}^{y-1}\left[\sum_{j=M-s_1}^{+\infty} f_2(j|\theta_2)\right] E_{\phi'_{i,1}(\theta_1)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,1}) f_1(s_1|\theta_1)\right\}\\
=&E_{\phi'_{i,1}(\theta_1)}\left\{\sum_{s_1}\sum_{s_{22}} v^{\mathcal{U}}_{i+1}(\phi'_{i+1,1}) f^{y+1}_{\mathcal{U}}(s_1, s_{22}|\theta_1)\right\},
\end{aligned}
$$

which implies that $y^{\mathcal{U}}_i \geq y^m_i$.

Following the same procedure, we can prove that under the $\mathcal{U}$ scenario, when the demand parameter θ_1 is known and $\alpha = 0$ but θ_2 is unknown, for any period i ($i = 1, \cdots, N$), given the same prior distribution $\phi'_{i,2}(\theta_2)$, learning about θ_2 requires $y^{\mathcal{U}}_i \leq y^m_i$. In this situation, we write $f^y_{\mathcal{U}}(s_1, s_{22}|\theta_2)$ as shorthand for $f^y_{\mathcal{U}}(s_1, s_{22}|\theta_1, \theta_2, \alpha)$. We can show that for $i = 1, \cdots, N-1$, $1 < y \leq M$, and any $\phi'_{i,2}(\theta)$, the following four relationships hold, based on which we can obtain $y^{\mathcal{U}}_i \leq y^m_i$:
(a') when $s_1 < y$ and $s_{22} < M - s_1$,

$$E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f^y_{\mathcal{U}}(s_1, s_{22}|\theta_2)\right\} = f_1(s_1|\theta_1)\cdot E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f_2(s_{22}|\theta_2)\right\};$$

(b') when $s_1 < y$ and $s_{22} = M - s_1$,

$$E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f^y_{\mathcal{U}}(s_1, M-s_1|\theta_2)\right\} = f_1(s_1|\theta_1)\cdot E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) \sum_{j=M-s_1}^{+\infty} f_2(j|\theta_2)\right\};$$

(c') when $s_1 = y$ and $s_{22} < M - y$,

$$E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f^y_{\mathcal{U}}(y, s_{22}|\theta_2)\right\} = \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1)\right]\cdot E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f_2(s_{22}|\theta_2)\right\};$$

(d') when $s_1 = y$ and $s_{22} = M - y$,

$$
\begin{aligned}
&E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) f^y_{\mathcal{U}}(y, M-y|\theta_2)\right\} = \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1)\right]\cdot E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) \sum_{j=M-y}^{+\infty} f_2(j|\theta_2)\right\}\\
&\leq \left[\sum_{i=y}^{+\infty} f_1(i|\theta_1)\right]\cdot E_{\phi'_{i,2}(\theta_2)}\left\{v^{\mathcal{U}}_{i+1}(\phi'_{i+1,2}) \left(f_2(M-y|\theta_2) + \sum_{j=M-y+1}^{+\infty} f_2(j|\theta_2)\right)\right\}.
\end{aligned}
$$

Proof of Proposition 5: Under the setting given in Proposition 5, D_2 and α are both known. Thus, we can write $f^y_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1)$ as shorthand for $f^y_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1, \theta_2, \alpha)$. As $M = 2$, the value of y can only be 1 or 2. Then, we can get the following relationship between the two likelihood functions $f^1_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1)$ and $f^2_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1)$ based on the assumption that $f_1(1|\theta_1) = 0$ for all $\theta_1 \in \Theta_1$:

$$
\begin{aligned}
&f^2_{\mathcal{U}}(s_1, s_{21}, s_{22}|\theta_1)\\
&= \begin{cases}
f_1(0|\theta_1)f_2(0), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 0,\\
f_1(0|\theta_1)f_2(1), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 1,\\
f_1(0|\theta_1)\sum_{j=2}^{+\infty} f_2(j), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 2,\\
\sum_{i=2}^{+\infty} f_1(i|\theta_1), & \text{if } s_1 = 2,\ s_{21} = 0 \text{ and } s_{22} = 0,
\end{cases}\\
&= \begin{cases}
f_1(0|\theta_1)f_2(0), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 0,\\
f_1(0|\theta_1)f_2(1), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 1,\\
f_1(0|\theta_1)\sum_{j=2}^{+\infty} f_2(j), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 2,\\
\sum_{i=2}^{+\infty} f_1(i|\theta_1)(1-\alpha)^{i-1}f_2(0) + \sum_{i=2}^{+\infty} f_1(i|\theta_1)(1-\alpha)^{i-1}\sum_{j=1}^{+\infty} f_2(j) & \\
\quad + \sum_{i=2}^{+\infty}\sum_{k=1}^{i-1} f_1(i|\theta_1)\binom{i-1}{k}\alpha^k(1-\alpha)^{i-k-1}, & \text{if } s_1 = 2,\ s_{21} = 0 \text{ and } s_{22} = 0,
\end{cases}\\
&= \begin{cases}
f^1_{\mathcal{U}}(0,0,0|\theta_1), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 0,\\
f^1_{\mathcal{U}}(0,0,1|\theta_1), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 1,\\
f^1_{\mathcal{U}}(0,0,2|\theta_1), & \text{if } s_1 = 0,\ s_{21} = 0 \text{ and } s_{22} = 2,\\
f^1_{\mathcal{U}}(1,0,0|\theta_1) + f^1_{\mathcal{U}}(1,0,1|\theta_1) + f^1_{\mathcal{U}}(1,1,0|\theta_1), & \text{if } s_1 = 2,\ s_{21} = 0 \text{ and } s_{22} = 0.
\end{cases}
\end{aligned}
$$

Following the same logic in the proof of Proposition 4, we can prove the "stock less" result.

Proof of Inequality (11): For $i = 1, \cdots, N-1$, we have that

$$
\begin{aligned}
&G^{\mathcal{O}}_i(y+1, \phi_i) - G^{\mathcal{O}}_i(y, \phi_i)\\
&= E_{\phi_i(\theta,\alpha)}\Bigg\{\pi(y+1|\theta_1,\theta_2,\alpha) - \pi(y|\theta_1,\theta_2,\alpha)\\
&\qquad + \delta\sum_{x_1=y+1}^{+\infty}\sum_{x_{22}=0}^{+\infty}\Big[\sum_{x_{21}=0}^{x_1-y-1} v^{\mathcal{OS}}_{i+1}(\phi_{i+1})\binom{x_1-y-1}{x_{21}}\alpha^{x_{21}}(1-\alpha)^{x_1-y-1-x_{21}}\\
&\qquad - \sum_{x_{21}=0}^{x_1-y} v^{\mathcal{OS}}_{i+1}(\phi_{i+1})\binom{x_1-y}{x_{21}}\alpha^{x_{21}}(1-\alpha)^{x_1-y-x_{21}}\Big] f_1(x_1|\theta_1)f_2(x_{22}|\theta_2)\Bigg\}.
\end{aligned}
$$

Similar to the proof of Proposition 4 in Chen and Plambeck (2008), we can show that for any period $i = 1, \cdots, N-1$, given the prior $\phi_i(\theta,\alpha)$,

$$
E_{\phi_i(\theta,\alpha)}\Bigg\{\sum_{x_1=y+1}^{+\infty}\sum_{x_{22}=0}^{+\infty}\sum_{x_{21}=0}^{x_1-y-1} v^{\mathcal{OS}}_{i+1}(\phi_{i+1})\binom{x_1-y-1}{x_{21}}\alpha^{x_{21}}(1-\alpha)^{x_1-y-1-x_{21}} f_1(x_1|\theta_1)f_2(x_{22}|\theta_2)\Bigg\}
$$

$$\leq E_{\phi_i(\theta,\alpha)}\Bigg\{\sum_{x_1=y+1}^{+\infty}\sum_{x_{22}=0}^{+\infty}\sum_{x_{21}=0}^{x_1-y} v_{i+1}^{\mathcal{OS}}(\phi_{i+1})\binom{x_1-y}{x_{21}}\alpha^{x_{21}}(1-\alpha)^{x_1-y-x_{21}}f_1(x_1|\theta_1)f_2(x_{22}|\theta_2)\Bigg\}.$$

Thus, the inequality (11) holds.

Proof of Proposition 7: We can see that (a) implies (b) by taking the maximum over the inequalities in (a). So, we only need to prove (a) here.

First, we use the backward induction to show that $\sum_{n=0}^{N-i}\delta^n\cdot\pi(y,\phi_i)\leq G_i^{\mathcal{U}}(y,\phi_i)$ for $i=1,\cdots,N$. When $i=N$, it holds for sure. Assume the result holds for period $i+1$ ($i=1,\cdots,N-1$), which means that $\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\leq G_{i+1}^{\mathcal{U}}(y,\phi_{i+1})$, and thus $\sum_{n=0}^{N-i-1}\delta^n\cdot\max\limits_{0<y\leq M}\pi(y,\phi_{i+1})\leq v_{i+1}^{\mathcal{U}}(\phi_{i+1})$. Now, for period i, we have

$$\begin{aligned}
&G_i^{\mathcal{U}}(y,\phi_i)\\
=&E_{\phi_i(\theta_1,\theta_2,\alpha)}\Bigg\{\pi(y|\theta_1,\theta_2,\alpha)+\delta\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)\\
&+\delta\sum_{s_1=0}^{y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)+\delta\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)\\
&+\delta\sum_{s_{21}=0}^{M-y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)+\delta v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\Bigg\}\\
\geq&E_{\phi_i(\theta_1,\theta_2,\alpha)}\Bigg\{\pi(y|\theta_1,\theta_2,\alpha)+\delta\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1}\left[\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\right]f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)\\
&+\delta\sum_{s_1=0}^{y-1}\left[\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\right]f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)\\
&+\delta\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1}\left[\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\right]f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)\\
&+\delta\sum_{s_{21}=0}^{M-y-1}\left[\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\right]f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)\\
&+\delta\left[\sum_{n=0}^{N-i-1}\delta^n\cdot\pi(y,\phi_{i+1})\right]f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\Bigg\}\\
\geq&E_{\phi_i(\theta_1,\theta_2,\alpha)}\{\pi(y|\theta_1,\theta_2,\alpha)\}\\
&+\sum_{n=1}^{N-i}\delta^n\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1}\pi(y,\phi_{i+1}(\theta_1',\theta_2',\alpha'|s_1,0,s_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\sum_{n=1}^{N-i}\delta^n\sum_{s_1=0}^{y-1}\pi(y,\phi_{i+1}(\theta_1',\theta_2',\alpha'|s_1,0,M-s_1-1,y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)\}
\end{aligned}$$

$$+\sum_{n=1}^{N-i}\delta^n\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1}\pi(y,\phi_{i+1}(\theta_1',\theta_2',\alpha'|y,s_{21},s_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)\}$$
$$+\sum_{n=1}^{N-i}\delta^n\sum_{s_{21}=0}^{M-y-1}\pi(y,\phi_{i+1}(\theta_1',\theta_2',\alpha'|y,s_{21},M-y-s_{21},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)\}$$
$$+\sum_{n=1}^{N-i}\delta^n\pi(y,\phi_{i+1}(\theta_1',\theta_2',\alpha'|y,M-y,0,y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\}$$
$$=\sum_{n=0}^{N-i}\delta^n\cdot\pi(y,\phi_i),$$

where the last equality is based on the law of total expectation, whose formal proof is similar to that of Lemma 1(a) in Chen (2010).

Next, we use the backward induction to show $G_i^{\mathcal{U}}(y,\phi_i)\le G_i^{\mathcal{O}}(y,\phi_i)$ $(i=1,\cdots,N)$. When $i=N$, it holds for sure. Assume that the result holds for period $i+1$ $(i=1,\cdots,N-1)$, which means that $G_{i+1}^{\mathcal{U}}(y,\phi_{i+1})\le G_{i+1}^{\mathcal{O}}(y,\phi_{i+1})$, and thus $v_{i+1}^{\mathcal{U}}(\phi_{i+1})\le v_{i+1}^{\mathcal{O}}(\phi_{i+1})$. Then, for period i, we have

$$G_i^{\mathcal{U}}(y,\phi_i)$$
$$=E_{\phi_i(\theta_1,\theta_2,\alpha)}\Bigg\{\pi(y|\theta_1,\theta_2,\alpha)+\delta\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)$$
$$+\delta\sum_{s_1=0}^{y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)+\delta\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)$$
$$+\delta\sum_{s_{21}=0}^{M-y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)+\delta v_{i+1}^{\mathcal{U}}(\phi_{i+1})f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\Bigg\}$$
$$=E_{\phi_i(\theta_1,\theta_2,\alpha)}\{\pi(y|\theta_1,\theta_2,\alpha)\}$$
$$+\delta\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1=s_1,x_{21}=0,x_{22}=s_{22},y,\phi_i))$$
$$\cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)\}$$
$$+\delta\sum_{s_1=0}^{y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1=s_1,x_{21}=0,x_{22}\ge M-s_1,y,\phi_i))$$
$$\cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)\}$$
$$+\delta\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\ge y+s_{21},x_{21}=s_{21},x_{22}=s_{22},y,\phi_i))$$
$$\cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)\}$$
$$+\delta\sum_{s_{21}=0}^{M-y-1}v_{i+1}^{\mathcal{U}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\ge y+s_{21},x_{21}=s_{21},x_{22}\ge M-y-s_{21},y,\phi_i))$$

$$
\begin{aligned}
&\qquad\qquad \cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)\}\\
&+\delta v_{i+1}^{\mathcal{U}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\geq M,x_{21}\geq M-y,x_{22}\geq 0,y,\phi_i))\\
&\qquad\qquad \cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\}\\
\leq & E_{\phi_i(\theta_1,\theta_2,\alpha)}\{\pi(y|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{s_1=0}^{y-1}\sum_{s_{22}=0}^{M-s_1-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1=s_1,x_{21}=0,x_{22}=s_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,s_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{s_1=0}^{y-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1=s_1,x_{21}=0,x_{22}\geq M-s_1,y,\phi_i))\\
&\qquad\qquad \cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(s_1,0,M-s_1-1|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{s_{21}=0}^{M-y-1}\sum_{s_{22}=0}^{M-y-s_{21}-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\geq y+s_{21},x_{21}=s_{21},x_{22}=s_{22},y,\phi_i))\\
&\qquad\qquad \cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},s_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{s_{21}=0}^{M-y-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\geq y+s_{21},x_{21}=s_{21},x_{22}\geq M-y-s_{21},y,\phi_i))\\
&\qquad\qquad \cdot E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,s_{21},M-y-s_{21}|\theta_1,\theta_2,\alpha)\}\\
&+\delta v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1\geq M,x_{21}\geq M-y,x_{22}\geq 0,y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{U}}^y(y,M-y,0|\theta_1,\theta_2,\alpha)\}\\
\leq & E_{\phi_i(\theta_1,\theta_2,\alpha)}\{\pi(y|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{x_1=0}^{y-1}\sum_{x_{22}=0}^{M-x_1-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1,x_{21},x_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{O}}^y(x_1,x_{21},x_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{x_1=0}^{y-1}\sum_{x_{22}=M-x_1}^{+\infty} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1,x_{21},x_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{O}}^y(x_1,x_{21},x_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{x_{21}=0}^{M-y-1}\sum_{x_1=y+x_{21}}^{+\infty}\sum_{x_{22}=0}^{M-y-x_{21}-1} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1,x_{21},x_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{O}}^y(x_1,x_{21},x_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{x_{21}=0}^{M-y-1}\sum_{x_1=y+x_{21}}^{+\infty}\sum_{x_{22}=M-y-x_{21}}^{+\infty} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1,x_{21},x_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{O}}^y(x_1,x_{21},x_{22}|\theta_1,\theta_2,\alpha)\}\\
&+\delta\sum_{x_{21}=M-y}^{+\infty}\sum_{x_1=y+x_{21}}^{+\infty}\sum_{x_{22}=0}^{+\infty} v_{i+1}^{\mathcal{O}}(\phi_{i+1}(\theta_1',\theta_2',\alpha'|x_1,x_{21},x_{22},y,\phi_i))E_{\phi_i(\theta_1,\theta_2,\alpha)}\{f_{\mathcal{O}}^y(x_1,x_{21},x_{22}|\theta_1,\theta_2,\alpha)\}\\
= & G_i^{\mathcal{O}}(y,\phi_i),
\end{aligned}
$$

where the last inequality can be formally proved by following the procedure stated in the proof of Lemma 1(b) in Chen (2010). Here, we omit the details.